TABLE OF CONTENTS November 2005

EDITORIAL

Dexamethasone to improve maternal outcome in women with hemolysis, elevated liver enzymes, and low platelets syndrome 1587
Baha M. Sibai, MD, John R. Barton, MD
Cincinnati, OH, and Lexington, KY

Dexamethasone does not improve maternal outcome in women with HELLP syndrome.

EDITORS’ CHOICE

Dexamethasone treatment does not improve the outcome of women with HELLP syndrome: A double-blind, placebo-controlled, randomized clinical trial 1591
Javier E. Fonseca, MD, MSc, Fabián Méndez, MD, PhD, Claudia Cataño, MD, Fernando Arias, MD, PhD
Cali, Colombia, and Toledo, OH

Dexamethasone treatment does not improve the outcome of women with HELLP syndrome.

CLINICAL OPINION

WHI clinical trial revisit: Imprecise scientific methodology disqualifies the study’s outcomes 1599
Adam Ostrzenski, MD, PhD, Katarzyna M. Ostrzenska, MD
St Petersburg, FL

The Women’s Health Initiative clinical trial did not provide evidence for or against the use of hormone replacement therapy in postmenopausal women because of the deficiency in the study’s methodology and its execution.

Commentary
In this issue of the Journal, Ostrzenski and Ostrzenska challenge some of the conclusions of the Women’s Health Initiative (WHI) Estrogen plus Progestin trial. Specifically, they question some of the methodology employed in the trial. Because of the importance of the study and the significant impact on women’s health care issue, we have invited the WHI investigators to respond to the conclusions of these authors. Barbara Cochrane et al responded, and their comprehensive reply can also be found in this issue. The WHI represents a massive undertaking, and investigators will continue to examine its conclusions. This effort will continue to enhance our understanding of the important findings of this trial.

Contents continued on page 5A
WHI response to Ostrzenski and Ostrzenska
Barbara B. Cochrane, PhD, RN, David H. Barad, MD, MS, Margery Gass, MD, Robert L. Brunner, PhD, Cora E. Lewis, MD, MSPH, Marcia L. Stefanick, PhD
Seattle, WA, Bronx, NY, Cincinnati, OH, Reno, NV, Birmingham, AL, and Palo Alto, CA

REVIEW ARTICLE
Evidence-based surgery for cesarean delivery
Vincenzo Berghella, MD, Jason K. Baxter, MD, MSCP, Suneet P. Chauhan, MD
Philadelphia, PA, and West Allis, WI

Cesarean delivery techniques that are supported by good quality recommendations should be performed routinely; cesarean delivery techniques that have lower quality recommendations deserve further research.

GENERAL OBSTETRICS AND GYNECOLOGY: GYNECOLOGY
Uterine artery embolization versus hysterectomy in the treatment of symptomatic uterine fibroids (EMMY trial): Peri- and postprocedural results from a randomized controlled trial
Wouter J. K. Hehenkamp, MD, Nicole A. Volkers, MD, Peter F. J. Donderwinkel, MD, Sjoerd de Blok, MD, PhD, Erwin Birnie, PhD, Willem M. Ankum, MD, PhD, Jim A. Reekers, MD, PhD
Amsterdam and Groningen, The Netherlands

The safety of uterine artery embolization in a randomized comparison to hysterectomy for the treatment of uterine fibroids that cause menorrhagia was demonstrated.

Ovarian cancer screening in the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial: Findings from the initial screen of a randomized trial
Saundra S. Buys, MD, Edward Partridge, MD, Mark H. Greene, MD, Philip C. Prorok, PhD, Douglas Reding, MD, Thomas L. Riley, Patricia Hartge, ScD, Richard M. Fagerstrom, PhD, Lawrence R. Ragard, MD, David Chia, PhD, Grant Izmirlian, PhD, Mona Fouad, MD, Christine C. Johnson, PhD, John K. Gohagan, PhD, for the PLCO Project Team
Salt Lake City, UT, Birmingham, AL, Bethesda and Rockville, MD, Marshfield, WI, Los Angeles, CA, and Detroit, MI

The initial ovarian cancer screening with transvaginal ultrasound and CA-125 in 28,816 women enrolled in the PLCO Trial identified 29 neoplasms, including 9 borderline tumors.

Endometrial cancer in women 45 years of age or younger: A clinicopathological analysis
Gilbert P. Pellerin, MD, Michael A. Finan, MD
New Orleans, LA

Prognostic factors and survival are analyzed in young women with endometrial cancer.
Laparoscopic staging in patients with incompletely staged cancers of the uterus, ovary, fallopian tube, and primary peritoneum: A Gynecologic Oncology Group (GOG) study
Nick M. Spirtos, MD, Scott M. Eisekop, MD, Guy Boike, MD, John B. Schlaerth, MD, James O. Cappellari, MD
Los Gatos, Tarzana, and Pasadena, CA, Saginaw and East Lansing, MI, and Winston-Salem, NC
Interval staging using laparoscopy can safely identify occult disease in selected patients; risks of laparotomy and visceral injury should not be understated.

Uterine innervation after hysterectomy for chronic pelvic pain with, and without, endometriosis
Gurprit Atwal, MB ChB, Daniel du Plessis, MD, Gordon Armstrong, MD, Richard Slade, MD, Martin Quinn, MB ChB, MD
Manchester, UK
Histopathologic features associated with reinnervation have been observed in the uterine isthmus after hysterectomy for chronic pelvic pain.

GENERAL OBSTETRICS AND GYNECOLOGY: OBSTETRICS
Maternal complications with vaginal birth after cesarean delivery: A multicenter study
George A. Macones, MD, MSCE, Jeffrey Peipert, MD, MPH, Deborah B. Nelson, PhD, Anthony Odibo, MD, Erika J. Stevens, MA, David M. Stamilio, MD, MSCE, Emmanuelle Pare, MD, Michal Elovitz, MD, Anthony Sciscione, DO, Mary D. Sammel, ScD, Sarah J. Ratchiffe, PhD
Philadelphia, PA, and Providence, RI
Most research on vaginal birth after cesarean delivery has focused on tertiary care institutions and has been of limited sample size.

Increased intratuterine frequency of *Ureaplasma urealyticum* in women with preterm labor and preterm premature rupture of the membranes and subsequent cesarean delivery
Armin Witt, MD, Angelika Berger, MD, Christian J. Gruber, MD, Ljubomir Petricevic, MD, Petra Apfalter, MD, Christof Worda, MD, Peter Husslein, MD
Vienna, Austria
In patients with preterm premature rupture of membranes and/or preterm labor and subsequent preterm cesarean delivery, intra- amniotic colonization with Ureaplasma urealyticum was increased significantly compared with patients with other indications for preterm cesarean delivery.

Persistance of adverse obstetric and neonatal outcomes in monochorionic twins after exclusion of disorders unique to monochorionic placentation
Line Leduc, MD, Larissa Takser, MD, PhD, Denyse Rinfret, RN
Montréal, Québec, Canada
Monochorionic placentation remains a risk factor for adverse obstetric and neonatal outcomes even after exclusion of disorders unique to monochorionic gestations.
Hypertensive disease in pregnancies complicated by systemic lupus erythematosus

Robert S. Egerman, MD, Risa D. Ramsey, PhD, Lu W. Kao, RN, Jay J. Bringman, MD, Andrew J. Bush, PhD, Jim Y. Wan, PhD
Memphis, TN

The percentage of parturients with systemic lupus erythematosus who develop preeclampsia is increased, regardless of the presence of underlying chronic hypertension.

Gestational age-specific predicted risk of neonatal respiratory distress syndrome using lamellar body count and surfactant-to-albumin ratio in amniotic fluid

Raymond Karcher, PhD, Elizabeth Sykes, MD, Daniel Batton, MD, Zi Uddin, PhD, Gary Ross, DO, Elaine Hockman, PhD, George H. Shade, Jr, MD
Royal Oak, Detroit, and Madison Heights, MI

Gestational age-specific predicted risk of respiratory distress syndrome in newborn from lamellar body count and surfactant-to-albumin ratio in amniotic fluid are equally accurate.

The mean weekly increment of amniotic fluid TDx-FLM II ratio is constant during the latter part of pregnancy

Ibrahim Bildirici, MD, Christopher N. Moga, MD, Ann M. Gronowski, PhD, Yoel Sadovsky, MD
St Louis, MO

The mean weekly increment in TDx-FLM II ratio that was obtained by serial amniocentesis is constant between 31 and 38 weeks of gestation.

Maternal plasma concentrations of IGF-1, IGFBP-1, and C-peptide in early pregnancy and subsequent risk of gestational diabetes mellitus

Chunfang Qiu, MD, MS, Surab Vadachkoria, MD, PhD, Lois Meryman, BA, Ihunnaya O. Frederick, MPH, Michelle A. Williams, ScD
Seattle, WA

Alterations of maternal plasma free IGF-1, IGFBP-1, and C-peptide concentrations in early pregnancy are associated with GDM risk.

Protein Z in patients with pregnancy complications

Florence Bretelle, MD, PhD, Dominique Arnoux, PhD, Raha Shojai, MD, Claude D’Ercole, MD, José Sampol, PhD, Françoise Dignat, PhD, Laurence Camoin-Jau, PhD
Marseille, France

An association between Protein Z deficiency and pregnancy complications such as intrauterine demise and intrauterine growth restriction has been found.

A free radical scavenger, edaravone, inhibits lipid peroxidation and the production of nitric oxide in hypoxic-ischemic brain damage of neonatal rats

Jesmin I. Noor, MD, Tomoaki Ikeda, MD, Yuto Ueda, MD, Tsuyomu Ikenoue, MD
Miyazaki, Japan

A potent free radical scavenger, edaravone, inhibits both lipid peroxidation and nitric oxide production in neonatal rats and may be a highly effective and potent drug for the prevention of hypoxic-ischemic encephalopathy in newborn infants.
TABLE OF CONTENTS continued

Embryogenesis of fused umbilical arteries in human embryos 1709
Shigehito Yamada, MD, Junzo Hamanishi, MD, Shozo Tanada, MD, Mitsuhiro Tachibana, MD, Rokuro Mimura, MD, Shingo Fujii, MD, Kohei Shiota, MD
Kyoto and Amagasaki, Hyogo, Japan

Embryogenesis of fused umbilical arteries was examined in early human embryos, and its relevance to fetal pathologic condition was discussed.

Insulin and fatty acids regulate the expression of the fat droplet-associated protein adipophilin in primary human trophoblasts 1716
Uriel Elchalal, MD, W. Timothy Schaiff, PhD, Steven D. Smith, BA, Eli Rimon, MD, Ibrahim Bildirici, MD, D. Michael Nelson, MD, PhD, Yoel Sadovsky, MD
St. Louis, MO

Insulin and fatty acids enhance adipophilin expression and the formation of fatty acid droplets in primary human trophoblasts.

Human preterm amnion cells cultured in 3-dimensional collagen I and fibrin matrices for tissue engineering purposes 1724
Grozdana Bilic, PhD, Heike Hall, PhD, Anne Greet Bittermann, PhD, Prisca Zammeretti, PhD, Tilo Burkhart, MD, Nicole Oechslein-Kindblom, MD, Roland Zimmermann, MD
Zurich, Switzerland

Three-dimensional fibrin matrices might be useful in amnion cell tissue engineering, including cell-matrix transplantation.

EDUCATION

The American Association of Obstetricians and Gynecologists Foundation Scholars Program: Additional data on research-related outcomes 1733
Georgine M. Pion, PhD, Charles B. Hammond, MD
Nashville, TN, and Durham, NC

MDs with AAOFG-sponsored research training have actively pursued research careers in academic medicine, as indicated by their faculty status, receipt of NIH funding, and publications.

CASE REPORTS

Ischiorectal abscess after sacrospinous ligament suspension 1740
Michael Hibner, MD, Jeffrey L. Cornella, MD, Javier F. Magrina, MD, Jacques P. Heppell, MD
Scottsdale, AZ

An ischiorectal abscess, an uncommon complication after sacrospinous fixation resolved after treatment with a perianal incision, drainage, and intravenous antibiotics.

Application of the three-dimensional maximum mode in prenatal diagnosis of Apert syndrome 1743
Tilman Esser, MD, Patrick Rogalla, MD, Christian Bamberg, MD, Karim D. Kalache, MD
Berlin, Germany

Three-dimensional maximum mode is an important tool for the correct prenatal diagnosis of Apert syndrome, because it helps to demonstrate the specific cranial deformities.
Placement of a temporary vena cava filter during labor
Steven L. Clark, MD, Duane D. Blatter, MD, G. Marc Jackson, MD
Salt Lake City, UT

We describe the placement of a removable vena cava filter during labor.

Maternal death caused by midgut volvulus after bariatric surgery
Paul V. Loar, III, MD, Luis Sanchez-Ramos, MD, Andrew M. Kaunitz, MD, Andrew J. Kerwin, MD, Jesus Diaz, MD
Jacksonville, FL

A pregnancy after laparoscopic gastric bypass surgery was complicated by midgut volvulus, intestinal ischemia, small bowel perforation, and death.

Vulvar cellular angiofibroma: A case report
Ryan Kerkuta, MD, Colleen M. Kennedy, MD, MS, Jo A. Benda, MD, Rudolph P. Galask, MD, MS
Iowa City, IA

We present a case of a woman with vulvar cellular angiofibroma treated with simple excision. There has been no recurrence after 10 months of follow-up.

A tale of 2 pedunculated myomas
Ihab M. Usta, MD, Elie M. Hobeika, MD, Anwar H. Nassar, MD
Beirut, Lebanon

Pedunculated myomas might have unusual configuration and behavior, which are factors that should be considered in decisions regarding their operative treatment.

Temporary balloon occlusion of the common iliac artery: New approach to bleeding control during cesarean hysterectomy for placenta percreta
Jin-Chung Shih, MD, Kao-Lang Liu, MD, Ming-Kwang Shyu, MD
Taipei, Taiwan

Temporary balloon occlusion of bilateral common iliac arteries is a simple and efficient technique that provides satisfactory bleeding control during cesarean hysterectomy for placenta percreta.

Episiotomy dehiscence that required intestinal diversion
Carl H. Rose, MD, Kristi L. Blessitt, MD, Farshid Araghizadeh, MD, John C. Morrison, MD
Jackson, MS

Episiotomy infection requires early recognition and thorough evaluation to exclude occult rectal injury.

Endometriotic umbilical port site metastasis after laparoscopy
Umut Barbaros, MD, Ahmet Cem Iybozkurt, MD, Mine Gulluoglu, MD, Merve Barbaros, MD, Yestim Erbil, Vaht Turanli, Selcuk Mercan
Istanbul, Turkey

Cyclic pain and color on umbilical region should alert the physicians to suspect either umbilical primary endometriosis or umbilical port site metastasis of endometriosis following a minimal invasive treatment of this disorder.

Contents continued on page 10A
CLASSIC PAGES IN OBSTETRICS AND GYNECOLOGY

On the Sites of the Negative and Positive Feedback Actions of Estradiol in the Control of Gonadotropin Secretion in the Rhesus Monkey
Y. Nakai, T. M. Plant, D. L. Hess, E. J. Keogh, Ernst Knobil

An excerpt from Endocrinology 1978;102:1008-1014 with a commentary by Lawrence D. Longo, MD, followed by Discovery of the hypothalamic gonadotropin-releasing hormone pulse generator and of its physiologic significance by Ernst Knobil, MD, reprinted from Endocrinology 1992;131:1005-1006.

TRANSACTIONS FOR THE 2005 COUNCIL ON RESIDENT EDUCATION IN OBSTETRICS AND GYNECOLOGY AND THE ASSOCIATION OF PROFESSORS OF GYNECOLOGY ANNUAL MEETING

Steadfastly forward
Timothy R. B. Johnson, MD
Ann Arbor, MI

“Every time I look back, I meet the eyes of my foreparents looking steadfastly forward.”
—Professor Roger Shinn

Active transformation of our specialty will ensure a positive future.

The essential elements of undergraduate medical education in obstetrics and gynecology: A comparison of the Association of Professors of Gynecology and Obstetrics Medical Student Educational Objectives and the National Board of Medical Examiners Subject Examination
Maya M. Hammoud, MD, Susan M. Cox, MD, Barbara Goff, MD, Alice Goepfert, MD, Aggie Butler, PhD, David B. Swanson, PhD, Kathleen Z. Holtzman, BS, Krista Albee, BA, Nadine T. Katz, MD, Sonya S. Erickson, MD
Ann Arbor, MI, Dallas, TX, Seattle, WA, Birmingham, AL, Philadelphia, PA, New York, NY, and Denver, CO

The Association of Professors of Gynecology and Obstetrics Medical Student Educational Objectives and the National Board of Medical Examiners Obstetrics and Gynecology Subject Examination agree on the essential elements of Undergraduate Medical Education in obstetrics-gynecology.

Medical students self-reported work hours: Perception versus reality
Colleen Casey, MD, Sangeeta Senapati, MD, Casey B. White, PhD, Larry D. Gruppen, PhD, Maya M. Hammoud, MD
Ann Arbor, MI

Medical students self-reported work hours during the obstetrics and gynecology third-year clerkship are higher than the actual number of hours they are scheduled to work.

Development and assessment of a Web-based evaluation and management coding curriculum for residents
Paul M. Lemen, MD
Milwaukee, WI

An electronic learning module on evaluation and management coding for residents, using scoring based on reimbursement schedules, is effective for instruction and assessment and well accepted by learners.
Teaching residents coding and documentation: Effectiveness of a problem-oriented approach

Sawsan As-Sanie, MD, MPH, Denniz Zolnoun, MD, MPH, Mary Ellen Wechter, MD, Georgine Lamvu, MD, MPH, Frank Tu, MD, MPH, John Steege, MD
Chapel Hill, NC, Ann Arbor, MI, and Chicago, IL

Problem-oriented interactive learning appears to be an effective method of teaching residents proper coding and documentation.

Implementation and evaluation of a genetics curriculum to improve obstetrician-gynecologist residents’ knowledge and skills in genetic diagnosis and counseling

Charles J. Macri, MD, Nancy D. Gaba, MD, Lauren M. Sitzer, MD, Lisa Freese, MS, Susanne L. Barthgate, MD, John W. Larsen Jr, MD
Washington, DC

To determine whether a genetics curriculum for obstetrician-gynecologist residents using needs assessment, pretest, educational intervention, Web-based resources, and case-based learning with standardized patients improves posttest performance.

Obstetrics-gynecology resident satisfaction

Bridgette A. Blazek, MD, Terrell W. Zollinger, DrPH, Katherine Y. Look, MD
Indianapolis, IN

Relevant training, collegiality, adequate resources, workload, care continuity, supportive coworkers, learning environment, autonomy, supportive faculty positively have an impact on obstetrics-gynecology resident satisfaction.

Residency attrition rate in obstetrics and gynecology: Are we losing more postgraduates today?

Maria Manriquez Gilpin, MD
Morgan Hill, CA

The resident attrition rate is higher today than in previous reports; however, most residents are transferring to other obstetrics and gynecology residencies.

Personality type and clinical evaluations in an obstetrics/gynecology medical student clerkship

Katrina R. Davis, MD, Joseph A. Banken, PhD
Little Rock, AR

Clinical evaluation data of medical students in an obstetrics/gynecology clerkship are not correlated with National Board of Medical Examiners scores but are correlated with Myers-Briggs type inventory extraversion.

Focused assessment of surgical performance: Difficulty with faculty compliance

Gabriella G. Gosman, MD, Hyagriv N. Simhan, MD, MSCR, Richard S. Guido, MD, Ted T. M. Lee, MD, Suketu M. Mansuria, MD, Joseph S. Sanfilippo, MD, MBA
Pittsburgh, PA

Faculty members had a low rate of compliance when asked to rate resident surgical performance with a validated instrument after every surgical case.
TABLE OF CONTENTS continued

Self-assessment of resident surgical skills: Is it feasible? 1817
Lynn S. Mandel, PhD, Barbara A. Goff, MD, Gretchen M. Lentz, MD
Seattle, WA
Resident self-assessment of surgical proficiency on the Objective Structured Assessment of Technical Skills correlates highly with faculty ratings of the same skills.

Resident job satisfaction: One year of duty hours 1823
Kirsten J. Lund, MD, Stephanie B. Teal, MD, MPH, Ruben Alvero, MD
Denver, CO
Accreditation Council for Graduate Medical Education duty-hour restrictions may not change overall resident job satisfaction significantly or increase time spent teaching, but they may increase satisfaction with specific aspects of training.

The influence of an audience response system on knowledge retention: An application to resident education 1827
Archana Pradhan, MD, MPH, Dina Sparano, BS, Cande V. Ananth, PhD, MPH
New Brunswick, NJ
The audience response system is an effective teaching tool for resident educators to increase knowledge retention of lecture material.

Obstetrics and gynecology residents as teachers of medical students: Predictors of excellence 1831
Joseph A. (Tony) Ogburn, MD, Eve L. Espey, MD, MPH, Maxine H. Dorin, MD, Chen Ming, MS, William F. Rayburn, MD
Albuquerque, NM
Work experience, age, and male gender, not measures of academic performance, are the major predictors of intern applicants later found to be excellent teachers.

A multicenter study to determine motivating factors for residents pursuing obstetrics and gynecology 1835
May Hsieh Blanchard, MD, Amy M. Autry, MD, Haywood L. Brown, MD, John R. Musich, MD, Leah Kaufman, MD, Dylan R. Wells, MD, Robert D. Stager, MD, Jennifer L. Swanson, MD, Kirsten J. Lund, MD, Donald W. Wiper III, MD, Jennifer L. Bailit, MD, MPH
Cleveland, OH, San Francisco, CA, Durham, NC, Royal Oak, MI, New Hyde Park, NY, Memphis, TN, Augusta, GA, Pittsburgh, PA, Denver, CO, and Portland, ME
Residents pursuing training in obstetrics and gynecology choose the field for surgical opportunities and variety in day-to-day practice.

Candid candidate comments: The relationship between residency program selection factors and match list placements from ranked applicants 1842
Marilyn J. Raymond, PhD, PT, Robert J. Sokol, MD, Louis A. Vontver, MD, MEd, Kenneth A. Ginsburg, MD
Detroit, MI, and Seattle, WA
Three residency program selection factors, meeting candidates’ career goals, faculty-resident relationships, and location, were associated with receipt of top match list ratings from ranked applicants.
Educational games in an obstetrics and gynecology core curriculum
Sharon O’Leary, MD, Lisa Diepenhorst, MD, Ruth Churley-Strom, MS, RN, Diane Magrane, MD
Ypsilanti and Ann Arbor, MI
Academic instruction using an interactive game to teach ectopic pregnancy content to third-year medical students is as effective and more enjoyable than traditional lecture.

Medical student identification of domestic violence as measured on an objective, standardized clinical examination
Susan E. Hoffstetter, PhD, WHNP, Robert J. Blaskiewicz, MD, Gail E. Furman, PhD, Jennifer A. McCabe, BS
St Louis, MO
A low rate of domestic violence screening using history by medical students was found in a standardized patient setting of nonspecific abdominal pain.

A new curriculum for hysteroscopy training as demonstrated by an objective structured assessment of technical skills (OSATS)
Amy L. VanBlaricom, MD, Barbara A. Goff, MD, Michael Chinn, MD, Melodie M. Icasiano, MD, Peter Nielsen, MD, Lynn Mandel, PhD
Seattle, WA
A hysteroscopy curriculum improved knowledge and technical skill in obstetrics and gynecology residents.

Measurement of endometrial stripe thickness by obstetrics and gynecology residents
Daniel M. Breitkopf, MD, Edward R. Smith, PhD, William N. P. Herbert, MD
Galveston, TX, and Charlottesville, VA
By the end of residency, obstetrics and gynecology residents can accurately measure the endometrial stripe thickness with transvaginal sonography.

Impact of 1996 Residency Review Committee obstetrics-gynecology primary care requirements on residency training and surgical procedures
Margaret W. Chu, MD, Martha J. Rall, MS, Linda M. Frazier, MD, MPH, Douglas V. Horbelt, MD, Travis W. Stembridge, MD
Wichita, KS
Recent residency graduates felt better trained in primary care and had adequate experience in major inpatient surgical procedures but thought obstetrics-gynecology is not a primary care specialty.

Improved performance and student satisfaction after implementation of a problem-based preclinical obstetrics and gynecology curriculum
Petra M. Casey, MD, Diane Magrane, MD, Timothy G. Lesnick, MS
Rochester, MN
Students’ performance on written examinations and key aspects of student satisfaction improved significantly when problem-based learning was introduced into preclinical obstetrics and gynecology.
TABLE OF CONTENTS continued

LETTERS TO THE EDITORS

Risk assessment for neonatal respiratory distress syndrome with FLM II combined with gestational age
Michael G. Pinette, MD, Joseph Wax, MD
Portland, ME

Reply
C. A. Parvin, PhD, L. A. Kaplan, PhD, J. F. Chapman, DrPH, T. G. McManamon, PhD, A. M. Gronowski, PhD
St. Louis, MO

Lymphatic mapping for endometrial cancer: Is hysteroscopic injection a safe technique for sentinel lymph node biopsy?
Emmanuel Barranger, MD, Serge Uzan, MD, Emile Darai, MD, PhD
Paris, France

Reply
Francesco Raspagliesi, MD, Antonino Ditto, MD, Shigeki Kusamura, MD, Maria L. Carcangi, MD, Francesca Vecchione, MD, Marco Maccauro, MD, Eugenio Solima, MD
Milan, Italy

Evidence does not support cervical preservation
Oz Harmanli, MD, Stephen A. Metz, MD
Springfield, MA

Reply
Todd R. Jenkins, MD
Charlotte, NC

Fetal monitoring with the ST analyser: Need for a long-term follow-up of the infants
Georges J. J. Boog, MD
Nantes, France

Reply
Robert Gagnon, MD, Kristina L. Dervaitis
London, Ontario, Canada

READER SERVICES

Information for Readers 17A
Professional and educational opportunities 19A
Change of address 1629
EDITORIAL

Dexamethasone to improve maternal outcome in women with hemolysis, elevated liver enzymes, and low platelets syndrome

Baha M. Sibai, MD,a,* John R. Barton, MDb

Department of Obstetrics & Gynecology, University of Cincinnati, Cincinnati, OH

The syndrome of hemolysis, elevated liver enzymes, and low platelets (HELLP) has been described in women with preeclampsia for several decades. For the past 50 years, it has been recognized that the presence of all these laboratory abnormalities in preeclampsia-eclampsia is usually associated with increased rates of maternal complications (Table).1-4 The rate of these complications will depend on the criteria used to establish the diagnosis of HELLP syndrome, the population studied (hospital based or tertiary referral center), and the presence of associated medical, surgical, or obstetric complications (abruption placenta, eclampsia, peripartum hemorrhage).

More than 2 decades since the term HELLP syndrome was first coined by Weinstein5 and after the publication of extensive literature in which the term HELLP syndrome has been described, there is still considerable disagreement with regard to the diagnosis and treatment of this syndrome.1 This disagreement is due in part to a lack of uniform diagnostic criteria for this syndrome and in part to a lack of well-designed, randomized studies for the treatment of this syndrome.1

The diagnostic criteria for HELLP syndrome, and its subclasses, are not as rigorous as one thinks and have varied among investigators from the same country and even between countries.1 For example, a significant percentage of published reports on the subject include patients who had no evidence of hemolysis, which is considered the hallmark of the triad of HELLP syndrome. The term elevated liver enzymes is also a point of confusion because there is no consistency in the literature with regard to which liver function test to use or what degree of elevation in the tests should be used to diagnose elevated liver enzymes. Some studies included elevated amino aspartate transferase (AST) only and some included either elevated AST or amino alanine transferase, and the test values considered abnormal ranged from 17 to 72 IU/L.1 In addition, different cut-off values were used for the diagnosis of low platelet count, ranging from less than 100,000 to less than 150,000/mm3.

The term HELLP syndrome is also used to describe a heterogenous group of patients with severe preeclampsia who have a wide spectrum of clinical findings, laboratory abnormalities, and maternal complications. Initial studies on HELLP syndrome included mainly antepartum patients with misdiagnosed and complicated preeclampsia who were symptomatic and had severe laboratory abnormalities at the time of hospitalization.5,6 More recently, however, the diagnosis of HELLP syndrome is often reported in asymptomatic pregnant women with hypertension or preeclampsia who are incidentally found to have increased liver function tests or thrombocytopenia.7 In essence, these patients have biochemical findings for HELLP syndrome discovered as a consequence of the increased frequency of obtaining laboratory evaluation in women with preeclampsia. Consequently, these studies referred to a variety of conditions that may have a similar definition of HELLP syndrome but may not necessarily

* Reprint requests: Baha M. Sibai, MD, Professor, University of Cincinnati, 231 Albert Sabin Way, Cincinnati, OH 45267.
E-mail: baha.sibai@uc.edu

0002-9378/S - see front matter © 2005 Mosby, Inc. All rights reserved.
doi:10.1016/j.ajog.2005.08.006
For delivery. Because the presence of this syndrome has been confirmed, a decision must be made concerning the need times sudden deterioration in the maternal condition. The clinical course of women with true HELLP syndrome is usually characterized by progressive and some-bidity and mortality, some authors considered its presence an indication for immediate delivery. 

There is considerable disagreement, however, as to the treatment of women with HELLP syndrome at or before 34 weeks of gestation when the maternal condition is stable except for mild to moderate abnormalities in blood test results and a reassuring fetal condition. In such patients, some authors recommend the administration of corticosteroids to accelerate fetal lung maturity followed by delivery within 48 hours, whereas others recommend administration of corticosteroids for both maternal and fetal benefit. In the latter case, some recommend corticosteroid use only antepartum, whereas others recommend only postpartum and some recommend both antepartum and postpartum use.

The maternal benefits of corticosteroids in women with HELLP syndrome were first reported by Thiagarajah et al. In this study, the authors noted improvement in laboratory tests in 5 patients treated with corticoste-roidal lung maturity enhancement. Since then corticosteroids have been suggested as safe and effective drugs for improving maternal and neonatal outcome in women with HELLP or partial HELLP syndrome. Reported maternal benefits of corticosteroids have included improvement in laboratory values, improvement in hourly urine output, improvement in blood pressure, shorter hospital stay, and an increased use of regional anesthesia. Most of the reported studies suggesting the benefit from maternal corticosteroids are retrospective and have critical design flaws, mainly in the inclusion criteria, choice of historical controls, and/or the clinical outcome reported. There are only 4 randomized trials that compare dexamethasone plus standard treatment versus standard treatment alone in HELLP syndrome. These studies included a total of 129 patients; only 25 of the randomized patients had the syndrome in the antepartum period. In addition, these studies were not placebo controlled. The results of these randomized trials demonstrated improved maternal laboratory values and improved urine output in patients receiving dexamethasone but no differences in serious maternal morbidity such as the need for transfusion, pulmonary edema, renal failure, or serious hepatic complications. Consequently, a recent Cochrane review on the subject concluded that there is insufficient evidence to determine whether adjunctive corticosteroid use in HELLP syndrome decreases major maternal morbidity.

In this issue of the journal, Fonseca et al report the results of the first randomized, double-blind, placebo-controlled clinical trial of the use of dexamethasone treatment to improve maternal outcome in patients with HELLP syndrome. A total of 132 women with HELLP syndrome (60 antepartum and 72 postpartum) were treated with either dexamethasone intravenously or a matching placebo. Pregnant women assigned dexamethasone received 10 mg intravenously every 12 hours until delivery and 3 additional doses after delivery. Postpar-tum women assigned dexamethasone received 310-mg doses after delivery. Only patients with platelet count less than 100,000/mm³, AST greater than 70 IU/L and lactic dehydrogenase greater than 600 IU/L were included in the study. In addition, most patients had elevated bilirubin values, and 49 (37%) had a platelet count 50,000 or less. The primary outcome of the trial was mean days of hospitalization from randomization to discharge. Secondary outcomes were time to recovery of laboratory parameters and maternal complications. The investigators found that the additional treatment with intravenous dexamethasone did not reduce the duration of hospitalization in patients with HELLP syndrome. In addition, the rates of platelets and fresh frozen plasma transfusions as well as maternal complications such as acute renal failure, pulmonary edema, and oliguria were not reduced by treatment with dexamethasone. A surprising finding of this trial was that the

<table>
<thead>
<tr>
<th>Table</th>
<th>Serious maternal complications in HELLP syndrome</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Cardiorespiratory: pulmonary edema, pleural effusions, adult respiratory distress syndrome</td>
<td></td>
</tr>
<tr>
<td>• Renal: acute renal failure, dialysis</td>
<td></td>
</tr>
<tr>
<td>• Hepatic: infarction, hemorrhage, subcapsular hematoma, liver failure</td>
<td></td>
</tr>
<tr>
<td>• Disseminated intravascular coagulopathy</td>
<td></td>
</tr>
<tr>
<td>• Need for blood products: packed red blood cells, platelets, and fresh frozen plasma</td>
<td></td>
</tr>
<tr>
<td>• Stroke, cerebral edema, ischemia, hemorrhage</td>
<td></td>
</tr>
<tr>
<td>• Death</td>
<td></td>
</tr>
</tbody>
</table>
time of recovery of laboratory tests was not shortened by treatment. In addition, in contrast to previous studies,\textsuperscript{9,11} AST recovery was slower in patients assigned to dexamethasone as compared with those assigned to placebo. Moreover, the results were similar whether dexamethasone was given to women with HELLP syndrome before delivery or to those who had postpartum HELLP. It is important to note that this is the first placebo-controlled trial and had the largest sample size. These results do not support the recommendations for routine use of high-dose dexamethasone in patients with HELLP syndrome.

The study by Fonseca et al\textsuperscript{19} warrants several comments. The primary outcome of the trial was duration of hospitalization rather than serious maternal morbidity or mortality. Second, although no statistically significant differences were found in the duration of hospitalization and the average time to platelet count recovery, subgroup analysis according to severity of disease showed that among patients with class 1 HELLP (platelet count less than 50,000 mm\textsuperscript{3}), there was a shorter average platelet count recovery and less duration of hospitalization in those receiving dexamethasone. Third, the authors provided no information about the rates of blood transfusions, regional anesthesia, or cesarean delivery in the study groups.

Several questions remain unanswered by this trial and previous trials with regard to corticosteroid use in HELLP syndrome.

1. Should corticosteroids be used to prolong gestation in women with HELLP syndrome before 32 weeks? Currently there are no randomized trials evaluating the benefit or safety of this therapy. The trial by Fonseca et al\textsuperscript{19} did not provide any answers in this regard. Despite the fact that some studies have shown laboratory values transiently improve during expectant management of HELLP syndrome (with or without dexamethasone),\textsuperscript{11,12,20,21} the safety of such an approach remains unknown. Indeed, we have found that laboratory abnormalities in HELLP syndrome do not correlate with the severity of hepatic histopathology findings\textsuperscript{22} or hepatic imaging findings.\textsuperscript{23}

2. Should dexamethasone be given to all women with HELLP syndrome or restricted to those with severe thrombocytopenia with evidence of hemolysis? It is well established that serious maternal morbidity and the need for transfusion of blood and blood products in women with HELLP syndrome is significantly higher in those with severe thrombocytopenia in association with frank hemolysis as compared with those with minimal laboratory abnormalities in the absence of hemolysis. The study by Fonseca et al\textsuperscript{19} suggested that there is a potential maternal benefit in those with class 1 HELLP syndrome; however, the authors admit that this was a post hoc analysis. Thus, the potential benefits of dexamethasone in women with class 1 HELLP syndrome remain unanswered.

3. Should dexamethasone be administered to patients with antepartum HELLP syndrome to make them suitable candidates for regional anesthesia? Regional anesthesia has been shown to avoid complications of exacerbated hypertension, aspiration, and failed intubation attributable to general anesthesia in this population. A retrospective study by O’Brien et al\textsuperscript{13} found that administration of dexamethasone increased the rate of regional anesthesia use in women with antepartum HELLP syndrome, particularly in those who achieved a latency of 24 hours before delivery.

Therefore, there is definite need for placebo-controlled trials with adequate sample size to answer these questions. Until then, the use of high-dose dexamethasone to improve maternal outcome in women with HELLP syndrome beyond 34 weeks’ gestation and/or in the postpartum period remains experimental.

References

10. Magann EF, Perry KO Jr, Meydrec EF, Harris RL, Chauhan SP, Martin JN Jr. Postpartum corticosteroids: accelerated recovery from the syndrome of hemolysis, elevated liver enzymes,


Dexamethasone treatment does not improve the outcome of women with HELLP syndrome: A double-blind, placebo-controlled, randomized clinical trial

Javier E. Fonseca, MD, MSc, Fabian Méndez, MD, PhD, Claudia Cataño, MD, Fernando Arias, MD, PhD

School of Public Health and School of Medicine, Universidad del Valle; Department of Gynecology and Obstetrics, Hospital Universitario del Valle, Cali, Colombia; Department of Obstetrics and Gynecology, The Toledo Hospital, Medical College of Ohio, Toledo, OH

Received for publication March 15, 2005; revised April 4, 2005; accepted July 5, 2005

Objective: The purpose of this study was to determine the efficacy of dexamethasone for treatment of HELLP (hemolysis, elevated liver enzymes and low platelet count) syndrome.

Study design: A prospective, double-blind clinical trial was conducted among 132 women with HELLP syndrome who were assigned randomly to treatment or placebo groups. Pregnant women in the experimental group received 10-mg doses of dexamethasone intravenously every 12 hours until delivery and 3 additional doses after delivery. Puerperal women received 3 10-mg doses of dexamethasone after delivery. The same schedule was used in the placebo group. The main outcome variable was the duration of hospitalization. In addition, we evaluated treatment effects on the time to recovery of laboratory and clinical parameters and on frequency of complications.

Results: The mean duration of hospitalization of patients who received dexamethasone therapy was shorter than in the placebo group (6.5 vs 8.2 days), but this difference was not statistically significant (P = .37). No significant differences were found in the time to recovery of platelet counts (hazard ratio, 1.2; 95% CI, 0.8-1.8), lactate dehydrogenase (hazard ratio, 0.9; 95% CI, 0.5-1.5), aspartate aminotransferase (hazard ratio, 0.6; 95% CI, 0.4-1.1) and to the development of complications. The results were found in both pregnant and puerperal women.

Conclusion: The results of this investigation do not support the use of dexamethasone for treatment of HELLP syndrome.

2005 Mosby, Inc. All rights reserved.

Supported in part by the Valle State Secretariat of Health; the dexamethasone and placebo were provided by Organon Laboratories, The Netherlands.

Reprints not available from the authors. Address correspondence to Javier E. Fonseca, Universidad del Valle, Obstetrics and Gynecology, Hospital Universitario del Valle, Cali, Valle 57, Colombia.

E-mail: jfonseca@univalle.edu.co

Multisystemic abnormalities that are associated with adverse maternal and fetal outcomes have been recognized for many years in women with preeclampsia; HELLP (hemolysis, elevated liver enzymes and low platelet count) syndrome is one of its most dangerous presentations. Women with HELLP syndrome may be classified by the degree of thrombocytopenia into...
Material and methods

This was a double-blind, placebo-controlled randomized clinical trial that involved pregnant and puerperal women who were admitted to the Hospital Universitario del Valle in Cali, Colombia, between October 2001 and September 2003. Women who were at >20 weeks of gestation or during the first 3 days of puerperium were asked to participate in the study if hypertension developed during the pregnancy or the puerperium and met the criteria for complete HELLP syndrome as defined by Sibai\textsuperscript{12}: platelet count, \( \leq 100,000/mm^3 \); aspartate aminotransferase (AST), >70 U/L; lactate dehydrogenase (LDH), >600 U/L. Exclusion criteria were oral temperature >37.5°C and diabetic ketoacidosis. Because of the potential for spontaneous recovery, puerperal women were excluded if randomization was not accomplished in the first 24 hours after diagnosis. The study was approved by the Institutional Review Boards of the Hospital and the Medical School.

Pregnant and puerperal women were assigned randomly to treatment or placebo groups, with the use of stratified and random permuted blocks of 4. The assignment was kept inside consecutively numbered opaque envelopes that were labeled as pregnant or puerperal and that were opened after informed consent had been obtained. Pregnant women in the experimental group received 10-mg doses of dexamethasone sodium phosphate (Oradexon) intravenously every 12 hours until delivery and 3 additional doses after delivery. Puerperal women received 3 10-mg doses after delivery. The same schedule was used in the control group to administer sterile water as placebo. Dexamethasone and placebo were packed in identical vials in sealed boxes that were labeled with the corresponding treatment codes. Codes were not broken until the end of the univariate analysis. Treatment was to be discontinued if oral temperature rose above 37.5°C.

All patients received 1 to 1.5 g/hr of magnesium sulfate intravenously. Nifedipine, 10 mg orally every 6 hours, was administered to women with diastolic arterial pressure >100 mm Hg. Another antihypertensive medication (clonidine and amlodipine) was administered if the diastolic pressure remained elevated. In addition, patients received 1000 mL of normal saline solution during the first 2 hours and 1000 mL of normal saline solution every 6 hours afterwards. If the urinary output remained <30 mL/hr, an additional 500 mL of normal saline solution was given during 1 hour; if oliguria persisted, 20 mg of furosemide was administered intravenously. Renal failure was diagnosed if the serum creatinine level was >1.5 mg/dL and if pulmonary edema was diagnosed by physical examination and chest radiography. When surgery was indicated, 8 units of platelets were transfused to women with platelet counts <50,000/mm\(^3\). Because the standard of care of the community is interruption of pregnancy after diagnosis of HELLP, induction of labor or cesarean delivery were performed, depending on the maternal and fetal condition. If the gestational age was between 26 to 34 weeks, betamethasone (12 mg intramuscularly) was given every 24 hours for up to 2 doses before delivery. Withholding steroids was considered unacceptable by the investigators and the Institutional Review Board.

Duration of hospitalization was measured from randomization to discharge. The duration of hospitalization of the 4 maternal deaths was excluded for the calculation of mean and median but was included and treated as censored data in survival analysis. Criteria for discharge included a platelet count >100,000/mm\(^3\), regardless of AST or LDH levels. However, if evidence of organ damage or other clinical complications was present, the patients remained in the hospital until recovery.

Measurements of blood pressure and urine output were carried out every 2 hours. Baseline and follow-up laboratory studies included platelet count, AST, LDH, and serum creatinine measurements every 12 hours. Platelet counts were performed by automated counting; LDH and AST were processed at 25°C, with a reference pattern of 120 to 240 U/L and 0 to 18 U/L, respectively.

Medical personnel were trained on protocol procedures for enrollment and follow-up of patients. Quality checks of clinical and laboratory forms were carried out before data entry. After entry, programs were run to verify the consistency of responses within each questionnaire. Any detected inconsistencies were resolved by correction against original data sheets.

Sample size was calculated by the duration of hospitalization, as defined earlier. We assumed an average hospital stay of 6 days in the control group and considered as significant a 33% decrease in stay for the experimental group, with a significance level of .01 and a power of 90%. The required sample size was
67 subjects per group. Analysis was carried out on the basis of intention to treat. A planned subgroup analysis was performed according to pregnant and puerperal strata. An additional, unplanned subgroup analysis was conducted by severity of clinical presentation at diagnosis. We carried out an interim analysis, at a sample size of 50 individuals, with no differences with final results. Continuous variables were analyzed with unpaired \( t \)-test or Mann-Whitney test, according to their distribution. If needed, transformations were carried out to allow for normal based statistics. Time to recovery of laboratory parameters (platelets, LDH, and AST) and duration of hospitalization was analyzed by Kaplan-Meier. Categoric variables were compared by chi-squared test or Fisher’s exact test. Multivariate analysis was performed with linear, logistic, or Cox regression, correspondingly. In addition to treatment (ie, the exposure of interest), other variables were considered in the final model if their probability values were < .2 in the univariate analysis.\(^{14-16}\) Antepartum use of betamethasone up to 2 weeks before randomization was also considered during modeling. Where appropriate, results are presented as relative risk with 95% CI, odds ratio, and hazard ratio.

**Results**

A total of 144 patients were considered eligible and were invited to participate in the study. Two patients (1.4%) were excluded because of fever, and 2 patients (1.4%)
declined to participate. Eight puerperal women (5.5%) were not allocated to treatment during the first 24 hours after diagnosis and were also excluded. After these exclusions, 132 women were eligible for randomization: 60 women were still pregnant, and 72 women were in the puerperal state. Two patients received only 2 doses of steroid after delivery because of death. Two women (in the experimental and control groups) received 1 dose of dexamethasone that was not provided by the study (Figure 1).

The mean age of enrolled women was 25.3 years (range, 14-44 years); the mean gestational age was 33.6 weeks (range, 20-41 weeks), and the mean parity was 2.4 pregnancies (range, 0-12 pregnancies). Platelet counts ranged from 13,000 to 100,000/mm$^3$, with mean and median values of 57,798/mm$^3$ and 59,200/mm$^3$, respectively. The mean LDH was 2183 U/L, with a median of 1701 (range, 623-12,600); the mean AST was 532 U/L, with a median of 280 (range, 71-2870). Baseline characteristics according to study groups were similar (Table I).

### Duration of hospitalization

The distribution of duration of hospitalization was transformed (1/duration of hospitalization) to allow the use of statistical methods that are based on the Gaussian distribution. The mean duration of hospitalization was shorter among patients who received dexamethasone therapy; however, this difference was not statistically significant. Median and interquartile ranges also were found to be no different (Table II). The univariate and multivariate analysis showed that a longer duration of hospitalization was associated with a lower urinary output ($<30$ mL/h) and higher LDH levels. Survival analysis of time to discharge showed no differences (Figure 2, A; hazard ratio, 1.3; 95% CI, 0.9-1.9).

### Time to recovery of laboratory tests

There was no statistically significant difference between dexamethasone-treated and placebo-treated patients with respect to the time that was required to achieve a platelet count of $>100,000/mm^3$ (Figure 2, B; Table III). Recovery of the platelet count was more likely to occur among patients with urinary output at enrollment of $<30$ mL/h and less likely among those with renal failure, although these findings were not significant at the multivariate analysis. Platelet counts did not reach levels of $>100,000/mm^3$ in all 4 maternal deaths.

LDH levels of $<600$ U/L were not reached before discharge by 72 patients (38 experimental patients and 34 control patients), which included the 4 deaths and 68 patients who were discharged when their platelet counts were $>100,000/mm^3$. There was no statistically significant difference between treated and control patients who reached an LDH of $<600$ U/L before discharge regarding their time to recovery (Figure 2, C; Table III) or other patient characteristics at enrollment.

### Table I: Baseline characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Placebo (n = 66)</th>
<th>Dexamethasone (n = 66)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)*</td>
<td>26.2 ± 7.20</td>
<td>24.5 ± 7.00</td>
</tr>
<tr>
<td>Gestational age (wk)*</td>
<td>33.5 ± 4.20</td>
<td>33.8 ± 4.50</td>
</tr>
<tr>
<td>Parity (n)*</td>
<td>2.0 ± 2.06</td>
<td>1.2 ± 1.25</td>
</tr>
<tr>
<td>Platelets (n/mm$^3$)*</td>
<td>58,446 ± 21,053</td>
<td>61,171 ± 18,912</td>
</tr>
<tr>
<td>AST (U/L)*</td>
<td>492 ± 579</td>
<td>573 ± 621</td>
</tr>
<tr>
<td>Alanine aminotransferase (U/L)*</td>
<td>229 ± 251</td>
<td>281 ± 300</td>
</tr>
<tr>
<td>LDH (U/L)*</td>
<td>2,242 ± 1,671</td>
<td>2,124 ± 1,849</td>
</tr>
<tr>
<td>Total bilirubin (mg/mL)*</td>
<td>3.7 ± 4.67</td>
<td>3.3 ± 3.20</td>
</tr>
<tr>
<td>Creatinine (mg/mL)*</td>
<td>0.9 ± 0.53</td>
<td>0.9 ± 0.49</td>
</tr>
<tr>
<td>Urinary output (mL/min)*</td>
<td>85 ± 65.50</td>
<td>92 ± 59.70</td>
</tr>
<tr>
<td>Systolic pressure (mm Hg)*</td>
<td>141 ± 21.01</td>
<td>145 ± 21.00</td>
</tr>
<tr>
<td>Diastolic pressure (mm Hg)*</td>
<td>93 ± 13.40</td>
<td>93 ± 12.61</td>
</tr>
</tbody>
</table>

### Table II: Duration of hospitalization according to treatment

<table>
<thead>
<tr>
<th>Duration of hospitalization (days)</th>
<th>Placebo (66)</th>
<th>Dexamethasone (66)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (D.S.)</td>
<td>8.2 (12.55)</td>
<td>6.5 (9.66)</td>
</tr>
<tr>
<td>Median</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Range</td>
<td>2-89</td>
<td>2-64</td>
</tr>
<tr>
<td>Interquartile range</td>
<td>3-8</td>
<td>3-6</td>
</tr>
</tbody>
</table>

Difference of mean: 1.7 ($-2.28561$). $P = .37$
AST levels of <70 U/L were not reached by 53 patients (32 experimental patients and 21 control patients), which included 2 of the maternal deaths and 51 patients who were discharged once their platelet counts reached >100,000/mm³. Among patients with an AST level of <70 U/L before discharge, there was a trend to faster recovery among those patients who received placebo (log rank test probability value, .07) that was not significant after adjustment for renal failure, ethnicity, and parity (hazard ratio, 0.7; 95% CI, 0.4-1.1; Figure 2, D; Table III).

Recovery of clinical parameters

No significant differences in urinary output were found between the 2 treatment groups. Furosemide was required in 23 patients: 13 patients received placebo, and 10 patients received dexamethasone therapy (relative risk, 0.8; 95% CI, 0.4-1.6). All patients were hypertensive, and 124 patients (93.9%) required nifedipine therapy; therefore, we were not able to evaluate changes in blood pressure, because they could have been associated with antihypertensive use. The addition of a second antihypertensive drug was necessary in 30 patients, 13 of whom received dexamethasone therapy (P = .32). A third antihypertensive drug was required in 8 patients, 4 per treatment group.

Complications and blood transfusion

There were 4 maternal deaths, 3 deaths in the dexamethasone group and 1 death in the placebo group. Three of the maternal deaths occurred in women with liver failure and severe hemolysis, with AST and LDH levels of >2600 U/L and >6450 U/L, respectively. The other death was due to a cerebrovascular accident. The treatment groups were not different regarding development of complications or transfusion need (Table IV). Interestingly, there were a higher number of infections among those patients who received placebo, and maternal death and pulmonary edema were more frequent among those women who received dexamethasone therapy, even after adjustment. Infections were found to be associated independently with admission to the intensive care unit (odds ratio, 10.4; 95% CI, 2.2-48.2), renal failure (odds ratio, 8.8; 95% CI, 1.9-38.8), and vaginal delivery (odds ratio, 6.7; 95% CI, 1.3-33.3). The higher odds of infection with vaginal delivery remained after adjustment for the occurrence of premature rupture of membranes and the use of antibiotics and steroids.
before randomization. Platelet transfusion was found to be associated with the development of renal failure (odds ratio, 4.0; 95% CI, 1.1-14.3) and AST levels at enrollment.

### Subgroup analysis by pregnant and puerperal patients

Stratified analysis of pregnant and puerperal groups showed no differences in the occurrence of complications, recovery of laboratory parameters, transfusion need, or duration of hospitalization. Among puerperal women, the mean duration of hospitalization tended to be lower in those women who received placebo than in those women who received dexamethasone therapy (6.8 vs 8.2 days), but this difference was not significant. Median duration was 4 days in both groups; the interquartile ranges were 3 to 9 days and 3 to 6 days, respectively. Among pregnant women, the duration of hospitalization was lower in the women who received dexamethasone therapy (4.5 vs 9.9 days), but this difference did not reach statistical significance.

### Table III

Determinants of duration of hospitalization, platelet count, LDH and AST recovery by univariate Cox regression

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Duration of hospitalization</th>
<th>Platelets*</th>
<th>LDH†</th>
<th>AST‡</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hazard ratio</td>
<td>95% CI</td>
<td>Hazard ratio</td>
<td>95% CI</td>
</tr>
<tr>
<td>Placebo</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Dexamethasone</td>
<td>1.3</td>
<td>0.87-1.94</td>
<td>1.2</td>
<td>0.80-1.77</td>
</tr>
<tr>
<td><strong>Steroid use up to 2 weeks before delivery</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No†</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Yes‡</td>
<td>0.9</td>
<td>0.60-1.46</td>
<td>0.9</td>
<td>0.59-1.43</td>
</tr>
<tr>
<td><strong>Delivery</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaginal§</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Cesarean§</td>
<td>0.8</td>
<td>0.51-1.27</td>
<td>0.7</td>
<td>0.47-1.09</td>
</tr>
<tr>
<td><strong>Class HELLP at admission</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HELLP 1§</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>HELLP 2</td>
<td>0.8</td>
<td>0.54-1.28</td>
<td>1.2</td>
<td>0.79-1.81</td>
</tr>
<tr>
<td><strong>Acute renal failure at admission</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No†</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Yes‡</td>
<td>0.3</td>
<td>0.15-0.44</td>
<td>0.5</td>
<td>0.22-0.99</td>
</tr>
<tr>
<td><strong>Urinary output at admission</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;30 cc/hour‡</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>31-100 cc/hour□</td>
<td>4.7</td>
<td>1.94-11.52</td>
<td>2.4</td>
<td>1.07-5.17</td>
</tr>
<tr>
<td>&gt;100 cc/hour□</td>
<td>5.8</td>
<td>2.13-15.80</td>
<td>2.6</td>
<td>1.07-6.22</td>
</tr>
</tbody>
</table>

* Platelets above 100,000/mm³.
† LDH below 600 U/L.
‡ AST below 70 U/L.
§ Non-adjusted Hazard ratio.
¶ Reference category.

### Table IV

Complications associated with HELLP syndrome according to steroids use

<table>
<thead>
<tr>
<th>Complication</th>
<th>Placebo n (%)</th>
<th>Dexamethasone n (%)</th>
<th>R.R. crude (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute renal failure*</td>
<td>8 (12.9)</td>
<td>6 (10.0)</td>
<td>0.8 (0.29-2.10)</td>
</tr>
<tr>
<td>Oliguria</td>
<td>4 (6.06)</td>
<td>5 (7.58)</td>
<td>1.3 (0.35-4.45)</td>
</tr>
<tr>
<td>Pulmonary edema*</td>
<td>1 (1.54)</td>
<td>3 (4.62)</td>
<td>3.1 (0.32-28.09)</td>
</tr>
<tr>
<td>Eclampsia</td>
<td>10 (15.15)</td>
<td>8 (13.79)</td>
<td>0.8 (0.34-1.90)</td>
</tr>
<tr>
<td>Infections</td>
<td>10 (15.15)</td>
<td>5 (7.58)</td>
<td>0.5 (0.18-1.38)</td>
</tr>
<tr>
<td>Dead</td>
<td>1 (1.52)</td>
<td>3 (4.62)</td>
<td>3.0 (0.32-28.1)</td>
</tr>
<tr>
<td>Platelets transfusión</td>
<td>10 (15.15)</td>
<td>12 (18.18)</td>
<td>1.2 (0.56-2.58)</td>
</tr>
<tr>
<td>Plasma transfusión</td>
<td>6 (9.09)</td>
<td>5 (7.58)</td>
<td>0.8 (0.27-2.60)</td>
</tr>
</tbody>
</table>

* Only included patients without the event before randomization.

### Figure 3

Interval between randomization and recovery of the platelet count to >100,000/mm³ by study group among patients with HELLP 1.
median duration was 4 days in both groups, and the interquartile ranges were 3 to 4.5 days and 3 to 7 days for dexamethasone therapy and placebo, respectively.

Subgroup analysis by severity

The time to recovery of the platelet count was found to be heterogeneous when the cases were stratified by HELLP class at the time of enrollment (Mantel-Cox test: chi-squared test, 4.76; \( P = .03 \)). Therefore, we performed a subgroup analysis according to severity at enrollment. No differences were found among patients who were classified as HELLP class 2 and control subjects; however, among 49 patients with HELLP 1 (28 patients with placebo and 21 patients with dexamethasone therapy), the conditional probability of platelet recovery was higher in those patients who received dexamethasone therapy (Figure 3), even after adjustment for potential confounders (hazard ratio, 3.4; 95% CI, 1.3-8.5). Also, the duration of hospitalization was shorter among women who received dexamethasone therapy when means (4.6 and 10.4 days), medians (3.5 and 5.0 days), and interquartile ranges were compared. This trend persisted after adjustment (\( P = .03 \)).

Comment

Several randomized clinical trials have been published to evaluate the effect of dexamethasone therapy in women with HELLP syndrome.9-11 Although they indicate that dexamethasone therapy is beneficial, the strength of this conclusion is limited because of small number of patients in each trial, the lack of blinding and placebo controls, the inclusion of women with mild forms of the disease, and the lack of an strict definition of the syndrome (Table V). Observational studies have also found better outcomes in patients with HELLP syndrome who received dexamethasone therapy. However, 2 of the studies were retrospective, with historic control groups,17,18 and the other 2 studies compared different steroids.19,20 To the best of our knowledge, this is the largest reported clinical trial to evaluate the use of dexamethasone therapy in HELLP syndrome and the first that is double blind and placebo controlled. This is also the first study among patients with HELLP syndrome that report sample size estimation. The estimate was based on the duration of hospitalization because it has been widely accepted that this outcome variable reflects the development of complications and the rate of recovery of clinical and laboratory variables and because it is a useful indicator for patients and clinicians. Other valuable features in this study that support its internal validity are the study design (stratified randomization in blocks and double blind) and the small number of protocol violations that result in compliance with the assigned treatment of >95%. The external validity of this study is also high probably because of the large number of eligible patients who accepted randomization, the adoption of a widely accepted dose of dexamethasone for the treatment group, the use of betamethasone in preterm deliveries, and the clinical relevance of the outcome measures.

A weakness of this study is that 28.03% of our patients received betamethasone during the 2 weeks before delivery for the purpose of accelerating fetal lung maturity and preventing neonatal intracranial bleeding. However, analyses were carried out to adjust by previous steroid administration (Table III). Furthermore, analyses were carried out that included only women who did not received other steroids before delivery (45 women in the placebo group and 50 women in the dexamethasone group; power, \( >90\% \); \( \alpha \), .05) with similar results; therefore, previous administration of betamethasone did not affect final results.

The results of this study indicate that the administration of dexamethasone in patients with complete class 1 and 2 HELLP syndrome, when compared with similar patients who received placebo, does not reduce the number of complications or the need for blood products administration or shorten platelets and LDH recovery. These results were consistent in global and planned subgroup analysis. Contrasted with previous studies, AST recovery was slower in patients who were assigned to dexamethasone therapy than to placebo, and this finding remained in the subgroup analysis. Unfortunately, we cannot speculate about the clinical implications of the last finding, given that patient follow-up evaluations finished at the time of platelets recovery, regardless of AST levels.

<table>
<thead>
<tr>
<th>Table V</th>
<th>Randomized clinical trials of corticosteroids in HELLP syndrome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author</td>
<td>Year</td>
</tr>
<tr>
<td>Magann</td>
<td>1994</td>
</tr>
<tr>
<td>Magann</td>
<td>1994</td>
</tr>
<tr>
<td>Vigil-De Gracia</td>
<td>1997</td>
</tr>
<tr>
<td>Yalcin</td>
<td>1998</td>
</tr>
<tr>
<td>Isler</td>
<td>2003</td>
</tr>
<tr>
<td>Present work</td>
<td>2005</td>
</tr>
</tbody>
</table>
Although no statistically significant differences were found in the planned analysis of the duration of hospitalization, those pregnant women who received placebo stayed, on average, twice as long as those women who received dexamethasone therapy (9.9 vs 4.5 days). This difference was due to 2 patients of the control group who stayed long periods (49 and 89 days), one of whom had rupture of the uterus and subsequent complications, and the other who refused blood transfusion, regardless of clinical indication (hemoglobin levels of 6.2 and 2.8 g/dL before and after cesarean delivery), and experienced respiratory distress syndrome and renal failure that prolonged the hospitalization. These 2 outliers affected the mean duration of hospitalization, but comparisons of median and interquartile range did not show differences between groups.

Subgroup analysis according to the severity of disease showed that, among patients with HELLP 1, there were a shorter average time to platelet recovery and less duration of hospitalization in women who received dexamethasone therapy. The importance of this finding is diminished because this was an unplanned analysis and the severity of the disease was not taken into account at randomization. It is accepted that the results of an unplanned subgroup analysis should be considered exploratory.21

In summary, the results of this investigation do not support the use of dexamethasone for treatment of HELLP syndrome. In women with class 1 HELLP, further studies would be required to evaluate the potential benefit of this intervention.

Acknowledgments

We thank the residents in Obstetrics and Gynecology at the University Hospital, Universidad del Valle, Cali, Colombia, for their help in this project.

References

We analyzed The Women’s Health Initiative (WHI) Study because it had a significant impact on clinical practice, both nationally and internationally. However, despite the widespread public and professional awareness of the results, an independent, nonbiased analysis of the quality of the methodology of the study has not been available. We find the study design and its execution question the validity of the results, making it difficult to apply the WHI results to healthy postmenopausal women, different ethnic groups, or as general postmenopausal prevention.

Initial publications of the Women’s Health Initiative (WHI) study results have had profound influence on clinical management of postmenopausal women. The WHI clinical trial is considered by many as a landmark study of postmenopausal hormone therapy. Published findings from this study have concentrated on interpretation and applicability of the results. Such interpretations of the results led to several recommendations for practitioners and women and were published by, but not limited to, the American College of Obstetrics and Gynecology. In order to accept results of any study, an objective scientific scrutiny of methodology and execution of the study design is essential. Our objective was to evaluate the WHI study’s methodology and its execution using Sackett’s procedures for identifying evidence-based medicine in reviewing the WHI published results.

A review of the existing world body of the literature through PubMED, ACOGNET, ProQUEST, OVID, Cochrane Collection, The Lancet on Line Collection, MDConsult, New England Journal of Medicine, American College of Physicians on Line Resources, Highwire Journal, and Citation Index Reference computerized databases was conducted on the following subjects: menopause, hormone replacement therapy, Prempro, estrogen replacement therapy, and Women’s Health Initiative Study. Also, a manual search and analysis of the literature on these subjects were performed. Both computerized and manual searches failed to identify a publication related to the WHI study scientific scrutiny of the methodology and its execution.
Study design

A review and analysis of the methodology and its execution were conducted based upon the original publications of the WHI study.

Results

WHI Study design

The HRT clinical primary prevention trial selected a total of 27,500 women for double-blind comparison in each of the 3 initial arms of the clinical trials. These women were randomly placed into the study groups. Sixteen thousand six hundred and eight postmenopausal women were placed, with a ratio of 1:1 for the Prempro (Wyeth Pharmaceuticals, Philadelphia, PA) group and the placebo product. The interventional group (n = 8506) took Prempro 1 tablet daily continuously (1 tablet of Prempro contains conjugated equine estrogens, 0.625 mg, and medroxyprogesterone acetate 2.5 mg). Both groups were on diet of low-fat dietary pattern (total dietary fat \( \leq 20\% \) and saturated fat \( \leq 7\% \)), with a calcium carbonate supplement tablet (1000 mg daily) and a dietary calcium intake of 500 mg daily for a total of 1500 mg/day, and a vitamin D\(_3\) supplementation of 400 IU daily.\(^3\)

Arms of the WHI Study

The following arms of the study were created: 1) unopposed estrogen and women who had a hysterectomy (study halted in 2004); 2) estrogen/progestin (Prempro) and the intact uterus (study halted in 2003); 3) unopposed estrogen (Premarin 0.625 mg [Wyeth Pharmaceuticals, Philadelphia, PA]) and the intact uterus (study halted in 1996). Three hundred and thirty-one subjects participated in this group and the participants were transferred to the Prempro group without additional randomizing.

The power of the clinical trial

In general, the power of the WHI was adequate in both groups for Caucasian women: the interventional group 8506 and placebo group 8102. In the interventional group, 7140 (83.94%) were Caucasian women, 549 (6.45%), Black women, 472 (5.54%) Hispanic women, 26 (0.30%) American Indian women, 194 (2.28%) Asian/Pacific Islander women, and 125 (1.46%) of unknown ethnicity.

The age of the subjects

The intervention group

The mean age of the subjects in the study was 63.2 years with 2-standard deviation \( \pm 7.1 \) years (range from 56.1 and 70.3). The women’s age group between 60 and 79 years as n = 5.667 (69.94%).\(^2\)

The placebo group

The mean age was 63.3 \( \pm 7.1 \) (56.2–70.4). The women’s age group between 60 and 79 years was 5.667 (69.94%).\(^2\)

Yearly dropout

The intervention group

The yearly dropout was 39.26%.

The placebo group

The yearly dropout was 47.29%.

Hormone replacement therapy used by the subjects before entering WHI Study

The interventional group

Current users at the screening were n = 548 (6.44%) and past users were 1674 (19.68%). Total number of subjects exposed to HRT was 2222 (26.12%) at the time of screening.

The placebo group

Current users were n = 487 (6.0%), past users were n = 1588 (19.6%), and total number of subjects exposed to HRT was n = 2075 (25.75%) at the time of screening.

Medical conditions at the screening

The interventional group

Body mass index (BMI) 25 to 29 was identified in 2992 subjects (35.17%); BMI \( \geq 30 \) was in 2899 (34.08%). The total number of subjects overweight and obese was 5891 women (69.25%).

Elevated systolic blood pressure was present in 127 subjects (1.5%); hypertension in 3039 (35.72%); diabetes in 944 (11.09%), subjects on a statin medication 590 (6.93%), total number of subjects being treated for hypercholesterolemia was 1543 women (18.03%). The number of subjects on daily aspirin was 1623 (19.08%). Total number of sick subjects at the screening of the WHI study was 6570 (77.23%).

The placebo group

Body mass index 25 to 29 was present in 2479 subjects (30.59%); BMI \( \geq 30 \) was in 2737 subjects (33.78%). The total number of subjects overweight and obese was 5487 women (69.25%).

Elevated systolic blood pressure was present in 127 subjects (1.57%); hypertension in 2949 women, (36.39%) diabetes in 962 subjects (11.87%), subjects on a statin medication 548 (6.76%), total number of subjects being treated for hypercholesterolemia was 1510 women (18.63%). The number of subjects on daily aspirin was 1631 (20.02%). Total number of sick women at the screening of the WHI study was 6450 (79.60%) subjects.

Medical history of illnesses

The intervention group

One hundred and thirty-nine (1.63%) subjects had a medical history of myocardial infarct; 238 (2.79%)
subjects had a history of angina; 95 (1.11%) subjects had coronary artery graft or percutaneous transluminal angioplasty. The total number of subjects included in the intervention group with a history of heart illnesses was 472 (5.54%). Medical history of deep vein thrombosis (DVT) or pulmonary embolism (PE) was present in 79 subjects (0.92%); 61 (0.71%) women had a history of stroke. The total number of subjects with history of illness was 612 (7.19%).

**The placebo group**

One hundred and fifty-seven (1.93%) subjects had a medical history of myocardial infarct; 234 (2.88%) had a history of angina; 120 (1.48%) subjects had coronary artery graft, percutaneous transluminal angioplasty, or stroke. The total number of women included in the placebo group with a history of heart illnesses was 511 (0.63%). Medical history of DVT or PE was present in 62 subjects (0.76%); 77 (0.95%) women had a history of stroke. The total number of subjects with history of illness was 650 (8.02%) subjects.

**Family breast cancer at the screening**

**The interventional group**
The number of female relatives diagnosed with breast cancer was 1286 (15.11%).

**The placebo group**
The number of female relatives diagnosed with breast cancer was 1175 (14.5%).

**Comment**

**Evidence-based methodology**

Before study results can be interpreted objectively, a scientific scrutiny of the methodology is imperative. A basic concept of any clinical trial should be that the method of design and execution of the study will hold up to accepted standards of evidence-based medicine. The qualities of implementation of methodology before and during the study will invariably determine the overall value of the study findings. The present and long-term implications of the WHI clinical trial should require the establishment of the reliability of the study results and their applicability to healthy individuals and for the use of Prempro for prevention. The original title of the WHI study publications were “Risk and benefits of estrogen plus progestin in healthy postmenopausal women…”1 and “Failure of estrogen and progestin therapy for prevention.”2 The words healthy and prevention in the titles implied that the WHI study would evaluate the effect of estrogen/progestin on healthy postmenopausal women, and appraise the use of Prempro for postmenopausal women health prevention.1,2 The JAMA report of the WHI study results1 left the impression that these clinical trial results can be applied to all healthy postmenopausal women as a preventive approach regardless of age and race. Scientific scrutiny of the WHI study’s methodology fails to support this suggestion. Based on the WHI study group, implementation of the results into clinical practice has little, if any, scientific basis. The outline of the WHI study methodologic weakness is presented below.

**WHI Study design**

**Low-fat dietary pattern and dietary supplements**

“Women are screened for participation in one or both of the components—dietary modification (DM) or hormone replacement therapy (HRT)...”3 Such compound regimens being incorporated into the study design created at least 3 scientific deficiencies: 1) heterogeneous medications tested (Prempro, low-fat diet and calcium carbonate, and vitamin D3 supplements); 2) inadequate outcome measure related only to Prempro; and 3) interpretation of the results couldn’t be specifically assigned to either Prempro or low-fat diet pattern or calcium carbonate and vitamin D3. Each remedy used in the study has an individual clinical effect on the outcome measure; therefore, the validity of the outcome measures and the degree of effect each medication had on the outcome measure could not be determined.

This lack of the study identifying which remedy had a clinical effect on outcome measures leaves room for speculation of the validity of the results. The WHI study primary outcome measure was coronary heart disease (CHD) associated with Prempro, low-fat dietary pattern treatment, and calcium carbonate and vitamin D3 supplements. This combination did not distinguish between HRT and the low-fat diet pattern, and to what degree each remedy influenced outcome of CHD and stroke. Also, nutrition can be a fundamental factor in breast or colon cancers development,3 and a low-fat diet pattern unquestionably will influence the outcome measures of breast and colon cancer in this clinical trial. The WHI study conclusion that Prempro decreased the colon cancer prevalence was inaccurate because of the presence of the other documented factors that might influence in the study outcome. The reporting of the breast cancer prevalence influences cannot be objectively interpreted for the same reasons.

Calcium carbonate and vitamin D3 supplementation have been documented as affecting the outcome of bone fracture prevalence, including hip fracture.3 The WHI study, by having included the use of these dietary supplements in both groups, invalidated the position that Prempro alone had a significant beneficial effect on the incident of bone fracture. Using several remedies in the studies (estrogen/progestin, calcium carbonate, vitamin D3, low-fat diet) significantly reduced the validity
that Prempro alone was the beneficial cause of bone fracture reduction.

The power of the study
The original study plan was to incorporate 20% of overall minority ethnic group enrollment; however, only 16.03% of the enrollment target was achieved. In addition, there were significant differences in the number of subjects representing minority groups. The single ethnic group in the WHI study did not meet the minimum quota of 348 subjects necessary to qualify for “grade A” evidence-based medicine analysis, when the yearly dropout of 39.26% in the interventional group was subtracted. In general, the power of the WHI study was not sufficient to represent ethnic groups as a whole, or any specific ethnic group. Such misrepresentation among ethnic ratio distribution makes the applicability of the results of the WHI study to minority groups impractical. The power of the WHI study was adequate for Caucasian women.

To satisfy the power of the WHI study for any given ethnic group required 348 subjects. Black postmenopausal women were the highest group by number among the minority group subjects (549, equal to 6.45% of the total study’s population). The number of subjects that dropped out indicated that the power of the study was not satisfied. The WHI study publication did not disclose numerical information for dropout of each specific ethnic group. Applying the general 39.26% dropout ratio to the black population, at least 214 black women dropped out from the study. The remaining 335 black women were submitted to global study analysis instead of creating a separate black woman subgroup for statistical analysis. As a result, this group neither satisfied a power study for black postmenopausal women, nor represented a separate subgroup within the overall WHI study results analysis. The poor accounting of ethnic ratio distribution is not sufficient to support the conclusion that the WHI study results can be applied to any ethnic or minority groups.

Validity of the WHI study’s results for any minority group is lacking for 3 reasons: 1) insufficient power of the study for minority groups; 2) the absence of a specific subgroup analysis; and 3) inadequate inclusion/exclusion criteria.

Reallocation of the subgroup results to a general population analysis of the study does not reflect a subgroup response to the tested multigents. Different ethnic groups will demonstrate different predispositions to medical conditions and how the use of Prempro and the low-fat diet pattern with supplements of calcium carbonate and vitamin D3 would affect women associated with medical conditions differently than healthy subjects. Inadequate exclusion/inclusion criteria, the absence of adequate subgroups, and insufficient power of the study for minority groups make WHI study results impossible to apply to any minority group.

The age of the subjects
The median age of menopause among women in the US is 51.1 years. Approximately 70% of the WHI participants were women between the age of 60 years and 79 years, with a mean age of 63.2 years ± 7.1 years. Consequently, women in the study below 56.1 years of age would fall out of a 2-standard deviation analysis and become a statistically insignificant group. Simply, the applicability of the WHI study’s findings to younger women, between age 51.1 and 56.1 years and younger, is unknown. The age gap itself between the median age of menopause and the median age of the WHI study participants has disqualified the WHI label of a primary prevention trial. The WHI study can honestly only be described as a clinical trial of older women, being that there is a 12-year difference in median age of menopausal women nationally and those who were in this trial.

Yearly dropout
In the intervention group and in the placebo group, yearly dropout at the end of the study was 39.26% and 47.29%, respectively. In the WHI study, the majority of women were free from vasomotor symptoms (VMS). This fact, in conjunction with undesirable side effects related to Prempro itself (breast tenderness, bloating, endometrial bleeding, or mood changes) can provide an explanation for 39.26% dropout from the intervention group. Because younger postmenopausal women experience more severe VMS, then, most likely such a younger group will adhere to the protocol closely, and this may have reduced the number of yearly dropouts. Although it is clearly stated in the WHI study’s reports that this trial was not designed to evaluate postmenopausal VMS, psychologic, or peripheral symptoms, the absence of the symptoms among subjects contributed to the higher rate of dropout from the study.

Arms of the WHI study
The unopposed estrogen group was selected among women with an intact uterus. This group was originally randomized between 1992 and 1996, and the study was halted in 1996. The justification for discontinuation of the study was given as “...long-term adherence to unopposed estrogen was not feasible in women with a uterus...” Consequently, this group was not rerandomized and arbitrarily was relocated from the Premarin group to the Prempro group. Such contamination of the study population disqualifies Prempro result interpretation because 2 agents were used by the same subject (estrogen alone and later estrogen/progestin regimen) at different times. The studies indicated that sufficient time lapsed (4 weeks) to eliminate the short-term effects of estrogen. This time period would not definitely eliminate the long-term effect of estrogen use. Each of the defined regimens will have its own long-term effect.
on the outcome measures in the WHI study. The clinical parameters outcome measure of the WHI study of 2 independent agents, estrogen alone between 1992 and 1996, and estrogen/progestin from 1996 to 2003, is unfeasible for results interpretation. Estrogen alone would have had an effect on clinically measurable outcome parameters, and estrogen/progestin has apparently different effect on measurable clinical parameters (CHD, DVT, stroke, breast cancer, hip fracture, colorectal cancer, etc).

Medical conditions and history of illnesses at the screening
In the interventional group, a total of 6645 (78.12%) women were sick or had a history of illnesses. These subjects were identified as having been treated for being overweight; obesity; elevated systolic, diastolic blood pressure; hypertension; diabetes; hypercholesterolemia; or use of daily aspirin (coronary heart disease or stroke prevention) at the initial screening of the WHI study. These medical conditions by themselves would predispose subjects to CHD, stroke, or DVT, or PE, without HRT. It is obvious that 78.12% of the subjects did not qualify as a healthy population. The 22% remaining women would have been the only valid participants to reflect the stated main outcome to appraise the effect of HRT on CHD.

The combination of both an older age group of participants and preexisting medical conditions invalidated the concept that the WHI studies were designed to evaluate the effect of estrogen/progestin on healthy women or an evaluation study for prevention. The high percentage (78.12%) of subjects with a preexisting illness included in the WHI study should have disqualified the results of this clinical trial application to a healthy population. Included in the clinical trial were subjects with previous medical conditions (myocardial infarct, angina, angioplasty, DVT or PE, or stroke). The total number of subjects with a history of illness was 612 (7.19%).

Also, 1286 (15.11%) female relatives had documented breast cancer. Numerous epidemiologic studies indicate that a family history of breast cancer, particularly the immediate relatives, has significant influence on the breast cancer incidence. At least 1286 (15.11%) of the subjects in the WHI study had a family predisposition for breast cancer before being exposed to Prempro. A special subgroup should have been created for evaluation of the effect of estrogen/progestin on the subjects with a family history of breast cancer, and this subgroup should not have been included in the general study population.

Evidence-based medicine has already established that the above-mentioned medical conditions or history of illnesses would constitute either indirect or direct contraindication in the use of HRT. Under no circumstance would this subject group meet the criteria for healthy candidates. All of these unhealthy populations in the WHI study should have been organized into separate subgroups according to the medical condition, and presented with separate and specifically designed informed consent and excluded from the general study population. The exclusion and inclusion criteria of the WHI study revealed serious scientific concerns; this absence of specific medical condition subgroups made the WHI study invalid for healthy postmenopausal women and unhealthy subjects as well because healthy and unhealthy subject were mixed in the results.

The WHI clinical trial can be considered as an enormous undertaking and effort; however, it should not be considered as a clinical landmark or revolutionary trial. The lack of a sound scientific design and execution invalidated the study from the primary selection and methodology used. The professional community and the public-at-large were presented with the results of the WHI study without scientific scrutiny of this clinical trial. Some professionals among medical authorities and the US media have concluded that a hormone replacement therapy is bad for women’s health based on the WHI study. After careful review of the WHI reports, adherence to the principles of evidence-based medicine, our analysis can neither confirm the notion that HRT is bad for postmenopausal women health, nor that HRT is good for postmenopausal woman’s health. Our analysis simply suggests that the experienced clinical scientists must evaluate the design and execution of the WHI study in order to establish effectiveness and benefit/risk of the HRT, if not completely redo the study itself.

Conclusions
Only 1 dose of Prempro has been tested in the WHI study and the russets can not be applied to the low dose of Prempro.

Using combination of conventional medicine (estrogen/progestin) and alternative remedies (calcium carbonate, vitamin D₃, low-fat diet) significantly obscured, if not completely eliminated, interpretation of the benefit/risk effect of Prempro on CHD, stroke, breast cancer, bone fracture.

Inadequate implementation of exclusion/inclusion criteria
Insufficient power of the study for different minority groups made WHI study results impossible to apply to any of the minority groups.

The applicability of the WHI findings to women between age of 51.1 and 56.1 years and younger is unknown. The age gap itself between the median age of menopause and the median age of the subjects in the study
disqualifies the WHI study’s label of a primary prevention trial.

The WHI study can only be described as a clinical trial of older women, being that there is a 12-year difference in the median age of menopause and the median age of those women who were participated in the WHI study.

The reduced or absence postmenopausal symptoms among older women contributed to the higher rate of dropout from the study.

The women from the Premarin group (the subjects with an intact uterus being on unopposed estrogen) were not rerandomized and arbitrarily relocated to the Prempro interventional group without creating a homogenous placebo group.

The WHI study’s outcome measure was contaminated by using 2 independent agents among the same subjects: estrogen alone (Premarin) between 1992 and 1996 and estrogen/progestin (Prempro) between 1996 and 2003. Such contamination made the WHI study’s results interpretation impractical.

The WHI clinical trial can be considered as an enormous undertaking and effort; however, it should not be considered as a clinical landmark or revolutionary trial.

Our analysis can neither confirm the notion that the HRT constitutes a health-hazard for postmenopausal women, nor that HRT is beneficial for postmenopausal women health.

The WHI clinical trial should be completely redesigned and redone in order to determine the HRT safety and effectiveness.

Acknowledgments

The authors are very appreciative to Joyce C. Burkhart, MS, and Robert W. Burkhart, MS, for assisting in the manuscript preparation.

References

WHI response to Ostrzenski and Ostrzenska

Barbara B. Cochrane, PhD, RN, a,* David H. Barad, MD, MS, b Margery Gass, MD, c Robert L. Brunner, PhD, d Cora E. Lewis, MD, MSPH, e Marcia L. Stefanick, PhD f

Public Health Sciences Division, Fred Hutchinson Cancer Research Center, a Seattle, WA; Department of Obstetrics and Gynecology and Women’s Health, Albert Einstein College of Medicine, b Bronx, NY; Department of Obstetrics and Gynecology, University of Cincinnati College of Medicine, c Cincinnati, OH; Department of Family and Community Medicine, University of Nevada School of Medicine, d Reno, NV; Division of Preventive Medicine, University of Alabama at Birmingham, e AL; Stanford Center for Research in Disease Prevention, Stanford University, f Palo Alto, CA

Received for publication June 14, 2005; accepted July 29, 2005

Contrary to Ostrzenski’s and Ostrzenska’s' premise, the Women’s Health Initiative (WHI) Estrogen plus Progestin (E+P) Trial has generated considerable scientific scrutiny, from a JAMA editorial describing the results as “strong evidence,” 3 to the American College of Obstetricians and Gynecologists characterizing the trial as “…the largest, most statistically valid, and well-analyzed research to evaluate the use of HRT in healthy postmenopausal women,”4 and the American Society for Reproductive Medicine writing “The data emerging from this methodologically sound randomized controlled trial appear incontrovertible.”5 Sackett, whose principles of evidence-based medicine Ostrzenski and Ostrzenska endorse, described the trial as a “rigorous” randomized clinical trial, designed and executed to answer the study question.7

We address below some of Ostrzenski’s and Ostrzenska’s claims about the WHI findings:

The majority of participants were “sick”

The presence of risk for chronic disease (eg, hypertension, obesity) does not mean that a person is “sick”; in fact, WHI participants at baseline were healthier than the general population of US women aged 50 to 79.8

This was a clinical trial of “older” women

Over 12% of participants were aged 50 to 54 at baseline (n = 2029), making it the largest randomized E+P trial ever conducted in this age group. Subgroup analyses did not indicate significant interactions with age for E+P effects on major outcomes.

Yearly dropout was high and compromised study power

The yearly dropout rates cited by Ostrzenski and Ostrzenska are not accurate. The dropout rates reported in the E+P trial (defined as stopping study pills at any time, whether or not prescription hormones were started or study pills were restarted) were cumulative (42% for E+P, 38% for placebo), 2 with first-year rates being similar to or lower than those reported in other hormone trials and community studies, and never exceeding 9.7% in any given year for either treatment arm (range
Despite these dropout rates, the study had adequate statistical power for testing the primary hypotheses.

**Dropout in the intervention group would have been lower if women had vasomotor symptoms**

In fact, many of these participants did have moderate or severe vasomotor symptoms at baseline (n = 574 or 28.3% of participants aged 50-54).12

**Power was not sufficient among minority groups to support the conclusions**

As with most medical studies, the results apply to the entire study population and are not specific to a particular racial/ethnic group. Although we did not report significant differences in outcomes among these groups, the 16% racial/ethnic minority participation does represent greater diversity than in previous hormone trials.9,10,13

**Reassignment of participants with a uterus from estrogen-alone to active E+P contaminated the intervention arm**

Only 331 women (4%) were reassigned, and findings did not differ substantially in analyses performed without these participants.2

We emphasize that in randomized clinical trials (RCTs), such as the WHI, study groups are balanced with regard to risk factors. As a result, RCTs are statistically more robust for determining treatment effects than other types of studies.14 The WHI investigators continue to examine data from the trial and invite informed discussions that enhance our understanding of the findings.

**References**

Evidence-based surgery for cesarean delivery

Vincenzo Berghella, MD, Jason K. Baxter, MD, MSCP, Suneet P. Chauhan, MD

Department of Obstetrics and Gynecology, Division of Maternal-Fetal Medicine, Jefferson Medical College of Thomas Jefferson University, Philadelphia, PA; Aurora Health Care, West Allis, WI

Received for publication January 21, 2005; revised March 18, 2005; accepted March 26, 2005

Objective: The purpose of this study was to provide evidenced-based guidance for surgical decisions during cesarean delivery.

Study design: We performed MEDLINE, PubMed, EMBASE, and COCHRANE searches with the terms cesarean section, cesarean delivery, cesarean, pregnancy, randomized trials, and each technical aspect of cesarean delivery. All randomized trials that covered a surgical aspect of cesarean delivery were included in the review. Each surgical step of cesarean delivery was reviewed separately.

Results: US Preventive Services Task Force recommendations favor blunt uterine incision expansion, prophylactic antibiotics (either ampicillin or first-generation cephalosporin for just 1 dose), spontaneous placental removal, non-closure of both visceral and parietal peritoneum, and suture closure or drain of the subcutaneous tissue when thickness is ≥2 cm.

Conclusion: Cesarean delivery techniques that are supported by good quality recommendations should be performed routinely. All technical aspects that have recommendations with lower quality should be researched with adequately powered and designed trials.

The word cesarean is derived from the Latin caesare, which means “to cut.” It is unlikely that Caesar was delivered by this technique. The first documented cesarean delivery (CD) was in 1020 AD.1 In 1882, the era of modern CD began when Saenger advocated closing all uterine incisions immediately after surgery. The lower uterine segment incision was introduced by Kronig in 1912 and popularized in the United States by DeLee in 1922. The transverse uterine incision was described by Munro-Kerr in 1926.2 CD has been associated with relatively low maternal mortality rates for >100 years. Safety has improved in the last 50 years, because these techniques have become more routine and antibiotics have been introduced.

CD is now the most common surgical procedure in the United States, with >1 million procedures performed per year. Its incidence increased to 27.6% of deliveries in 2003.3 This increase was fueled, at least in part, by the increased incidence of multiple gestations and the decreased incidences of vaginal births after CD and vaginal breech deliveries. Recently, women’s demands for elective CD have increased because complications from the procedure have diminished, women are having fewer children, and fear and concerns about vaginal delivery have not abated.

The aim of this manuscript was to review the evidence for the technique of CD and to offer recommendations based (where available) on randomized trials. Proper
technique, which has proved to be associated with the least complications, will minimize the morbidity and possibly the deaths that can be associated with CD. We aspire to stimulate better clinical treatment, promote education, foster research trials in areas of uncertainty, and focus on the prevention of possible complications rather than on treatment. We wanted to provide obstetricians with evidenced-based guidance for surgical decisions that are made in the operating room during a CD. We did not seek to review other aspects of CD (such as preoperative and postoperative care). We assume that during CD obstetricians would adhere to the best surgical techniques to decrease adhesions, to minimize tissue trauma, and to avoid ischemia and inflammation. We present one aspect of the CD technique at a time to show the effect of each step of the procedure individually. Therefore, we will not discuss proposed general approaches that involve the comparison of several steps at the same time, because direct comparison of the effects of each individual step is not possible.4,5

**Material and methods**

To achieve our aim, we performed MEDLINE, PubMed, EMBASE, and COCHRANE searches with the terms cesarean section, cesarean delivery, cesarean, pregnancy, randomized trials, plus each technical aspect (eg, lateral tilt, skin cleansing). The search was between 1966 and 2004 and was not restricted by language.

Each retrieved manuscript or Cochrane review was carefully evaluated, and any pertinent references from the manuscripts were obtained and reviewed. Our evidence-based review integrates individual clinical expertise with the best available external clinical evidence from systematic research.6 All randomized trials that covered surgical and selected nonsurgical aspects of CD were included in the review. In the absence of trials that adequately covered the aspect, analytic data were reviewed. In the absence of experimental or analytic data, observational data were evaluated. Each step of CD was reviewed separately. After each technical step was reviewed, evidence levels and recommendation levels, according to the new method outlined by the US Preventive Services Task Force,7 were reported (Table I).

**Results**

*Lateral tilt* involves tilting the woman towards her left side 10 to 15 degrees to avoid vena caval compression by the gravid uterus. Lateral tilt is compared with the supine position in 3 methodologically poor randomized trials that involved 293 women6-10 and that were summarized in a Cochrane Review.11 Lateral tilt did not provide fetal/neonatal benefit in Apgar scores (odds ratio [OR], 0.53; 95% CI, 0.25-1.16) or umbilical artery pH (mean difference, 0.03; 95% CI, 0.01-0.04). Fetal oxygen saturation was improved in a nonrandomized study of women in labor12 (recommendation: I; evidence: poor; Table I).

**Skin cleansing** techniques for CD have been studied only once in a randomized trial.13 One hundred women were assigned randomly to parachlorometaxylenol scrub for 5 minutes or not, after which all women received 7.5% povidone-iodine scrub and then povidine-iodine 10% solution. In the women who had received prophylactic antibiotics for CD, no significant differences in incidences of endometritis (OR, 2.7; 95% CI, 0.8-8.9) or wound infection (2/25 vs 0/25 women; OR, not calculable) were found between the 2 skin cleansing groups. This study was limited by a possible type II error, given the small numbers. This initial aspect of skin cleansing before CD is grossly understudied. Skin is impossible to sterilize. In nonpregnant adults, a meta-analysis of several trials found no differences in wound infection with different types and times of scrubs.14 Therefore, the use of an iodine solution alone (better than saline solution) is considered reasonable. Some investigators insist on the importance of letting the iodine dry for best infection prophylaxis (recommendation: I; quality: poor; Table I).

**Adhesive drapes** for CD were studied in 2 randomized trials, which included 1943 women.15,16 The 2 studies each reported no benefit of adhesive drapes for the prevention of wound infection. No Cochrane Review is available. In our own meta-analysis, the incidence of wound infection was higher in the adhesive drape group (13.8%; 133/967 women) compared with the control group (10.7%; 104/996 women; relative risk [RR] 1.34; 95% CI, 1.02-1.76). Therefore, adhesive drapes should not be recommended for the prevention of wound infection at CD (recommendation: D; quality: fair; Table I).

**Skin incision** techniques for CD were studied separately from other aspects of CD in 2 randomized trials, including 411 women.17,18 In general, a transverse skin incision is recommended, because this is associated with less postoperative pain and improved cosmetic effect compared with a vertical incision19 (recommendation: B; quality: fair; Table I).

The Pfannenstiel incision (slightly curved, 2 to 3 cm or 2 fingers above the symphysis pubis, with the mid portion of the incision within the shaved area of the pubic hair) and Joel-Cohen incision (straight, 3 cm below the line that joins the anterior superior iliac spines and therefore slightly more cephalad than the Pfannenstiel) are the preferred transverse incisions. These 2 incisions were compared in the 2 trials that have been published; no Cochrane Review is available.17,18 In an earlier trial on this issue,20 the Pfannenstiel group had the peritoneum closed, and the Joel-Cohen group did...
not, which makes a comparison of this single aspect of CD impossible. In the larger, better designed trial, no differences in total operative time (32 vs 33 minutes), intra- and postoperative complications, and neonatal outcomes were found. The extraction time was 50 seconds shorter for the Joel-Cohen group (median, 240 seconds [range, 50-600 seconds] vs 190 seconds [range, 60-600 seconds]). Considering the absence of clinical benefits to the mother and fetus, these authors concluded that there is no clear indication for the performance of a Joel-Cohen incision. In contrast, the smaller, less-well designed trial showed significantly shorter operating times and reduced blood loss and postoperative discomfort associated with the Joel-Cohen incision compared with the Pfannenstiel incision (recommendation: C; quality: fair; Table I).

Skin incision length has not been studied in a trial. Two non-randomized studies suggest that abdominal surgical incision size should provide at least 15 cm (size of a standard Allis clamp) of exposure to assure optimal outcome of both mother and fetus (recommendation: I; quality: poor; Table I).

Changing to a second scalpel after the first scalpel has been used for skin incision versus no such change has never been evaluated in a trial or in any obstetric literature. One randomized study in general surgery patients did not show any benefit from discarding the first knife after skin incision and concluded that one scalpel is adequate to use throughout the whole surgical procedure (recommendation: D; quality: fair; Table I).

Subcutaneous incision/opening has not been studied separately in a trial. Most clinicians use the scalpel as little as possible, opening layers bluntly from medial to lateral to avoid injury to tissue and the inferior epigastric vessels. Blunt dissection has been associated with shorter operating times. There are no trials to evaluate the safety or efficacy of electrosurgery, electrocautery, or diathermy (Bovie) during CD (recommendation: I; quality: poor; Table I).

Fascial incision has not been studied separately in a trial. Most experts recommend a transverse incision that is performed with the scalpel and then extended with scissors. Some clinicians have advocated digital extension, which can be accomplished by separating the forefingers in a cephalad-caudad direction after the fingers are inserted into a small, midline transverse fascial incision (recommendation: I; quality: poor; Table I).

Rectus muscle cutting has been studied in 3 trials that included 313 women; no Cochrane Review is available. These women were assigned randomly to either Maylard (muscle cutting) or Pfannenstiel (no muscle cutting) techniques. Transecting the rectus muscle was not associated with any difference in operative morbidity, difficult deliveries, postoperative complications, or pain scores in these studies. One study showed that abdominal muscle strength at 3 months was also similar, with a trend for better strength in the Pfannenstiel group. Therefore, rectus muscle cutting is probably not necessary (recommendation: D; quality: fair; Table I).

Dissection of fascia off the recti muscles has not been studied separately in a trial. Several investigators have cast doubt on the necessity of this commonly used technical step of CD (recommendation: I; quality: poor; Table I).
Opening of the peritoneum has not been studied separately in a trial. The peritoneum usually is opened carefully with blunt or sharp dissection and blunt expansion, high above the bladder, which avoids injury to the organs below (recommendation: I; quality: poor; Table I).

Bladder flap development versus no such development was studied in one randomized trial. This trial randomly assigned 102 women to either incision and opening of the bladder flap or to direct incision 1 cm above the bladder fold. Bladder flap development was associated with longer incision to delivery interval (7 vs 5 minutes; \( P < .001 \)), longer total operating time (40 vs 35 minutes; \( P = .004 \)), and greater change in hemoglobin level (1 vs 0.5 g/dL; \( P = .009 \)). Forming a bladder flap was also associated with more postoperative microhematuria (47% vs 21%; \( P < .01 \)) and greater need for analgesia (55% vs 26%; \( P = .006 \)) at 2 days after CD. There was only 1 postoperative fever in each group. No long-term effects (eg, adhesions, bladder function, fertility) were evaluated. Because bladder injury at CD is an uncommon event (1-3/1000 births), a sample size of >40,000 women would be required to show a difference in this outcome. Given these results and considerations, not developing a bladder flap at CD may be justified (recommendation: D; quality: fair; Table I).

The use of a bladder blade to protect the bladder has not been studied separately in a trial (recommendation: I; quality: poor; Table I).

Uterine incision type has not been studied separately in a trial. The transverse incision in the lower uterine segment is recommended by most experts and by retrospective case-control studies. Some experts advocate the classic vertical or at least low-vertical incision if the lower uterine segment is not large enough to allow a transverse incision (eg, for the very preterm [<28 weeks of gestation] uterus, fibroids), but this has been associated with increased blood loss compared with low transverse incision (recommendation: B; quality: fair; Table I).

Uterine stapling device versus traditional opening and closure of the uterine scar has been evaluated in 4 trials, which involved 526 women and was summarized in a Cochrane Review. The Cochrane Review states that there was no difference in total operating time between the stapling technique and other techniques to extend the incision (weighted mean difference, 1.17 minutes; 95% CI, −3.57-1.22). However, stapling devices increased the time that was needed to deliver the baby (weighted mean difference, 0.85 minutes; 95% CI, 0.48-1.23). Blood loss was lower with the use of staples (weighted mean difference, −41.22 mL; 95% CI, −50.63 to −31.8). No significant differences between stapling and other techniques were detected for other perinatal morbidity outcomes. The Cochrane reviewers concluded that there is not enough evidence to justify the routine use of stapling devices to extend the uterine incision at lower segment cesarean delivery. There is a possibility that stapling could cause harm, by prolonging the time to deliver the baby (recommendation: D; quality: fair; Table I).

Expansion of uterine incision either bluntly or by scissors was evaluated in 2 randomized trials, with no Cochrane Review. The first trial involved 147 sharp (scissors) versus 139 blunt (fingers) uterine incision expansions. The incidence of unintended expansions (14% vs 12%; \( P = .61 \)) and postpartum hemoglobin levels (9.9 vs 10.3 g/dL; \( P = .12 \)) were similar. No serious morbidities occurred. In the second larger and more recent trial, 470 sharp versus 475 blunt uterine incision expansions were compared. Sharp uterine incision expansions were associated with increased estimated blood loss (886 vs 843 mL; \( P = .001 \)), change in hematocrit (6.1% vs 5.5%; \( P = .003 \)), incidence of postpartum hemorrhage (13% vs 9%; RR, 1.23; 95% CI, 1.03-1.46), need for transfusion (2% vs 0.4%; RR, 1.65; 95% CI, 1.25-2.21), and total number of extensions (RR, 1.66; 95% CI, 1.47-1.86). In summary, these 2 trials show that sharp expansion significantly increases blood loss and the need for transfusion and is associated with more extensions. Because it is also quicker and is associated with less risk of inadvertently cutting the neonate or cord, blunt expansion should be preferred to sharp expansion of the uterine incision (recommendation: A; quality: good; Table I).

Instrumental delivery of the fetal head by either vacuum or forceps compared with manual means was evaluated only in a pilot randomized trial of 44 women with cephalic presentation who underwent CD. No controlled data are available on instrumental delivery of the aftercoming head at CD with breech presentation. Therefore, no firm recommendations can be made. Because instrumentation has been associated with maternal (especially for forceps) or fetal (especially for vacuum) harm in vaginal deliveries, the principle of primum non nocere (first do no harm) should be applied in this setting; therefore, manual delivery of the fetal head should be favored whenever possible until further data are available (recommendation: I; quality: poor; Table I).

Prophylactic antibiotics for CD have been evaluated in at least 81 randomized trials. This is certainly the best-studied aspect of CD. The comprehensive Cochrane Review shows benefit in both elective (non-laboring) and nonelective (laboring) CD. The decrease in incidence of endometritis was >60% in both elective (RR, 0.38; 95% CI, 0.22-0.64) and nonelective (RR, 0.39; 95% CI, 0.34-0.46) CD, with >2000 women who were randomly assigned in each analysis. The decrease in wound infection was approximately 25% (RR, 0.73; 95% CI, 0.53-0.99) in elective CD and approximately 65% (RR, 0.36; 95% CI, 0.26-0.51) in
nonelective CD. Overall, fever and urinary tract infections are also markedly decreased. These results justify a recommendation of prophylactic antibiotics in women who undergo any (elective or nonelective) CD (recommendation: A; quality: good; Table I).

Fifty-one trials have evaluated the appropriate antibiotic to give. The efficacy of ampicillin is equivalent to that of first-generation cephhalosporins, such as cefazolin (Ancef); later-generation, more expensive broad-spectrum agents do not improve efficacy further40 (recommendation: A; quality: good; Table I).

Systemic versus lavage routes of antibiotic administration seem to have similar efficacy40 (recommendation: I; quality: good; Table I).

Multiple systemic doses do not improve efficacy over a single dose40 (recommendation: D; quality: good; Table I).

Timing of antibiotic administration (at cord clamp versus preoperative) has been evaluated in 3 separate, underpowered studies that have shown no difference in infectious morbidity rates41-43 (recommendation: I; quality: poor; Table I).

If ampicillin or a first-generation cephhalosporin has already been given in labor, there may be no need for additional prophylactic antibiotics at CD (recommendation: I; quality: poor; Table I).

Prevention of uterine atony and postpartum hemorrhage has not been studied for CD but has been studied extensively for the third stage of labor after vaginal delivery.44 In the setting of vaginal delivery, both intravenous and intramuscular oxytocin effectively reduce postpartum hemorrhage and the need for therapeutic uterotonic by at least 40%, compared with placebo or no routine prophylactic agent. Oxytocin is as effective and has fewer side effects than ergot alkaloids (recommendation: I; quality: poor; Table I).

Regarding oxytocin infusion rates, in the setting of CD, only 1 blinded trial compared 10 units of oxytocin in 500 mL (infusion rate, 333 mU/min; n = 163 women) vs 80 units of oxytocin in 500 mL (infusion rate, 2667 mU/min; n = 158 women) lactated Ringer’s saline solution over 30 minutes after cord clamping in 321 women.45 The lower infusion was associated with an increased need for another uterotonics by at least 40%, compared with placebo or no routine prophylactic agent. Oxytocin is as effective and has fewer side effects than ergot alkaloids (recommendation: I; quality: poor; Table I).

Carbetocin as a single 100-µg dose was compared with 8- or 16-hour oxytocin infusion for the prevention of uterine atony at CD in 2 Canadian trials that included >750 women.46,47 They reported more effective prevention of uterine atony and lower need for additional uterotonics with carbetocin. Carbetocin (where available) may be recommended over oxytocin for prevention of uterine atony (recommendation: C; quality: fair; Table I).

Placental removal options of either spontaneous (with gentle cord traction) or manual placental removal at CD have been studied in 6 randomized trials that included >1700 women.48-54 The Cochrane Review of this topic has not been substantively updated since the publication of the 3 most recent studies.55 Our meta-analysis of the data shows a significant reduction in postoperative endometritis (OR, 0.62; 95% CI, 0.48-0.80) for spontaneous removal in the 5 studies that reported on this outcome (n = 1699 women).49-54 No reduction of wound infection for the spontaneous removal group was shown (OR, 0.43; 95% CI, 0.15-1.26) in the only study that reported on this outcome (n = 335 women).53 The few studies that recorded blood loss or changes in hemoglobin/hematocrit level usually reported a benefit for spontaneous removal, which included less feto-maternal hemorrhage. Most investigators hypothesize that blood loss is increased in manual removal because dilated sinuses in the uterine wall are not closed yet. Bacterial contamination of the lower uterine segment and incision may contaminate the surgeon’s dominant hand36; therefore, the upper segment in manual removal or the glove itself may be contaminated. Spontaneous placental removal should be preferred to manual removal, given the significant decrease in endometritis (recommendation: A; quality: good; Table I).

One trial evaluated the benefit of changing the operator’s glove before manual removal of the placenta, but this did not alter the incidence of endometritis12 (RR, 1.0, 95% CI, 0.8-1.3; recommendation: D; quality: fair; Table I).

Uterine exteriorization has been compared with leaving the uterus infra- abdominally for uterine incision repair in 6 randomized trials that involved 1221 women,50,51,57-60 which were summarized in a Cochrane Review.61 Exteriorization was associated with a significant decrease in fever for >3 days (OR, 0.41; 95% CI, 0.17-0.97). There were no other statistically significant differences for other important outcomes, which included bleeding. For most outcomes, relatively few studies contributed data. The balance of the benefits and harms is too close to justify a general recommendation (recommendation: C; quality: fair; Table I).

Cleaning any placental remnants or blood clots from the uterus with a sponge or other means is a technique that is used frequently after placental removal but that has not been studied in any trial (recommendation: I; quality: poor; Table I).

Closure of uterine incision with either a single or double layer of suture was studied in 2 trials that
involved 1006 women\textsuperscript{62,63} and was summarized in a Cochrane Review.\textsuperscript{64} The larger trial (n = 906 women)\textsuperscript{63} used chromic catgut and found a significant decrease in operating time of 5.6 minutes (P = .0001). Blood loss, need for transfusion, and endometritis rates were not different in the 2 groups in this study. The smaller study found a lower incidence of abnormal scar (smoother healing) for the single-layer group during hysterography at the 3-month follow-up examination (26% vs 88%; RRR, 0.30; 95% CI, 0.18-0.48).\textsuperscript{62} Blinding was not mentioned in this last study. Chapman et al\textsuperscript{65} followed 145 of the 906 women who experienced subsequent labor in the Hauth study.\textsuperscript{63} One of 70 women in the 1-layer versus none of 75 women in the 2-layer group had dehiscence; none of the women had uterine rupture; and all newborn infants did well. Unfortunately, these numbers are still too small to detect a significant difference in these rare, but extremely important, long-term outcomes. Three larger retrospective reviews yielded contradictory results.\textsuperscript{66-68} The first 2 series\textsuperscript{66,67} used chromic catgut as the suture, with the first layer interlocked, and the third series\textsuperscript{68} used polyglactin 910 suture (0-Vicryl). Continuous nonlocking single-layer closure was associated with decreased operating times and no adverse outcomes, compared with 2-layer repair of a locking suture that was followed by an imbricating layer in a case-control study with no long-term follow-up information.\textsuperscript{59} Because there is as of yet no trial that has demonstrated a benefit from 2-versus 1-layer uterine closure, it might be reasonable to omit the second layer if the woman is planning no more pregnancies (eg, receives tubal ligation). For women who are planning future pregnancies, the uterus could be closed in 2 layers. In a retrospective study, a single-layer surgical closure, compared with a double-layer closure, has been found to be a highly significant risk factor for uterine rupture during attempted vaginal birth after cesarean delivery.\textsuperscript{67} There appear to be no advantages for the routine use of single-layer closure compared with 2-layer closure, except perhaps a shorter operation time\textsuperscript{64} (recommendation: C; quality: fair; Table I).

Experts usually advocate the incorporation of all of the muscle up to the serosa in a 1-layer closure to avoid bleeding from edges, but this aspect of CD has never been studied properly (recommendation: I; quality: poor; Table I). It is unclear whether one needs to incorporate the decidua or not (recommendation: I; quality: poor; Table I). In a small case-control study (n = 81 women), continuous single-layer closure saved operating time and reduced blood loss compared with interrupted single-layer closure\textsuperscript{70} (recommendation: B; quality: fair; Table I).

Closure of the uterus, peritoneum, and rectus sheath with blunt or sharp needles was evaluated in 1 randomized trial that involved 203 women.\textsuperscript{71} No significant difference in outcome was found in these 2 groups (recommendation: I; quality: poor; Table I).

Peritoneal closure versus nonclosure was evaluated in 9 trials that included 1811 women\textsuperscript{72-80} and was summarized in a Cochrane Review.\textsuperscript{81} Some of these studies evaluated either visceral or parietal peritoneal closure individually, although some of the studies evaluated both together. Nonclosure of the peritoneum reduced operating time whether both or either peritoneal layer was not sutured. For both layers, the operating time was reduced by 7.33 minutes (95% CI, –8.43 to –6.24). There was significantly less postoperative fever and reduced postoperative hospital stay for visceral peritoneum and for both layer nonclosure. There were no other statistically significant differences. The trend for analgesia requirement and wound infection tended to favor nonclosure. The long-term follow-up\textsuperscript{82} of one trial\textsuperscript{77} after 7 years showed no differences in pain, fertility, urinary symptoms, and adhesions. A recent review, which also included a separate review of general surgery and gynecologic data, concluded that “we encourage clinicians not to close both parietal and visceral peritoneum.”\textsuperscript{83} Observational studies have shown that the peritoneum regenerates in 5 to 6 days. Long-term studies after CD are limited, but data from other surgical procedures are reassuring. The hypothetic benefits of closing these layers for anatomic barrier, reduction of wound dehiscence, and minimization of adhesion have not been proven and, in fact, have been invalided by these trials. There is, at present, no evidence to justify the time taken and cost of peritoneal closure (recommendation: A; quality: good; Table I).

Intra-abdominal irrigation with 500 to 1000 mL of normal saline solution versus no irrigation before abdominal wall closure to decrease maternal morbidities was evaluated in 1 trial that involved 196 women.\textsuperscript{84} No significant differences in blood loss, intrapartum complications, hospital stay, return of gastrointestinal function, or incidence of infectious complications were found (recommendation: D; quality: fair; Table I).

Reapproximation of rectus muscles has not been studied in any trial. Most clinicians agree that the muscles find the right anatomic location by themselves and that suturing them together can cause unnecessary pain when the woman starts to move after surgery (recommendation: I; quality: poor; Table I).

Techniques of fascial closure have not been studied in any trial of CD. Most experts suggest continuous nonlocking closure with delayed-absorbable suture (recommendation: I; quality: poor; Table I).

Irrigation of the subcutaneous tissue to minimize wound infections and other complications has not been studied versus no irrigation in a randomized CD trial. The type of irrigation, with saline or antibiotic solution, has also not been studied in a trial (recommendation: I; quality: poor; Table I).
<table>
<thead>
<tr>
<th>CD technical aspect</th>
<th>Recommendation*</th>
<th>Quality*</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lateral tilt</td>
<td>I</td>
<td>Poor</td>
<td>15 degrees to left</td>
</tr>
<tr>
<td>Skin cleaning</td>
<td>I</td>
<td>Poor</td>
<td>Iodine†</td>
</tr>
<tr>
<td>Adhesive drapes</td>
<td>D</td>
<td>Fair</td>
<td>Not recommended</td>
</tr>
<tr>
<td>Skin incision</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>C</td>
<td>Fair</td>
<td>Pfannenstiel or Joel-Cohen</td>
</tr>
<tr>
<td>Length</td>
<td>I</td>
<td>Poor</td>
<td>15 cm</td>
</tr>
<tr>
<td>Changing to a second scalpel</td>
<td>D</td>
<td>Fair</td>
<td>Not recommended</td>
</tr>
<tr>
<td>Subcutaneous incision</td>
<td>I</td>
<td>Poor</td>
<td>†</td>
</tr>
<tr>
<td>Fascial incision</td>
<td>I</td>
<td>Poor</td>
<td>†</td>
</tr>
<tr>
<td>Rectus muscle cutting</td>
<td>D</td>
<td>Fair</td>
<td>Not recommended</td>
</tr>
<tr>
<td>Dissection of fascia off rectus</td>
<td>I</td>
<td>Poor</td>
<td>†</td>
</tr>
<tr>
<td>Opening of peritoneum</td>
<td>I</td>
<td>Poor</td>
<td>†</td>
</tr>
<tr>
<td>Bladder flap</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Development</td>
<td>D</td>
<td>Fair</td>
<td>†</td>
</tr>
<tr>
<td>Use of bladder blade</td>
<td>I</td>
<td>Poor</td>
<td>Operator preference</td>
</tr>
<tr>
<td>Uterine incision</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>B</td>
<td>Fair</td>
<td>Transverse</td>
</tr>
<tr>
<td>Stapling device</td>
<td>D</td>
<td>Fair</td>
<td>Not recommended</td>
</tr>
<tr>
<td>Expansion of uterine incision</td>
<td>A</td>
<td>Good</td>
<td>Blunt</td>
</tr>
<tr>
<td>Instrumental delivery</td>
<td>I</td>
<td>Poor</td>
<td>†</td>
</tr>
<tr>
<td>Prophylactic antibiotics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes or no</td>
<td>A</td>
<td>Good</td>
<td>Yes (all CDs)</td>
</tr>
<tr>
<td>Antibiotic type</td>
<td>A</td>
<td>Good</td>
<td>†</td>
</tr>
<tr>
<td>Route of administration</td>
<td>A</td>
<td>Good</td>
<td>Systemic/lavage (equivalent)</td>
</tr>
<tr>
<td>Multiple systemic doses</td>
<td>D</td>
<td>Good</td>
<td>Not recommended</td>
</tr>
<tr>
<td>Timing</td>
<td>I</td>
<td>Poor</td>
<td>†</td>
</tr>
<tr>
<td>Prevention of uterine atony</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxytocin vs placebo</td>
<td>I</td>
<td>Poor</td>
<td>Oxytocin†-‡</td>
</tr>
<tr>
<td>Oxytocin infusion rate</td>
<td>I</td>
<td>Poor</td>
<td>Optimal rate unclear†</td>
</tr>
<tr>
<td>Carbetocin vs oxytocin†</td>
<td>C</td>
<td>Fair</td>
<td>Carbetocin, if available†</td>
</tr>
<tr>
<td>Placental removal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spontaneous vs manual</td>
<td>A</td>
<td>Good</td>
<td>Spontaneous</td>
</tr>
<tr>
<td>Glove change</td>
<td>D</td>
<td>Fair</td>
<td>Not recommended</td>
</tr>
<tr>
<td>Uterine exteriorization</td>
<td>C</td>
<td>Fair</td>
<td>Operator preference</td>
</tr>
<tr>
<td>Cleaning of uterus</td>
<td>I</td>
<td>Poor</td>
<td>†</td>
</tr>
<tr>
<td>Closure of uterine incision</td>
<td>C</td>
<td>Fair</td>
<td>†</td>
</tr>
<tr>
<td>Two vs one layers</td>
<td>C</td>
<td>Fair</td>
<td>†</td>
</tr>
<tr>
<td>Incorporation decidua/serosa</td>
<td>I</td>
<td>Poor</td>
<td>†</td>
</tr>
<tr>
<td>Continuous vs interrupted</td>
<td>B</td>
<td>Fair</td>
<td>Continuous</td>
</tr>
<tr>
<td>Sharp vs blunt needles</td>
<td>I</td>
<td>Poor</td>
<td>†</td>
</tr>
<tr>
<td>Peritoneal closure</td>
<td>D</td>
<td>Good</td>
<td>Not recommended (for both parietal and visceral)</td>
</tr>
<tr>
<td>Intra-abdominal irrigation:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saline vs none</td>
<td>D</td>
<td>Fair</td>
<td>Not recommended</td>
</tr>
<tr>
<td>Reapproximation of rectus muscles</td>
<td>I</td>
<td>Poor</td>
<td>†</td>
</tr>
<tr>
<td>Techniques of fascial closure</td>
<td>I</td>
<td>Poor</td>
<td>†</td>
</tr>
<tr>
<td>Irrigation of the subcutaneous tissue</td>
<td>I</td>
<td>Poor</td>
<td>†</td>
</tr>
<tr>
<td>Subcutaneous tissue</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any thickness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Closure vs non-closure</td>
<td>D</td>
<td>Fair</td>
<td>Closure not recommended</td>
</tr>
<tr>
<td>Drain vs no drain</td>
<td>D</td>
<td>Fair</td>
<td>Drain not recommended</td>
</tr>
<tr>
<td>≥ 2 cm thickness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Closure vs nonclosure</td>
<td>A</td>
<td>Good</td>
<td>Closure recommended</td>
</tr>
<tr>
<td>Drain vs no drain</td>
<td>A</td>
<td>Good</td>
<td>Drain recommended</td>
</tr>
<tr>
<td>Closure vs drain</td>
<td>C</td>
<td>Fair</td>
<td>†</td>
</tr>
<tr>
<td>Closure of skin: Staples vs subcuticular suture</td>
<td>I</td>
<td>Poor</td>
<td>†</td>
</tr>
</tbody>
</table>

* Level of evidence was based on the USPSTF recommendations (Table I).†
† See text for more detail.
‡ Ampicillin or first-generation cephalosporin for just 1 dose before or after cord clamping.
§ Based on trials of women with vaginal delivery, not CD.
**Subcutaneous tissue closure** versus nonclosure has been evaluated in 6 studies\(^{85-90}\) and was summarized in a recent meta-analysis\(^{91}\) and in a Cochrane Review.\(^{92}\) The Cochrane Review did not analyze the trials by the thickness of the subcutaneous tissue, which is how most trials were done. Some studies have evaluated drainage of subcutaneous tissue.\(^{93}\) Most studies used 3-0 Vicryl for closure. These studies and their results should be stratified by subcutaneous tissue thickness, because results differ markedly by this variable.

Suture closure versus nonclosure of subcutaneous fat in women with **any subcutaneous thickness** was evaluated in 3 trials that involved 875 women.\(^{35,87,90}\) Wound disruption was decreased in the closure group (RR, 0.56; 95% CI, 0.36-0.86). Inability to blind represents a possible source of bias. The 2 trials that separately reported on subcutaneous depth of <2 cm did not show a difference (RR, 1.01; 95% CI, 0.46-2.20).\(^{87,89}\) Therefore, routine subcutaneous tissue closure in women with a depth <2 cm cannot be recommended (recommendation: D; quality: fair; Table I).

Drainage of subcutaneous tissue in women with any thickness was evaluated in 1 trial that involved 242 women.\(^{93}\) No prophylactic antibiotics were given. The women in the study group received a 2-cm corrugated rubber drain, which was left to drain open, coming out of 1 end of incision. The drain was removed the next day. A trend towards increased wound infection was seen (RR, 1.91; 95% CI, 0.81-4.50). Therefore, routine subcutaneous tissue drainage in women with a depth <2 cm cannot be recommended (recommendation: D; quality: fair; Table I).

Suture closure versus nonclosure of subcutaneous fat in women with **≥2-cm subcutaneous thickness** has been evaluated in 5 trials that involved 887 women.\(^{84-90}\) Suture closure was associated with a significant decrease in wound disruptions (RR, 0.66; 95% CI, 0.48-0.91). Wound disruptions usually were defined as any wound complication that required intervention. Suture closure was associated with a significant decrease in seromas (RR, 0.42; 95% CI, 0.24-0.75) in the 4 trials that involved 852 women in which it was evaluated. The evidence supports routine subcutaneous suture closure in women with a depth ≥2 cm (recommendation: A; quality: good; Table I).

Drainage of subcutaneous fat versus no drainage or versus suture closure in women with ≥2-cm thickness was studied in 2 trials, with the use of a 7-mm Jackson-Pratt drain with closed suction cup.\(^{88,90}\) Compared with no drain or no closure, drainage was associated with a decrease in wound complications (OR, 0.44; 95% CI, 0.26-0.74; n = 600 women). Drainage was associated with an incidence of wound complications that was similar to suture closure (OR, 0.67; 95% CI, 0.39-1.20; n = 573 women). Therefore, although suture closure or drainage in women with ≥2-cm thickness is associated with benefit compared with no suture or no drainage, it is not yet clear whether any of these prophylactic interventions is superior to the other. Drainage is recommended over no drainage (recommendation: A; quality: good; Table I).

**Closure of skin** by either staples or subcuticular suture was evaluated in only 1 trial\(^{94}\) and was reported in a Cochrane Review.\(^{95}\) The 50 women (with complete data) who had received a Pfannenstiel incision were randomized equally to either staples or 4-0 Vicryl subcuticular suture. Staples were associated with decreased operative time (<1 vs 10 minutes; \(P < .001\)) but with increased number of pain pills in the hospital (24.6 vs 19.7; \(P = .008\)) and increased pain scale at hospital discharge (6.6 vs 5.1; \(P = .003\)) and after delivery (2.0 vs 0.5; \(P < .001\)). Staples were associated with similar appearance by the linear trend test, as rated by physician (7/25 vs 2/25 patients; OR, 4.47; 95% CI, 0.83-24.19) and by patient (4/25 vs 0/25 patients; OR, 10.67; 95% CI, 0.54-209.66) in the short-term (6-week) assessment. There was no blinding possible. No long-term outcome was reported. The general surgery literature seems to support the evidence that subcuticular suture is associated with less pain and better cosmesis. CD is considered a clean-contaminated procedure, with a relatively high risk of wound infection. In clean-contaminated procedures, the general surgical literature would suggest avoidance of a continuous method of wound closure. In contrast with interrupted closure (with staples or sutures), continuous wound closure does not allow selective minimal wound opening in cases of infection or collections. Therefore, there is no conclusive evidence about how the skin should be closed after CD (recommendation: I; quality: poor; Table I).

**Comment**

Approximately 2 CDs are started every minute in the United States alone; this is the most frequent major operation that is performed in this country. Good fetal and maternal outcome is more important than any preset incidence of CD. As CDs are performed more frequently, it is imperative to use the CD technique that is safest for mother and fetus. Because several major and minor morbidities are possible at CD, adherence to proper technique allows the prevention of complications. Instead, even after dozens of trials and hundreds of excellent articles, expert opinion frequently guides decisions about the technical details of CD. Too often we do not take advantage of all the knowledge that is already available, either because of ignorance or of feelings of superiority. Complications that occur because of improper technique put both mother and fetus at unnecessary but significant risks.

Table II summarizes our evidence-based recommendations. Clinically, good-quality recommendations with benefit compared with no suture or no drainage, it is not yet clear whether any of these prophylactic interventions is superior to the other. Drainage is recommended over no drainage (recommendation: A; quality: good; Table I).
favor blunt uterine incision expansion, prophylactic antibiotics (either ampicillin or first-generation cephalosporin for just 1 dose), spontaneous placental removal, nonclosure of both visceral and parietal peritoneum, and suture closure or drain of the subcutaneous tissue when the thickness is $\geq 2$ cm. These CD techniques should be performed routinely. All technical aspects that have recommendations with less than good quality should be researched properly with adequately powered and designed trials. Obstetricians who are in practice or training should be educated to be better aware of these clinical and research recommendations.

References


Uterine artery embolization versus hysterectomy in the treatment of symptomatic uterine fibroids (EMMY trial): Peri- and postprocedural results from a randomized controlled trial

Wouter J. K. Hehenkamp, MD,a,* Nicole A. Volkers, MD,b Peter F. J. Donderwinkel, MD,c Sjoerd de Blok, MD, PhD,e Erwin Birnie, PhD,d Willem M. Ankum, MD, PhD,a Jim A. Reekers, MD, PhDb

Department of Gynecology, a Department of Radiology, b Department of Public Health Epidemiology, d Academic Medical Center, Amsterdam, The Netherlands; Department of Gynecology, e Martini Hospital, Groningen, The Netherlands; Department of Gynecology, Department of Onze Lieve Vrouwe Gasthuis, Amsterdam, The Netherlands

Received for publication February 14, 2005; revised March 23, 2005; accepted May 3, 2005

KEY WORDS
Uterine artery embolization 
Fibroids 
Menorrhagia 
Randomized controlled trial 
Safety 
Hysterectomy

Objective: This was a randomized controlled trial to evaluate the safety of uterine artery embolization (UAE) compared with hysterectomy.

Study design: Twenty-eight Dutch hospitals recruited 177 patients with symptomatic uterine fibroids and menorrhagia who were eligible for hysterectomy. Patients were randomized to UAE (n = 88) or hysterectomy (n = 89). In this paper we evaluate the peri- and postprocedural complications, length of hospital stay, unscheduled visits, and readmission rates up to 6 weeks’ post-intervention. Analysis was by intention to treat.

Results: Bilateral UAE failure occurred in 4 patients (4.9%). Major complications occurred in 4.9% (UAE) and 2.7% (hysterectomy) of cases (P = .68). The minor complication rate from discharge until 6 weeks after was significantly higher in the UAE group than in the hysterectomy group (58.0% vs 40.0%; RR 1.45 [1.04-2.02]; P = .024). UAE patients were more often readmitted (11.1% vs 0%; P = .003). Total length of hospital stay was significantly shorter in UAE patients (mean [SD]: 2.5 [2.7] vs 5.1 [1.3], P < .001).

Conclusion: UAE is a procedure similar to hysterectomy with a low major complication rate and with a reduced length of hospital stay. Higher readmission rates after UAE stress the need for careful postprocedural follow-up.

© 2005 Mosby, Inc. All rights reserved.

The Emmy study is funded by ZonMw ‘Netherlands Organisation for Health Research and Development’ (grant application number 945-01-017), and supported by Boston Scientific Corporation, The Netherlands.

Drs Hehenkamp and Volkers contributed equally to this paper.

* Reprint requests: W. J. K. Hehenkamp, MD, Academic Medical Center, Department of Gynecology, Meibergdreef 9; 1105 AZ Amsterdam, The Netherlands.

E-mail: w.j.k.hehenkamp@amc.uva.nl

002-9378/S - see front matter © 2005 Mosby, Inc. All rights reserved.
doi:10.1016/j.ajog.2005.05.017
Uterine artery embolization (UAE) for the treatment of heavy menstrual bleeding caused by uterine fibroids was first described in 1995. Since then, several large case series have been published describing the risks and benefits of UAE. These reports suggest that UAE may have advantages over surgery, but are hampered by the inclusion of patients with strong treatment preferences and the lack of a control group. Obviously, this seriously affects the validity and generalizability of their results.

To evaluate the safety and efficacy of UAE in comparison to the standard treatment, ie, hysterectomy, we initiated a prospective, multicenter, randomized controlled trial comparing UAE with hysterectomy for the treatment of menorrhagia caused by uterine fibroids. In the trial, patients were followed until 2 years after the intervention. In this report, we present the baseline and procedural characteristics, peri- and postprocedural complications, duration of hospital stay, unscheduled visits, and readmissions up to 6 weeks’ post-intervention.

Material and methods

Study design

The EMbolization versus hysterectoMY (EMMY) study is a multicenter, randomized controlled trial, conducted in The Netherlands. Five university hospitals and 29 general hospitals participated in the trial.

Patients visiting the gynecologic outpatient clinics were asked to participate if they met the following criteria: 1) the clinical diagnosis of uterine fibroids had been confirmed by ultrasonography; 2) menorrhagia (subjectively reported by the patient as increased or prolonged menstrual blood loss which causes dysfunction in daily life) was their predominant complaint, among other possibly fibroid-related signs and symptoms; 3) they were premenopausal; and 4) they were to be scheduled for a hysterectomy. Whenever other treatment options were still available, women were not asked to participate, but were treated otherwise.

Women were excluded if: 1) preservation of the uterus was warranted for future pregnancy; 2) renal failure (creatinine >150 mmol/L), active pelvic infection, or clotting disorders were clinically established; 3) they were allergic to contrast material; 4) uterine malignancy was suspected; 5) submucosal fibroids with 50% of their diameter within the uterine cavity or dominant pedunculated serosal fibroids were present.

After written informed consent had been obtained the attending gynecologist contacted the trial bureau by telephone, where the patient was registered and randomly assigned (1:1) to UAE or hysterectomy, using a computer-based minimization scheme (‘balancing procedure’), and stratified for study center. The randomization result was recorded electronically.

According to Dutch guidelines, the study was approved by the Central Committee Involving Human Subjects (www.ccmo.nl) and by local ethics committees of participating hospitals.

Preassessment

All clinical data were prospectively recorded in a standardized case record form during the entire study period. All patients underwent a pelvic ultrasound either transvaginally or transabdominally. The uterus and the largest fibroid were measured in 3 dimensions, ie, longitudinal (D1), anterior-posterior (D2), and transverse (D3). Volumes were calculated using the formula (0.5233 × D1 × D2 × D3).

Procedures

Uterine artery embolization

Patients were advised to discontinue any GnRH analogues treatment at least 1 month before the UAE. UAE was performed in all participating hospitals. The first 2 to 3 procedures were supervised by an interventional radiologist (J.R.) with ample experience in UAE. All radiologists were experienced in interventional radiology, including various embolization techniques in general. At the start of the study UAE was not a routine procedure for all radiologists. Seven radiologists were considered experienced in UAE group (having performed >10 UAE procedures), and 19 interventional radiologists had less experience in UAE (having performed <10 UAE procedures). Patients received an intravenous line and a Foley catheter before UAE. UAE was performed under local or epidural/spinal anesthesia. The use of analgesics and antibiotics was not standardized. Femoral artery access could be unilateral or bilateral. A 4-F or 5-F catheter was introduced into the femoral artery and advanced over the aortic bifurcation to the contralateral internal iliac artery to identify the origin of the uterine artery. In case of spasm, the policy was to wait, but a microcatheter and/or spasmyotics could be used within the study protocol. When catheters were placed correctly, the actual embolization was carried out. Polyvinyl alcohol particles (PVA, Contour, Boston Scientific, Beck, The Netherlands) with a size of 355 to 500 μm, were used. Only if an anastomosis with the ovarian artery was observed were 500 to 700 μm particles used. PVA, mixed with contrast medium and saline, was injected into each uterine artery until parenchyma filling of the fibroids had stopped (target embolization), or until the main uterine artery was blocked with stasis of contrast (selective embolization). After the procedure, groin pressure was applied for 10 to 15 minutes.

According to the Cardiovascular and Interventional Radiology Society of Europe guidelines, UAE was
considered successful whenever bilateral UAE was established; unilateral UAE was only considered a successful procedure if single-sided uterine arterial flow to the fibroids was present.  

If a uterine artery was absent and flow to the fibroids came solely from the ovarian artery, the procedure was stopped because of risk for ovarian damage, and considered unsuccessful. Also, in case of extensive collaterals to the cervix and vaginal wall, the procedure was stopped and considered unsuccessful.

Unsuccessful procedures may not always result from the technical inability to selectively catheterize the uterine artery. Therefore, we also calculated the true technical failure rate as the total number of arteries that could be embolized (ie, arteries were present without extensive collaterals with the cervico-vaginal vascular system), but which were not embolized because of technical inabilities to do so.

The type of anesthesia, type of UAE, the amount of PVA vials used, the amount of blood loss, the procedural complications, and the duration of the procedure were recorded. After the procedure, women were admitted to the gynecology ward for further care. All patients were advised to stay in hospital for at least 1 night. At discharge, all patients were no longer using opiates and received clear instructions on pain medication regiments. They also received written instruction with contact numbers to contact their gynecologist whenever uncontrollable pain, persistent fever, or expulsion of fibroids occurred.

**Hysterectomy**

The type of hysterectomy and the route of access were left at the discretion of the attending gynecologist in order to keep as close to daily practice as possible. The following procedures were allowed: abdominal hysterectomy, either by median or a pfannenstiel incision, vaginal hysterectomy, laparoscopically assisted vaginal hysterectomy (LAVH), and laparoscopic hysterectomy. Both supravaginal and total hysterectomies were allowed. We used no guidelines for: antibiotic prophylaxis; type of anesthesia; removal or ablation of endocervical tissue in the supravaginal hysterectomy group; concomitant adnexal surgery; wound closure; evaluation and treatment of fever; or hospital discharge criteria. Prospectively recorded were: prescription of antibiotics, type of anesthesia, type of hysterectomy, removal of the cervix, ovaries, or other procedures, complications, blood loss, and duration of procedure. At discharge, patients were instructed in a similar fashion as for the UAE patients.

**Follow-up**

Complications were classified as “major” when the events were potentially life-threatening, could lead to permanent sequelae, or required surgical intervention. Other complications were listed as “minor.” Nausea, pain, and fever were considered “general” complications. Whenever a definite cause of fever was identified (eg, urinary tract infection), this was listed under minor or major complications, using the criteria described above.

Complications were separately listed for 2 time intervals: the hospitalization period (ie, occurring during and after the procedure) and the first 6 weeks thereafter (ie, between discharge and first routine visit at 6 weeks after the procedure). Complication rates were expressed as the occurrence of at least 1 complication within a patient and calculated for minor and major complications separately in both time intervals and overall.

All UAE patients were routinely telephoned by the gynecologist 1 week after discharge to inquire about their health status.

At the first routine visit (6 weeks after the procedure), complications after discharge, unscheduled visits, readmissions, and reinterventions were recorded.

**Sample size and end points**

The primary end point of this trial was the elimination of menorrhagia after a follow-up period of 2 years. UAE was considered equivalent to hysterectomy when menorrhagia resolved in at least 75% of patients, with preservation of the uterus and no significant differences in major complications between both procedures. To reject the null hypothesis that UAE and hysterectomy are not clinically equivalent (expected effectiveness of UAE = 0.875; expected effectiveness of hysterectomy = 0.999; threshold value \( \Delta = 0.25; \alpha = 0.05 \) un-sided; \( 1 - \beta = 0.90 \)), at least 2 patients had to be included.

The objective of the present study was to compare the following end points between both interventions: technical failures, procedure safety, complications, duration of hospital stay (discharge date minus procedure date), and the occurrence of unscheduled visits, readmissions, and reinterventions. For this analysis, no separate power calculation was made.

**Statistical analysis**

All data entries were visually double checked by an independent second investigator. Analyses were done using SPSS statistical software (version 11.5.1, Chicago, IL).

Study outcomes were analyzed according to original treatment assignment (intention to treat). Differences in baseline characteristics were tested with multiple logistic regression analysis. Differences in complications between groups were expressed in absolute numbers, rates, and relative risks (RR) with 95% CI. Confidence intervals were calculated with Statcalc (EpiInfo version 5, Centers for Disease Control and Prevention, Atlanta, GA). Differences in hospital stay were tested with the
Mann-Whitney U test. Differences in categorical data were compared with $\chi^2$-tests or Fisher exact tests if appropriate. We also investigated the effect of experience of the radiologist and hospitals performing UAE on technical failure, complications, and readmissions. A $P$ value of < .05 was considered statistically significant.

**Results**

**Patients**

Patients were enrolled between March 2002 and February 2004. Twenty-eight of the 34 participating hospitals included patients. Of 349 eligible patients, 177 were randomized: 88 were allocated UAE and 89 hysterectomy (Figure). The majority of patients refusing participation did so for a strong preference for hysterectomy (58%) or for UAE (21%). After randomization 7 patients in the UAE group and 14 patients in the hysterectomy group refused the allocated treatment. Patients who refused the assigned treatment were comparable to participating patients in terms of: age, race, BMI, parity, symptoms, and duration of symptoms (data not shown). The mean age was 44.6 years (UAE group) and 45.4 years (hysterectomy group). Participants were predominantly white: 61.4% and 64.0% for UAE and hysterectomy respectively (Table I). Table II shows that most patients (85.3%) had already received 1 or more treatments for symptomatic uterine fibroids before study enrollment. Patients suffered from menorrhagia for a median of 24 months. Other symptoms besides menorrhagia were prevalent. The majority of women had multiple fibroids. Fibroid volumes were higher in the hysterectomy group. Logistic regression analysis did not reveal baseline characteristics that could predict randomization outcome, confirming successful randomization.

**Procedures**

UAE was successfully performed in 72 of 81 patients, 5 of whom had a unilateral procedure because of single-sided arterial blood flow to the fibroid (procedural success rate: 88.9%). The remaining 11.1% consisted of 5 patients (6.2%) with a unilateral procedure (caused by technical failure on the other side) and 4 patients (4.9%) with bilateral unsuccessful UAE. The bilateral impossibility to embolize resulted from bilateral absence of uterine artery flow to the fibroids (n = 2), bilateral technical failures (n = 1), and extensive anastomoses with the cervix/vagina on 1 side and a technical failure on the other (n = 1). These 4 patients subsequently underwent hysterectomy, but were analyzed in the UAE group. The total number of arteries that could potentially be embolized in the 88 UAE patients was 152. Of these, 8 arteries were not embolized because of technical inability (technical failure rate: 5.3%).

Table III displays the characteristics of both treatments. In most cases (86.1%), target embolization was carried out. For technically successful UAE, a median of 1 vial (range 0.1-3) of PVA was used for each artery. In the hysterectomy group, all operations were technically successful. Four conversions took place: 3 procedures (1 LAVH, 1 vaginal, and 1 laparoscopic hysterectomy) were converted to a laparotomy. In 1 abdominal hysterectomy, the cervix could not be removed.
as planned because of adhesions, and a supravaginal hysterectomy was carried out instead. Furthermore, in 1 vaginal hysterectomy, morcellation was necessary for a large fibroid. Most hysterectomies were performed transabdominally (84.0%).

UAE procedures on average took shorter than hysterectomy procedures (79.0 vs 95.4 minutes, \( P = .007 \)). Patients subject to UAE had significantly less blood loss than those undergoing hysterectomy (30.9 and 436.1 mL, respectively; \( P < .001 \)). Total admission time was significantly (\( P < .001 \)) shorter in the UAE group (mean 2.0 days; SD 2.1; range 0-13 days) than in the hysterectomy group (mean 5.1 days; SD 1.3; range 2-8 days).

### Complications during hospital stay

Table IV lists complications occurring during and after the procedures.

Intraprocedural complications were uncommon in both groups. In the UAE group, 7 minor complications occurred: 5 postpuncture hematomas, 1 blood clot in the gluteal artery, which resolved spontaneously, and 1 case of nausea during the procedure. In the hysterectomy group, 2 minor complications occurred: 1 allergic reaction to an anesthetic agent and 1 small tear in the rectus muscle.

During hospital stay febrile morbidity was significantly less common in the UAE group (4.9%) than after hysterectomy (20.0%; \( P = .006; \text{RR 0.25; 95%CI 0.09-0.72} \)). Postintervention fever occurred less frequently in patients who received antibiotics for both the hysterectomy (16.4% vs 50.0%; \( P = .046; \text{RR 0.33; 95%CI 0.14-0.79} \)) and UAE group (3.4% vs 5.8%; \( P = .99 \)). Hematomas occurred significantly more frequently after UAE, while the hysterectomy group experienced more urinary tract infections and urinary retention. No patients in the UAE group required a blood transfusion, compared with 10 patients (13.3%) in the hysterectomy group. The minor complication rates were 22.2% (95%CI 13.7-32.8) in the UAE group and 30.7% (95%CI 20.5-42.4) in the hysterectomy group (RR 0.72; 95%CI 0.43-1.23; \( P = .23 \)). Major complications were rare and concerned 2 cases of pulmonary embolisms, 1 in each group. The major complication rate was 1.2% (95%CI 0.03-7.2) and 1.3% (95%CI 0.03-7.2) for UAE and hysterectomy respectively (RR 0.93; 95%CI 0.06-14.54; \( P = .99 \)). Both minor and major complication rates did not differ significantly between the 2 groups.

### Follow-up

Table V describes the unscheduled visits within the first 6 weeks after discharge. In the UAE group, 30 patients (37.0% with a total of 46 visits) consulted a physician, mainly for pain and/or fever. In the hysterectomy group, 19 patients (25.3% with a total of 24 visits) consulted a physician after discharge for various reasons. This difference was not significant (RR 1.45; 95%CI 0.90-2.37; \( P = .12 \)).

Readmissions (Table VI) were significantly more common in the UAE group: 9 patients versus 0 patients in the hysterectomy group (\( P = .0032 \)). In the UAE

<table>
<thead>
<tr>
<th>Table I</th>
<th>Baseline characteristics: patient demographics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UAE (n = 88)</td>
</tr>
<tr>
<td>Age (y)</td>
<td>No. (%)</td>
</tr>
<tr>
<td>&lt; 35</td>
<td>1 (1.1)</td>
</tr>
<tr>
<td>35-40</td>
<td>17 (19.3)</td>
</tr>
<tr>
<td>40-45</td>
<td>28 (31.8)</td>
</tr>
<tr>
<td>45-50</td>
<td>33 (37.5)</td>
</tr>
<tr>
<td>&gt; 50</td>
<td>9 (10.2)</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>44.6 (4.8)</td>
</tr>
<tr>
<td>Body mass index (weight [kg]/length [m²])</td>
<td></td>
</tr>
<tr>
<td>&lt;18.5</td>
<td>2 (2.3)</td>
</tr>
<tr>
<td>18.5-24.9</td>
<td>33 (37.5)</td>
</tr>
<tr>
<td>25-29.9</td>
<td>32 (36.4)</td>
</tr>
<tr>
<td>&gt;30</td>
<td>21 (23.9)</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>26.7 (5.6)</td>
</tr>
<tr>
<td>Parity</td>
<td>No. (%)</td>
</tr>
<tr>
<td>0</td>
<td>30 (34.1)</td>
</tr>
<tr>
<td>≥1</td>
<td>58 (65.9)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>24 (27.3)</td>
</tr>
<tr>
<td>White</td>
<td>54 (61.4)</td>
</tr>
<tr>
<td>Other</td>
<td>10 (11.4)</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>16 (18.2)</td>
</tr>
<tr>
<td>Married</td>
<td>55 (62.5)</td>
</tr>
<tr>
<td>Living apart together</td>
<td></td>
</tr>
<tr>
<td>5 (5.7)</td>
<td>4 (4.5)</td>
</tr>
<tr>
<td>Divorced</td>
<td>12 (13.6)</td>
</tr>
<tr>
<td>Widow</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Employment status</td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>68 (77.3)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>20 (22.7)</td>
</tr>
<tr>
<td>Smoking status</td>
<td></td>
</tr>
<tr>
<td>Current smoker</td>
<td>21 (23.9)</td>
</tr>
<tr>
<td>Former smoker</td>
<td>11 (12.5)</td>
</tr>
<tr>
<td>Nonsmoker</td>
<td>56 (63.6)</td>
</tr>
<tr>
<td>Highest educational level</td>
<td></td>
</tr>
<tr>
<td>Elementary school</td>
<td>3 (3.4)</td>
</tr>
<tr>
<td>Lower vocational, lower secondary school</td>
<td>29 (33.0)</td>
</tr>
<tr>
<td>Intermediate and higher vocational, higher secondary school</td>
<td>26 (29.5)</td>
</tr>
<tr>
<td>College/University</td>
<td>28 (31.8)</td>
</tr>
<tr>
<td>Other</td>
<td>2 (2.3)</td>
</tr>
</tbody>
</table>

Data were available for all or all but 1 patient, unless stated otherwise. Logistic regression analysis did not reveal baseline characteristics that could predict randomization outcome. * Missing: 2
group, 7 of the 9 (77.8%) readmissions occurred within the first week after discharge from the hospital. Patients were readmitted for pain (22.2%), fever (22.2%), or a combination of both (44.4%). One patient (11.1%) was readmitted for expulsion of a necrotic fibroid. Hystero-
scopic removal was attempted, but failed because of
cervical dilation which interfered with uterine dilatation. Antibiotics were administered intravenously and the

<table>
<thead>
<tr>
<th>Table II</th>
<th>Baseline characteristics: symptoms, previous treatment and uterus/fibroid characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UAE (n = 88)</td>
</tr>
<tr>
<td>Previous treatment</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>11 (12.5)</td>
</tr>
<tr>
<td>Hormonal</td>
<td>59 (67.0)</td>
</tr>
<tr>
<td>Nonsteroidal anti-inflammatory drugs/tranexaminac</td>
<td>45 (51.1)</td>
</tr>
<tr>
<td>Iron-supplement/blood transfusion</td>
<td>50 (56.8)</td>
</tr>
<tr>
<td>Surgical procedures*</td>
<td></td>
</tr>
<tr>
<td>Hysteroscopic myomectomy</td>
<td>6 (6.8)</td>
</tr>
<tr>
<td>Laparoscopic myomectomy</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Laparotomic myomectomy</td>
<td>7 (8.0)</td>
</tr>
<tr>
<td>Hysteroscopic endometrium resection</td>
<td>3 (3.4)</td>
</tr>
<tr>
<td>Curettage</td>
<td>3 (3.4)</td>
</tr>
<tr>
<td>Symptoms</td>
<td></td>
</tr>
<tr>
<td>Menorrhagia</td>
<td>88 (100)</td>
</tr>
<tr>
<td>Dysmenorrhea</td>
<td>47 (53.4)</td>
</tr>
<tr>
<td>Pain (not during menstruation)</td>
<td>15 (17.0)</td>
</tr>
<tr>
<td>Urinary symptoms</td>
<td>13 (14.8)</td>
</tr>
<tr>
<td>Defecation problems</td>
<td>5 (5.7)</td>
</tr>
<tr>
<td>Anemia</td>
<td>43 (48.9)</td>
</tr>
<tr>
<td>Pressure symptoms</td>
<td>23 (26.1)</td>
</tr>
<tr>
<td>Other symptoms</td>
<td>6 (6.8)</td>
</tr>
<tr>
<td>Duration of symptoms (m)</td>
<td></td>
</tr>
<tr>
<td>Median (range)</td>
<td>24 (3-250)</td>
</tr>
<tr>
<td>Duration of menstruation (d)</td>
<td></td>
</tr>
<tr>
<td>Total days (median, range)</td>
<td>7 (4-28)</td>
</tr>
<tr>
<td>Heavy days (median, range)</td>
<td>3 (1-28)</td>
</tr>
<tr>
<td>Number of fibroids*</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>35 (39.8)</td>
</tr>
<tr>
<td>2</td>
<td>13 (14.8)</td>
</tr>
<tr>
<td>3</td>
<td>17 (19.3)</td>
</tr>
<tr>
<td>&gt; 3</td>
<td>18 (20.5)</td>
</tr>
<tr>
<td>Median (range)</td>
<td>2 (1-20)</td>
</tr>
<tr>
<td>Uterine volume (cm³)*</td>
<td></td>
</tr>
<tr>
<td>0-250</td>
<td>33 (37.9)</td>
</tr>
<tr>
<td>251-500</td>
<td>26 (29.9)</td>
</tr>
<tr>
<td>501-1000</td>
<td>19 (21.8)</td>
</tr>
<tr>
<td>&gt; 1000</td>
<td>9 (10.3)</td>
</tr>
<tr>
<td>Median (range)</td>
<td>321 (31-3005)</td>
</tr>
<tr>
<td>Fibroid volume (dominant fibroid, cm³)*</td>
<td></td>
</tr>
<tr>
<td>0-100</td>
<td>55 (63.2)</td>
</tr>
<tr>
<td>101-200</td>
<td>14 (16.1)</td>
</tr>
<tr>
<td>201-400</td>
<td>11 (12.6)</td>
</tr>
<tr>
<td>&gt; 400</td>
<td>7 (8.9)</td>
</tr>
<tr>
<td>Median (range)</td>
<td>59 (1-673)</td>
</tr>
</tbody>
</table>

Number of fibroids and uterine/fibroid volume were calculated by ultrasound unless stated otherwise. Data were available for all or all but 1 patient, unless stated otherwise. Logistic regression analysis did not reveal baseline characteristics that could predict randomization outcome.

* The surgical treatments do not add up because some patients had several treatments.
1 UAE missing: 5, hysterectomy missing: 11.
2 UAE missing: 1, hysterectomy missing: 9.
3 UAE missing: 1, hysterectomy missing: 11.
4 MRI measurements were used in 5 patients.
5 1 patient in the UAE group because of missing ultrasound data.
The patient stayed in the hospital until fever and pain had subsided. The mean admission time for UAE increased from 2.0 to 2.5 days (SD 2.7; range 0-16 days) as a result of readmissions, but remained significantly shorter compared with hysterectomy ($P < .001$).

Complications and symptoms between discharge from the hospital and the first routine visit at 6 weeks are shown in Table IV. UAE patients complained of vaginal discharge in 21.0% compared with 8.0% of the hysterectomy patients ($P = .022$). A percentage (14.8%) of UAE patients experienced vaginal loss of fibroid tissue. Hot flashes were present in 19.8% (UAE) and 20.0% (hysterectomy) of patients. Four cases of pain and/or fever that required readmission were classified as minor complications because the definition of major complications which we used (as described in the methods section) did not apply here.

Three patients (3.7%) in the UAE group had major complications: pneumonia in a patient with a history of recurrent pneumonia caused by asthmatic disease ($n = 1$); reintervention because of an incomplete fibroid expulsion ($n = 1$); and septicemia ($n = 1$). One patient (1.3%) in the hysterectomy group was diagnosed with a vesicovaginal fistula, which was surgically repaired beyond the 6 weeks’ follow-up period (not reported in Table VI).

The minor complication rate in the first 6 weeks after discharge was significantly higher in the UAE group than in the hysterectomy group: 58.0% (95% CI 46.5-68.9) and 40.0% (95% CI 28.9-52.0), respectively ($RR = 1.45; 95% CI 1.04-2.02; P = .024$). The major complication rate in the first 6 weeks after discharge was 3.7% (95% CI 0.8-10.4) and 1.3% (95% CI 0.03-7.2) for UAE and hysterectomy, respectively ($RR = 2.78; 95% CI 0.30-26.13; P = .62$), and did not differ significantly.

The overall minor complication rate (ie, from the procedure until the 6-week routine visit) was 64.2% (95% CI 52.8-74.6) (52 patients) in the UAE group compared with 56.0% (95% CI 44.1-67.4) (42 patients) in the hysterectomy group (RR 1.12; 95% CI 0.87-1.46; $P = .38$). The overall major complication rate was 4.9% (95% CI 1.4-12.2) (4 patients) in the UAE group compared with 2.7% (95% CI 0.3-9.3) (2 patients) in the hysterectomy group (RR 1.85; 95% CI 0.35-9.82; $P = .68$). Both findings were not statistically significant. Also, when only abdominal hysterectomies were compared with UAE, overall major and minor complication rates did not differ significantly ($P = .28$ and $P = .70$). The difference in hospitalization time remained statistically significant ($P < .001$). Radiologists’ experience with UAE was not associated with the technical failure rate. Less experienced hospitals were not associated with higher complication or readmission rates.

### Table III Procedural characteristics

<table>
<thead>
<tr>
<th>Type of UAE</th>
<th>UAE (n = 81)</th>
<th>Hysterectomy (n = 75)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target embolization*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left uterine artery</td>
<td>65</td>
<td>–</td>
</tr>
<tr>
<td>Right uterine artery</td>
<td>59</td>
<td>–</td>
</tr>
<tr>
<td>Selective embolization*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left uterine artery</td>
<td>8</td>
<td>–</td>
</tr>
<tr>
<td>Right uterine artery</td>
<td>12</td>
<td>–</td>
</tr>
<tr>
<td>Type of hysterectomy (n = 4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal hysterectomy</td>
<td>2</td>
<td>63</td>
</tr>
<tr>
<td>Pfannenstiel incision</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>Median incision</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>Vaginal hysterectomy</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Vaginal hysterectomy with morcellator</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>LH with morcellator</td>
<td>–</td>
<td>2</td>
</tr>
<tr>
<td>LAVH</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>Cervix</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservation of cervix</td>
<td>2</td>
<td>22</td>
</tr>
<tr>
<td>Other procedures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Removal of hydrosalpinx</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>Adhesiolysis</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>Salpingo-oophorectomy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unilateral</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Bilateral</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>Anesthesia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local</td>
<td>71</td>
<td>–</td>
</tr>
<tr>
<td>Epidural</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Spinal</td>
<td>1 (+1)</td>
<td>3</td>
</tr>
<tr>
<td>General anesthesia</td>
<td>2</td>
<td>52</td>
</tr>
<tr>
<td>General and epidural</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>General and spinal</td>
<td>–</td>
<td>2</td>
</tr>
<tr>
<td>Duration of procedure (min)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>79 (30.5)</td>
<td>95.4</td>
</tr>
<tr>
<td>Median (range)</td>
<td>75 (30-165); 90 (90-195);</td>
<td>90 (45-175)</td>
</tr>
<tr>
<td>Blood loss (mL)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>30.9 (23.8)</td>
<td>436.1</td>
</tr>
<tr>
<td>Median (range)</td>
<td>20 (5-150); 300 (850-3000);</td>
<td>300 (10-2500);</td>
</tr>
<tr>
<td>Antibiotics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antibiotics administered</td>
<td>29 (35.8%); 67 (4 (100.0%));</td>
<td>(89.3%);</td>
</tr>
</tbody>
</table>

**Abbreviations:** LAVH, Laparoscopic-assisted vaginal hysterectomy; LH, laparoscopic hysterectomy. Characteristics of hysterectomies performed after bilaterally failed embolizations are presented in (bold) in the UAE column.

* For successful procedures.

$^1 P = .007$, compared with hysterectomy group.

$^2 P < .001$, compared with hysterectomy group.

### Comment

Present knowledge on UAE derives from numerous uncontrolled case series and only 1 small pre-consent
Table IV  Complications until the first scheduled visit (6 weeks after the procedure)

<table>
<thead>
<tr>
<th>Complication</th>
<th>Hospital stay 6 weeks after discharge</th>
<th>6 weeks after discharge</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UAE (n = 81)</td>
<td>Hyst. (n = 75)</td>
</tr>
<tr>
<td>General</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nausea</td>
<td>52</td>
<td>42</td>
</tr>
<tr>
<td>Pain</td>
<td>72</td>
<td>71</td>
</tr>
<tr>
<td>Febrile morbidity (&gt;38.5°C)</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>Minor complications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaginal discharge</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Pain requiring readmission</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Pain/fever requiring readmission</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Fibroid expulsion not requiring reintervention</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Hematoma</td>
<td>13</td>
<td>4</td>
</tr>
<tr>
<td>Wound abscess</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Wound dehiscence</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Urinary tract infection</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Urinary retention</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Urinary incontinence</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Endometritis</td>
<td>0</td>
<td>–</td>
</tr>
<tr>
<td>Hot flashes</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Anemia requiring transfusion</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Hypertension</td>
<td>7^h</td>
<td>1</td>
</tr>
<tr>
<td>Hypotension</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>23 (in 18 patients)</td>
<td>26 (in 23 patients)</td>
</tr>
</tbody>
</table>

Major complications

<table>
<thead>
<tr>
<th>Complication</th>
<th>Relative risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pneumonia</td>
<td>1</td>
</tr>
<tr>
<td>Ileus</td>
<td>1</td>
</tr>
<tr>
<td>Thrombosis</td>
<td>0</td>
</tr>
<tr>
<td>Vesicovaginal fistula</td>
<td>0</td>
</tr>
<tr>
<td>Pulmonary embolism</td>
<td>1</td>
</tr>
<tr>
<td>Intra-abdominal infection</td>
<td>0</td>
</tr>
<tr>
<td>Sepsis</td>
<td>0</td>
</tr>
<tr>
<td>Fibroid expulsion requiring re-intervention</td>
<td>0</td>
</tr>
<tr>
<td>Death</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>1 (in 1 patient)</td>
</tr>
</tbody>
</table>

N/A, Not applicable.

- The UAE group comprises both failed and successful embolizations.
- Complications of patients with hysterectomies after failed embolizations are described for the 6 weeks after discharge after their hysterectomy procedure.
- P = .016.
- P = .022.
- P = .03.
- Occurred in a hysterectomy performed after bilaterally failed UAE.
- P = .022.
- Including 1 patient that was admitted to the medium care unit for extreme hypertension.
- Spontaneous blood clot in gluteal artery during procedure.
- Small tear of m. rectus abdominis during surgery, allergic reaction to anesthetic agent during surgery.
- Gout attack, liquor spill after epidural anesthesia.
- Headache after epidural anesthesia.
- Complication led to readmission in 1 patient.
- Readmission, attempt to remove necrotic fibroid hysteroscopically, which only partly succeeded.

Hehenkamp et al 1625
randomized trial of moderate quality. According to the National Institute of Clinical Excellence (NICE), the limitations of the available literature only allow tentative conclusions about the safety and efficacy of UAE, especially since highly selected patient inclusion and high loss to follow-up bias the results. NICE strongly recommends the initiation of randomized controlled trials, which is exactly what we did here.

We deliberately chose to compare UAE with hysterectomy, not with myomectomy, for several reasons. First, hysterectomy is the standard procedure of choice to eliminate all fibroid-related complaints. In our view, myomectomy should be preserved for those women with symptomatic uterine fibroids with a strong desire for future pregnancy. Because UAE is considered to be contraindicated for women desiring pregnancy, a randomized comparison with myomectomy might even be considered unethical at this stage. In the absence of randomized data, we judged it more ethical to perform a study at the other end of the clinical spectrum, ie, in women facing hysterectomy as the last resort for their fibroid-related complaints. Although hysterectomy is an absolute cure for menorrhagia, possible sequelae, eg, incontinence, vaginal vault prolapse, risk for premature ovarian failure, long recovery, high costs, and the desire of some patients to preserve their uterus, justify serious consideration of alternative therapies such as UAE. The procedural success rate (88.9%) was comparable to the results of the aforementioned small semi-randomized trial, but lower than the success rates reported in most other studies, thereby illustrating the necessity of randomized data collection.

The technical failure rate (5.3%) was higher than the 0.5% to 2.5% reported in large case series, but similar to the technical failure rate of 5.0% reported in the only semi-randomized trial, because of several possible reasons. First, our study is a mix of both academic and nonteaching hospitals, whereas UAE in the reported case series was mostly performed in highly specialized single-centers, decreasing the generalizability of those results. However, in our study, experience of the interventional radiologist was not associated with outcome. Moreover, one series performed a second embolization attempt after initial technical failure, which obviously improves technical success rates. Generally, results from randomized controlled trials can

<table>
<thead>
<tr>
<th>Contact</th>
<th>Symptom(s)</th>
<th>UAE (n = 81) Number of contacts</th>
<th>Hysterectomy (n = 76) Number of contacts</th>
</tr>
</thead>
<tbody>
<tr>
<td>General physician</td>
<td>Pain</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Fever</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Vaginal bleeding</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Groin hematoma</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Constipation</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Blood pressure issues</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>3*</td>
<td>4†</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>13 (in 10 patients)</td>
<td>14 (in 12 patients)</td>
</tr>
<tr>
<td>Gynecologist</td>
<td>Fever</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Fever and pain</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Fever and vaginal bleeding</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Pain</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Pain and vaginal discharge</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Pain and vaginal bleeding</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Vaginal bleeding</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Vaginal discharge</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Wound dehiscence</td>
<td>0</td>
<td>2†</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>0</td>
<td>1†</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>30 (in 22 patients)</td>
<td>10 (in 8 patients)</td>
</tr>
<tr>
<td>Lung specialist</td>
<td>Fever, dyspnea: pneumonia</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>3 (in 1 patient)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Total number of visits</td>
<td>45</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Total number of patients</td>
<td>30†</td>
<td>19</td>
</tr>
</tbody>
</table>

* Checking hemoglobin level; gout attack; sensitive breast.
† Coughing, dizzy, and constipation; stomach pain; urge incontinence complaints; vaginal itch.
‡ Same patient.
§ Severe hair loss.
†† RR 1.45 (95%CI 0.90-2.37; \( P = .12 \)), compared with number of patients in the hysterectomy group.
substantially differ from those in case series because of publication bias and patient selection criteria.23,24

Mean hospital stay was significantly shorter for UAE than for hysterectomy. Mean hospital stay for UAE was longer than in some studies, but in our experience most patients need more care.25,26 Therefore, we would not recommend performing UAE as an outpatient procedure. Surprisingly, the number of unscheduled visits was higher in the UAE group. Readmission rates after UAE within the first 6 weeks (11.1%) were higher compared with other reports (2.9%-5.0%), although the reasons for readmission were similar.17,27 The experimental status of the UAE procedure could be the reason why physicians were more inclined to see patients and readmit them more quickly. Most readmissions in our study occurred within the first week after discharge (77.8%), underlining the need for adequate follow-up during this period. None of our hysterectomy patients were readmitted within the first 6 weeks, while Pinto et al found a readmission rate of 5.0% after hysterectomy.17 Because most hysterectomy patients were still in the hospital when most readmissions in the UAE group occurred, the comparison is not completely fair: if UAE patients would stay in the hospital as long as hysterectomy patients, only 2 readmissions would have occurred.

Overall major and minor complication rates in both groups were comparable, but minor complications in the period between discharge and the first 6 weeks’ visit were significantly higher in the UAE group. Our study, therefore, cannot support the suggestion made by others28 that UAE has a lower complication rate than hysterectomy, again stressing the need for randomized studies. A detailed comparison of complication rates with other studies is hampered by the fact that various studies apply different classification systems for reporting complications. Major complications were rare in our study. Although many series reported emergency hysterectomy rates up to 1.3% within the first weeks after UAE, no such procedures occurred in our patients.3,4,21,27

There are several limitations to our study. First, 21 (11.9%) patients withdrew from the study after randomization before treatment. Their baseline characteristics, however, did not differ from those being treated. Second, given the low major complication rates, our study size was too small to detect any difference in major complication rates, and definite conclusions, therefore, cannot be drawn. In contrast, we did find differences in minor complication rates and length of hospital stay, so lack of power was not an issue here. We used no objective criteria for menorrhagia but relied on subjective appreciations of our patients. By doing so, the generalizability of our findings is probably enhanced: included patients represent those seen in daily practice where the decision to perform a hysterectomy is not based on objective measurements (eg, pictorial charts) either. We could not find any differences in major complication rates between UAE and hysterectomy. Unsuccessful UAE procedures, however, seem to occur more often than previously reported. Hospital stay is significantly shorter for UAE. The higher minor complication rate after discharge in the UAE group, as well as the readmission rates and unscheduled visits,
emphasize the necessity for careful follow-up and clear instructions to the patient. Although the study results are supportive for UAE, the question as to whether UAE is a good alternative for hysterectomy depends on the balance of efficacy, costs, and quality of life, and still remains to be answered.

Acknowledgments

We would like to thank all participating patients, EMMY-trial group members, nurses, and other contributors who made the trial possible. The EMMY-trial participants and hospitals: J. Reekers, W. Ankum, M. Burger, G. Bonsel, Erwin Birnie, W. Hekenkamp, N. Volkers (Academic Medical Center, Amsterdam); S. de Blok, C. de Vries (Onze Lieve Vrouwe Gasthuis, Amsterdam); T. Salemans, G. Veldhuyzen van Zanten (Atrium Medical Centre, Heerlen); D. Tinga, T. Prins (Groningen University Hospital, Groningen); P. Sleijffers, M. Rutten (Bosch Medical Centre, Den Bosch); M. Smeets, N. Aarts (Bronovo Hospital, The Hague); P. van der Moer, D. Vroegindeweij (Medical Centre Rijnmond-Zuid, Rotterdam); F. Boekkooi, L. Lampmann (St. Elisabeth Hospital, Tilburg); G. Klei- verda, (Flev Hospital, Almere); R. Dik, J. Marsman (Gooi-Noord Hospital, Laren); C. de Nooijer, I. Hendriks, G. Guit (Kennemer Gasthuis, Haarlem); H. Ottervanger, H. van Overhagen, (Leyenburg Hospital, The Hague); A. Thurkow (St. Lucas/Andreas Hospital, Amsterdam); P. Donderwinkel, C. Holt (Martini Hospital, Groningen); A. Adriaanse, J. Wallis, (Medical Center Alkmaar, Alkmaar); J. Hirdes, J. Schutte, W. de Rhoet, (Medical Center Leeuwarden, Leeuwarden); P. Paaymans, R. Schepers-Bok (Hospital Midden-Twente, Hengelo); G. van Doorn, H. Franke, J. Krabbe, A. Huisman, (Medisch Spectrum Twente, Enschede); M. Hermans, R. Dallinga (Reinier de Graaf Gasthuis, Delft); F. Reijnouds, J. Spithoven, (Slingeland Hospital, Doetichem); W. Jager, P. Veekmans, (St. Hans Gasthuis, Weert); P. van der Heijden, M. Veerschild, J. van den Hout, (Twenteborg Hospital, Almelo); I. van Seumeren, A. Heinz, R. Lo, W. Mali (University Hospital Utrecht, Utrecht); J. Lind, Th. de Rooy (Westende Hospital, The Hague); M. Bulstra, F. Sanders (Diakonessenhuis Utrecht, Utrecht); J. Doornbos (De Heel Hospital, Zaandam); P. Dijkhuizen, M. van Kints (Rijnstate Hospital, Arnhem); Ph. Engelen, R. Heijboer (Slotervaart Hospital, Amsterdam).

References


Objective: Ovarian cancer screening with transvaginal ultrasound (TVU) and CA-125 was evaluated in the Prostate, Lung, Colorectal and Ovarian (PLCO) Trial.

Study design: This was a randomized controlled trial of screening versus usual care. Baseline screening results are reported.

Results: Of 39,115 women randomized to receive screening, 28,816 received at least 1 test. Abnormal TVU was found in 1338 (4.7%), and abnormal CA-125 in 402 (1.4%). Twenty-nine neoplasms were identified (26 ovarian, 2 fallopian, and 1 primary peritoneal neoplasm). Nine were tumors of low malignant potential and 20 were invasive. The positive predictive value for invasive cancer was 3.7% for an abnormal CA-125, 1.0% for an abnormal TVU, and 23.5% if both tests were abnormal.

Conclusion: The effect of screening on ovarian cancer mortality in the PLCO cohort has yet to be evaluated and will require longer follow-up. Screening identified both early- and late-stage neoplasms, and the predictive value of both tests was relatively low.

© 2005 Mosby, Inc. All rights reserved.
Advances in the treatment of ovarian cancer in the past decade have led to incremental improvements in survival. For example, it is now known that initial surgery by an experienced gynecologic oncologist increases the likelihood of optimal debulking and is associated with an improved long-term survival rate. The introduction of platinum-based, multiagent chemotherapy regimens, the more recent addition of the taxanes into clinical practice, the widespread use of salvage chemotherapy for recurrent cancer, and the increasing use of chemotherapy in less advanced stage disease have all contributed to improved survival. Despite these advances, ovarian cancer remains a fatal disease for most women in whom it is diagnosed, and it has the highest mortality rate of all the gynecologic malignancies. Ovarian cancer is diagnosed in 25,400 women and results in 14,300 deaths annually in the US. Most cases present at an advanced stage, and long-term survival is achieved in less than a third of patients. However, early stage ovarian cancer has a much higher survival rate, and it is possible that early detection through screening could significantly reduce mortality.

Screening for early disease is useful if the disease in question has a presymptomatic stage in which treatment is more effective than treatment administered for symptomatic disease. Effective screening also requires the availability of screening procedures acceptable to patients that can be performed at a reasonable cost. Finally, screening tests must have a sensitivity high enough to detect a significant fraction of all existing cases in the population and, at the same time, a sufficiently high specificity to avoid generating an excessive number of false-positive screens. False-positive test results are a particular problem in diseases with low prevalence in the target population and in diseases for which further evaluation of an abnormal screen often includes an invasive surgical procedure. Both concerns are true for ovarian cancer. One method to address the problem of balancing sensitivity and specificity is to utilize more than 1 screening test in combination, either in parallel or sequentially.

Bimanual palpation of the ovaries is widely used to detect ovarian pathology, but it is insensitive for detection of early stage cancer, and its widespread application has not resulted in a significant shift to earlier-stage ovarian cancer. Both CA-125 and ultrasound, particularly transvaginal ultrasound (TVU), have been advocated as potential ovarian cancer screening modalities. TVU screening has generally been associated with a high rate of false-positive screens, leading to a large number of surgical procedures that did not identify cancer. CA-125 is reported to be elevated in about half of women with early-stage ovarian cancer, but it is also increased in other cancers and in various nonmalignant conditions such as liver disease and heart failure. Studies to date have not clarified the efficacy of TVU and CA-125, performed either separately or together, for ovarian cancer screening. In this report, we describe the characteristics of the 39,115 women randomized to the screened arm of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial and report the results of the initial (baseline) ovarian cancer screening examinations in 28,816 women. TVU and CA-125 were performed at study entry and were administered concurrently.

**Material and methods**

**Study design**

The design of the PLCO Trial has been described in detail elsewhere. Briefly, the objectives are to determine in healthy subjects aged 55-74 at entry whether: 1) screening with flexible sigmoidoscopy can reduce mortality from colorectal cancer in males and females; 2) screening with chest x-ray can reduce mortality from lung cancer in males and females; 3) screening with digital rectal examination plus serum prostate specific antigen (PSA) can reduce mortality from prostate cancer; and 4) screening with CA-125 and transvaginal ultrasound can reduce mortality from ovarian cancer.

The study is a 2-armed trial in which half of subjects were randomized to receive screening, and half to usual care. Enrollment was initiated in the fall of 1993 and completed in the summer of 2001. Ten screening centers are participating: the University of Colorado Health Sciences Center; Lombardi Cancer Research Center of Georgetown University; Pacific Health Research Institute, Honolulu; Henry Ford Health System; University of Minnesota School of Public Health/Virginia L. Piper Cancer Institute; Washington University School of Medicine; University of Pittsburgh, Pittsburgh Cancer Institute and Magee-Women’s Hospital; University of Utah Health Sciences Center; Marshfield Clinic Research Foundation; and the University of Alabama at Birmingham. Each institution obtained local Institutional Review Board approval to carry out the study. CA-125 and PSA testing is performed centrally at the University of California, Los Angeles Immunogenetics Center. A biorepository for the collection and storage of blood samples and tissue is an integral component of the trial. Participants will be followed for at least 13 years from entry.

Women in the intervention arm are screened for ovarian cancer using CA-125 annually for 6 years and TVU annually for 4 years. Both studies are performed concurrently at entry into the study. Baseline ovarian cancer screening tests were performed on the first randomized subject on November 15, 1993 and on the last subject on December 13, 2001. The screening protocol originally included bimanual physical exam of the ovaries as one of the screening modalities. This
procedure was dropped in 1998 after review of the data determined that no ovarian cancers had been detected with this modality alone. This experience is consistent with previous reports suggesting that pelvic examination is not a satisfactory screening tool for ovarian cancer. In addition, more than two thirds of all women entering the study had routine bimanual pelvic examinations as part of their ongoing medical care, thus compromising our ability to compare the impact of bimanual examination on ovarian cancer mortality between the screened and unscreened arms.

**Eligibility**

The target population for the study included subjects from 55 to 74 years of age who had not been diagnosed previously with prostate, lung, colorectal, or ovarian cancer. Criteria for exclusion included current treatment for cancer other than basal cell and squamous cell skin cancer, and enrollment in another cancer screening or prevention trial. Beginning on April 15, 1995, individuals who had received a colonoscopy, sigmoidoscopy, or barium enema in the past 3 years were excluded. Women on tamoxifen were initially excluded because, at the time of study initiation, the relationship between tamoxifen and ovarian cancer risk was unknown. This restriction was lifted in April 1999 when it became clear that this was a null association. Individuals with previous surgical removal of 1 lung or the entire colon were also excluded. Initially, women who had undergone oophorectomy were ineligible but in 1996 this restriction was lifted because low accrual of women threatened to jeopardize screening end points for lung and colon cancer. Women who had undergone oophorectomy were not offered either ovarian cancer screening test.

**Screening procedures**

TVU was performed by qualified sonographers using a 5-7.5 MHz transvaginal probe. The examiner imaged both ovaries in the transverse and longitudinal planes. At least 5 minutes were spent looking for each ovary to ensure an adequate search; however, if the iliac vessels were visualized without ovaries being seen the examiner concluded the search for the ovaries. Ovaries were measured along the major and minor axes in both transverse and longitudinal planes, and the prolate ellipsoid formula (width × height × thickness × 0.523) was used to calculate the volume of each ovary and/or cyst. The following TVU test results were classified as abnormal (positive): ovarian volume > 10 cm³; cyst volume > 10 cm³; any solid area or papillary projection extending into the cavity of a cystic ovarian tumor of any size; or any mixed (solid/cystic) component within a cystic ovarian tumor. CA-125 was measured on serum obtained and frozen within 2 hours of blood draw at each of the 10 screening centers, then shipped to the UCLA Immunogenetics Center on dry ice. Samples were stored at −70°C and aliquots thawed for assay. Assays were performed according to instructions from the manufacturer (Centocor, Inc, Malvern, PA). Samples were run in duplicate. Any sample with a result over 35 U/mL was reanalyzed to verify the value. Samples showing discrepant results between duplicate results (coefficient of variation over 10%) were reanalyzed. Quality assurance for the measurement of CA-125 was done in accordance with the manufacturer’s suggested protocol using manufacturer-supplied samples, as well as additional control samples obtained from Bio-Rad (Hercules, CA). The assay precision is represented by its coefficient of variation (CV). The CVs (and 95% CIs) were found to be 4.07% (3.92-4.22) at the lower concentration of 52.7 U/mL and 3.78% (3.64-3.92) at the higher concentration of 106.5 U/mL. These results are in good agreement with those reported by the manufacturers on the product inserts.

The original Centocor CA-125 radioimmunoassay (RIA) assay was replaced with the Centocor CA-125II RIA assay on October 1, 1995. All samples tested using the original CA-125 assay were retested using the CA-125II assay. Of 5371 samples analyzed with both tests, the mean value rose from 9.4 to 13.1 U/mL. One hundred and twenty-two samples changed screening results, with 109 converting from negative to positive with the new assay (2.0%). Three of these subjects were diagnosed with cancer after the baseline screen; all 3 had an abnormal TVU. Four of the other 106 subjects also had an abnormal TVU but did not have ovarian cancer. Thirteen samples converted from positive to negative (0.24%). None of these subjects were diagnosed with ovarian cancer. The original CA-125 had a positivity rate of 0.6%, and the CA-125II assay had a positivity rate of 2.4% in these 5371 samples.

CA-125 results ≥ 35 U/mL were classified as abnormal (positive).

Results of both screening tests were sent to participants and their personal physicians within 3 weeks of specimen submission. Evaluation and follow-up of women with an abnormal screening test were at the discretion of the participant’s physician.

**Follow-up of abnormal screening tests**

Medical records of all procedures done to evaluate an abnormal screen were obtained by study personnel and recorded on standardized reporting forms. Pathology reports from all ovarian neoplasms were abstracted by trained certified tumor registrars and were reviewed by one of the authors (E.P.). Neoplasms with an ICD-O-2 behavior code of 3 (malignant neoplasms) and diagnosed in the screened arm within 12 months of a positive baseline screen are included in this report. (The NCI’s Surveillance, Epidemiology and End Results Program
Target Population: Men and women ages 55-74 with no personal history of prostate, lung, colorectal or ovarian cancer

154,942 Randomized

78,237 Female

39,122 Control Group

5,386 Did not receive either screen

4,913 Prior oophorectomy *

28,803 Received initial CA125
28,519 Received initial TVU
28,816 Received at least one test
28,506 Received both tests

Figure  Flow of participants into the PLCO Trial. *Ineligible for screening.

considers those neoplasms with a behavior of either 2 or 3 to be reportable as cancer.) Ovarian tumors occurring among women in the control group are not reported here. The purpose of this paper is to report on the performance characteristics of CA-125 and TVU in an initial screen.

Results

Enrollment

Enrollment into the PLCO Trial is shown in the Figure. Of 78,237 females enrolling into the trial, 39,115 were randomized to the intervention arm. Of these, 4913 women reported previous oophorectomy and were not eligible for ovarian screening, and an additional 5386 did not receive either screen. Of the 5386 who had neither screen, 3 were diagnosed with ovarian cancer before receiving screening; 17 died before any screening; 105 withdrew from the study; 5232 refused all T0 PLCO screening procedures; and 29 underwent other PLCO T0 procedures but refused ovarian cancer screening. Thirteen underwent TVU but not phlebotomy, and 297 had CA-125 measured but no TVU. Thus, a total of 28,816 women received at least 1 screening test and 28,506 received both. This report describes the characteristics of the 39,115 women randomized to the intervention arm, and summarizes the results in 28,803 women who received the initial CA-125 measurement and 28,519 who received the initial TVU.

Baseline characteristics

At study entry, subjects completed a baseline questionnaire that included age, race, educational level, and specific risk factors for ovarian cancer, including previous
gynecologic surgery, use of oral contraceptives, parity, personal history of breast cancer, and family history of breast and ovarian cancer in first-degree relatives. Data from the baseline questionnaire are shown for the 39,115 women randomized to the intervention arm (Table I).

Table I is divided into 3 groups: those who received any

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Screened</th>
<th></th>
<th>Not screened</th>
<th></th>
<th>Previous oophorectomy</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>%</td>
<td>Number</td>
<td>%</td>
<td>Number</td>
<td>%</td>
</tr>
<tr>
<td>Age group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55–59</td>
<td>9974</td>
<td>34.6</td>
<td>1698</td>
<td>31.5</td>
<td>1787</td>
<td>36.4</td>
</tr>
<tr>
<td>60–64</td>
<td>8871</td>
<td>30.8</td>
<td>1533</td>
<td>28.5</td>
<td>1370</td>
<td>27.9</td>
</tr>
<tr>
<td>65–69</td>
<td>6242</td>
<td>21.7</td>
<td>1251</td>
<td>23.2</td>
<td>1095</td>
<td>22.3</td>
</tr>
<tr>
<td>70–74</td>
<td>3729</td>
<td>12.9</td>
<td>904</td>
<td>16.8</td>
<td>661</td>
<td>13.5</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>25,618</td>
<td>88.9</td>
<td>3868</td>
<td>71.8</td>
<td>4342</td>
<td>88.4</td>
</tr>
<tr>
<td>Black</td>
<td>1533</td>
<td>5.3</td>
<td>356</td>
<td>6.6</td>
<td>281</td>
<td>5.7</td>
</tr>
<tr>
<td>Hispanic</td>
<td>417</td>
<td>1.4</td>
<td>93</td>
<td>1.7</td>
<td>95</td>
<td>1.9</td>
</tr>
<tr>
<td>Asian</td>
<td>984</td>
<td>3.4</td>
<td>159</td>
<td>3.0</td>
<td>116</td>
<td>2.4</td>
</tr>
<tr>
<td>Other</td>
<td>200</td>
<td>0.7</td>
<td>51</td>
<td>0.9</td>
<td>32</td>
<td>0.7</td>
</tr>
<tr>
<td>Missing response</td>
<td>64</td>
<td>0.2</td>
<td>885</td>
<td>15.9</td>
<td>47</td>
<td>1.0</td>
</tr>
<tr>
<td>Education level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;High school</td>
<td>1685</td>
<td>5.8</td>
<td>483</td>
<td>9.0</td>
<td>352</td>
<td>7.2</td>
</tr>
<tr>
<td>12 y/completed high school</td>
<td>11,427</td>
<td>39.7</td>
<td>1866</td>
<td>34.6</td>
<td>2006</td>
<td>40.8</td>
</tr>
<tr>
<td>Some college</td>
<td>6587</td>
<td>22.9</td>
<td>1086</td>
<td>20.2</td>
<td>1238</td>
<td>25.2</td>
</tr>
<tr>
<td>College graduate</td>
<td>4624</td>
<td>16.0</td>
<td>545</td>
<td>10.1</td>
<td>659</td>
<td>13.4</td>
</tr>
<tr>
<td>Postgraduate</td>
<td>4405</td>
<td>15.3</td>
<td>523</td>
<td>9.7</td>
<td>609</td>
<td>12.4</td>
</tr>
<tr>
<td>Missing response</td>
<td>88</td>
<td>0.3</td>
<td>885</td>
<td>16.4</td>
<td>49</td>
<td>1.0</td>
</tr>
<tr>
<td>Had previous pelvic surgery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bilateral oophorectomy</td>
<td>2</td>
<td>0.0</td>
<td>3</td>
<td>0.1</td>
<td>59</td>
<td>1.2</td>
</tr>
<tr>
<td>Hysterectomy</td>
<td>7768</td>
<td>27.0</td>
<td>1098</td>
<td>20.4</td>
<td>199</td>
<td>4.1</td>
</tr>
<tr>
<td>Bilateral oophorectomy and hysterectomy</td>
<td>83</td>
<td>0.3</td>
<td>105</td>
<td>1.9</td>
<td>4601</td>
<td>93.6</td>
</tr>
<tr>
<td>Neither</td>
<td>20,811</td>
<td>72.2</td>
<td>3295</td>
<td>61.2</td>
<td>8</td>
<td>0.2</td>
</tr>
<tr>
<td>Missing response</td>
<td>152</td>
<td>0.5</td>
<td>885</td>
<td>16.4</td>
<td>46</td>
<td>0.9</td>
</tr>
<tr>
<td>Ever taken oral contraceptives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>13,168</td>
<td>45.7</td>
<td>2327</td>
<td>43.2</td>
<td>2127</td>
<td>43.3</td>
</tr>
<tr>
<td>Yes</td>
<td>15,554</td>
<td>54.0</td>
<td>2175</td>
<td>40.4</td>
<td>2736</td>
<td>55.7</td>
</tr>
<tr>
<td>Missing response</td>
<td>94</td>
<td>0.3</td>
<td>884</td>
<td>16.4</td>
<td>50</td>
<td>1.0</td>
</tr>
<tr>
<td>Number of times pregnant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None has</td>
<td>2084</td>
<td>7.2</td>
<td>384</td>
<td>7.1</td>
<td>371</td>
<td>7.6</td>
</tr>
<tr>
<td>1</td>
<td>1640</td>
<td>5.7</td>
<td>253</td>
<td>4.7</td>
<td>309</td>
<td>6.3</td>
</tr>
<tr>
<td>2</td>
<td>5091</td>
<td>17.7</td>
<td>782</td>
<td>14.5</td>
<td>887</td>
<td>18.1</td>
</tr>
<tr>
<td>3-4</td>
<td>11,729</td>
<td>40.7</td>
<td>1738</td>
<td>32.3</td>
<td>1988</td>
<td>40.5</td>
</tr>
<tr>
<td>5-9</td>
<td>7625</td>
<td>26.5</td>
<td>1244</td>
<td>23.1</td>
<td>1232</td>
<td>25.1</td>
</tr>
<tr>
<td>10 or more</td>
<td>539</td>
<td>1.9</td>
<td>114</td>
<td>2.1</td>
<td>72</td>
<td>1.5</td>
</tr>
<tr>
<td>Missing response</td>
<td>108</td>
<td>0.4</td>
<td>871</td>
<td>16.2</td>
<td>54</td>
<td>1.1</td>
</tr>
<tr>
<td>Personal history of breast cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>27,718</td>
<td>96.2</td>
<td>4381</td>
<td>81.3</td>
<td>4718</td>
<td>96.0</td>
</tr>
<tr>
<td>Yes</td>
<td>1040</td>
<td>3.6</td>
<td>148</td>
<td>2.7</td>
<td>150</td>
<td>3.1</td>
</tr>
<tr>
<td>Missing response</td>
<td>58</td>
<td>0.2</td>
<td>857</td>
<td>15.9</td>
<td>45</td>
<td>0.9</td>
</tr>
<tr>
<td>Family history of breast and ovarian cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neither</td>
<td>23,517</td>
<td>81.6</td>
<td>3732</td>
<td>69.3</td>
<td>3964</td>
<td>80.7</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>3623</td>
<td>12.6</td>
<td>547</td>
<td>10.2</td>
<td>629</td>
<td>12.8</td>
</tr>
<tr>
<td>Ovarian cancer</td>
<td>948</td>
<td>3.3</td>
<td>128</td>
<td>2.4</td>
<td>175</td>
<td>3.6</td>
</tr>
<tr>
<td>Breast and ovarian cancer</td>
<td>120</td>
<td>0.4</td>
<td>13</td>
<td>0.2</td>
<td>30</td>
<td>0.6</td>
</tr>
<tr>
<td>Missing response</td>
<td>608</td>
<td>2.1</td>
<td>966</td>
<td>17.9</td>
<td>115</td>
<td>2.3</td>
</tr>
<tr>
<td>Had baseline questionnaire</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>54</td>
<td>0.2</td>
<td>855</td>
<td>15.9</td>
<td>45</td>
<td>0.9</td>
</tr>
<tr>
<td>Yes</td>
<td>28,762</td>
<td>99.8</td>
<td>4531</td>
<td>84.1</td>
<td>4868</td>
<td>99.1</td>
</tr>
<tr>
<td>Total</td>
<td>28,816</td>
<td>100.0</td>
<td>5386</td>
<td>100.0</td>
<td>4913</td>
<td>100.0</td>
</tr>
</tbody>
</table>
baseline ovarian cancer screening; those who did not receive either baseline ovarian cancer screen; and those who were ineligible for screening because of previous oophorectomy. Women provided information about previous oophorectomy both as part of the baseline questionnaire (“had previous pelvic surgery” on Table I) and at the time ovarian cancer screening was scheduled. Women who at that time reported that they were not eligible for ovarian cancer screening because of previous oophorectomy (“previous oophorectomy” on Table I) did not receive ovarian cancer screening regardless of their report on the baseline questionnaire. For this reason there are some inconsistencies in Table I. Women in the “not screened” group were on average older than those in the other 2 groups and were less educated. They also had a higher rate of noncompliance with the baseline questionnaire (15.9%) than women in the screened (0.2%) or oophorectomy (0.9%) groups. Enrollment declined with advancing age, with subjects aged 55 to 59, 60 to 64, 65 to 69, and 70 to 74 years comprising 34.4%, 30.1%, 22.0%, and 13.5% of participants, respectively. Subjects were primarily white (86.5%) and educated (over 50% reported having at least some college education). Although some women did not complete the baseline questionnaire or did not answer questions about gynecologic surgery, data were collected from over 97% of the participants.

### Compliance

Compliance rates for the baseline screening procedures were 83.4% for TVU and 84.2% for CA-125. The rate of compliance with each screening test declined slightly with advancing age; among subjects 55 to 59 and 70 to 74 years of age, compliance rates were 84.5% and 79.5%, respectively, for TVU, and 85.5% and 80.5%, respectively, for CA-125. Women who did not have a screen because of previous oophorectomy were excluded from the compliance calculations.

### Screening results

The baseline TVU was abnormal in 1338 (4.7%) of 28,519 baseline examinations. As previously reported, the ability of the examiner to visualize the ovaries by TVU decreased with advancing age of the participant, but age otherwise had no effect on TVU results (data not shown). A small number of examinations (1.9%) were classified as inadequate for interpretation.

Abnormal values for CA-125 were found in 402 (1.4%) of 28,803 baseline measurements. Age had a minimal effect on CA-125 concentration. Three fourths of the CA-125 values in each age group fell between 5 and 15 U/mL. A small number of samples (0.3%) contained insufficient serum to perform the assay.

At the individual level, there was very little overlap in abnormal results for the 2 screening tests (Table II). Among 28,506 women with results for both tests, 1703 had at least 1 abnormal test; 1338 had an abnormal TVU, 399 had an abnormal CA-125, and only 34 (2% of those with an abnormal screen) had abnormalities in both.

As a result of the baseline screening examination and subsequent follow-up procedures, 29 malignant neoplasms were identified (Table III). Twenty-six arose in the ovary, 9 of which were tumors of low malignant potential, 1 was an ovarian stromal tumor (granulosa cell), and 16 were invasive epithelial ovarian cancers. Three additional epithelial neoplasms were identified, 2 arising in the fallopian tube and 1 in the peritoneum. All 29 are designated “ovarian neoplasms” for subsequent analyses.

### Table II

<table>
<thead>
<tr>
<th>CA-125</th>
<th>TVU</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>34</td>
<td>365</td>
</tr>
<tr>
<td>−</td>
<td>1304</td>
<td>26,803</td>
</tr>
<tr>
<td>Total</td>
<td>1338</td>
<td>27,168</td>
</tr>
</tbody>
</table>

Only women having results for both TVU and CA-125 appear in this table. Three hundred and ten subjects were excluded from table because they had only 1 of the screening tests.

### Table III

<table>
<thead>
<tr>
<th>Neoplasm type</th>
<th>CA-125</th>
<th>TVU</th>
<th>#</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Malignant Potential</td>
<td>−</td>
<td>+</td>
<td>8</td>
<td>27.6</td>
</tr>
<tr>
<td>Granulosa</td>
<td>−</td>
<td>+</td>
<td>1</td>
<td>3.4</td>
</tr>
<tr>
<td>Invasive Ovarian</td>
<td>+</td>
<td>+</td>
<td>7</td>
<td>24.1</td>
</tr>
<tr>
<td></td>
<td>−</td>
<td>+</td>
<td>4</td>
<td>13.8</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>−</td>
<td>5</td>
<td>17.2</td>
</tr>
<tr>
<td>Fallopian Tube</td>
<td>+</td>
<td>−</td>
<td>1</td>
<td>3.4</td>
</tr>
<tr>
<td>Peritoneal</td>
<td>+</td>
<td>−</td>
<td>1</td>
<td>3.4</td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td></td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

Table III shows the relationship between CA-125 and TVU results and ovarian neoplasms. Invasive cancers and tumors of low malignant potential are shown separately. CA-125 alone was abnormal in 7 (24%), and TVU alone was abnormal in 13 (45%). Both tests were abnormal in 9 subjects (31%).

An additional 25 subjects had abnormalities in both TVU and CA-125 during the initial examination. Seventeen of these had surgery with benign findings, and the
who did not have surgery were not diagnosed with cancer during the subsequent 12 months.

Table IV shows the characteristics of tumors detected during the baseline screen. The 9 tumors of low malignant potential were all stage I, as was the malignant granulosa cell tumor. The stages of the 19 invasive epithelial tumors included 1 stage I, 3 stage II, 13 stage III, and 2 stage IV.

Table IV also shows the time interval between the positive screening test and the diagnosis of cancer. The PLCO protocol allowed up to 3 weeks between the screening test and results being sent to the physician. Further evaluation and follow-up resulted in a lengthy delay in surgery for some subjects; however, only 5 of the 29 cases had a delay of over 120 days from the initial screen to surgery. These included a stage IA serous cystadenoma (221 days), a stage IA malignant granulosa tumor (316 days), a stage IIA serous cystadenocarcinoma (213 days), a stage IIIA endometrioid adenocarcinoma (157 days), and a stage IV serous cystadenocarcinoma (237 days). In each of these cases the TVU result was positive but the CA-125 was normal. This delay could have contributed to the high stage of some of the neoplasms, particularly the stage IV serous cystadenocarcinoma.

The positive predictive value (PPV) of each test, and of the 2 tests together, was calculated (Table V). For example, 28,803 women had a CA-125 test, 402 of which were abnormal (1.4%). Sixty-two women underwent biopsy (15.4% of positive screens), 16 of which (25.8% of biopsies) were diagnosed with a neoplasm, for a PPV of 4.0% (16 neoplasms in 402 positive screens). Similar calculations demonstrate a PPV of 1.6% for TVU, and 26.5% if both tests were abnormal. Table V also shows similar data excluding tumors of low malignant potential (indicated as “# invasive cancer”). With this restriction, the PPV was 3.7% for an abnormal CA-125, 1.0% for an abnormal TVU, and 23.5% if both tests were abnormal.

---

**Table IV**  Histopathologic type, histopathologic grade, and stage by screening results

<table>
<thead>
<tr>
<th>Primary site</th>
<th>Invasive</th>
<th>Histopathology</th>
<th>Grade</th>
<th>Stage</th>
<th>Screen results</th>
<th>CA-125 result</th>
<th>Days between pos screen and diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ovary</td>
<td>No</td>
<td>Serous cystadenoma</td>
<td>LMP*</td>
<td>IA</td>
<td>TVU positive</td>
<td>15</td>
<td>221</td>
</tr>
<tr>
<td>Ovary</td>
<td>No</td>
<td>Serous cystadenoma</td>
<td>LMP*</td>
<td>IA</td>
<td>CA-125 &amp; TVU +</td>
<td>51</td>
<td>79</td>
</tr>
<tr>
<td>Ovary</td>
<td>No</td>
<td>Serous cystadenoma</td>
<td>LMP*</td>
<td>IA</td>
<td>TVU positive</td>
<td>18</td>
<td>55</td>
</tr>
<tr>
<td>Ovary</td>
<td>No</td>
<td>Serous cystadenoma</td>
<td>LMP*</td>
<td>IA</td>
<td>TVU positive</td>
<td>26</td>
<td>30</td>
</tr>
<tr>
<td>Ovary</td>
<td>No</td>
<td>Serous cystadenoma</td>
<td>LMP*</td>
<td>IA</td>
<td>TVU positive</td>
<td>8</td>
<td>91</td>
</tr>
<tr>
<td>Ovary</td>
<td>No</td>
<td>Serous cystadenoma</td>
<td>LMP*</td>
<td>IC</td>
<td>TVU positive</td>
<td>15</td>
<td>39</td>
</tr>
<tr>
<td>Ovary</td>
<td>No</td>
<td>Mucinous cystadenoma</td>
<td>LMP*</td>
<td>IA</td>
<td>TVU positive</td>
<td>17</td>
<td>46</td>
</tr>
<tr>
<td>Ovary</td>
<td>No</td>
<td>Serous cystadenoma</td>
<td>LMP*</td>
<td>IC</td>
<td>TVU positive</td>
<td>28</td>
<td>23</td>
</tr>
<tr>
<td>Ovary</td>
<td>No</td>
<td>Serous cystadenoma</td>
<td>LMP*</td>
<td>IC</td>
<td>TVU positive</td>
<td>26</td>
<td>80</td>
</tr>
<tr>
<td>Ovary</td>
<td>Yes</td>
<td>Malignant granulosa</td>
<td>Grade III</td>
<td>IA</td>
<td>TVU positive</td>
<td>8</td>
<td>316</td>
</tr>
<tr>
<td>Ovary</td>
<td>Yes</td>
<td>Serous cystadenocarcinoma</td>
<td>Grade III</td>
<td>IB</td>
<td>CA-125 &amp; TVU +</td>
<td>793</td>
<td>40</td>
</tr>
<tr>
<td>Ovary</td>
<td>Yes</td>
<td>Serous cystadenocarcinoma</td>
<td>Grade II</td>
<td>IA</td>
<td>TVU positive</td>
<td>13</td>
<td>213</td>
</tr>
<tr>
<td>Ovary</td>
<td>Yes</td>
<td>Papillary serous &amp; endometrial</td>
<td>Grade III</td>
<td>IIIC</td>
<td>CA-125 &amp; TVU +</td>
<td>456</td>
<td>42</td>
</tr>
<tr>
<td>Ovary</td>
<td>Yes</td>
<td>Endometrioid adenocarcinoma</td>
<td>Grade III</td>
<td>IIIA</td>
<td>TVU positive</td>
<td>15</td>
<td>157</td>
</tr>
<tr>
<td>Ovary</td>
<td>Yes</td>
<td>Serous cystadenocarcinoma</td>
<td>Grade III</td>
<td>IIIB</td>
<td>TVU positive</td>
<td>9</td>
<td>72</td>
</tr>
<tr>
<td>Ovary</td>
<td>Yes</td>
<td>Adenocarcinoma, NOS</td>
<td>Grade III</td>
<td>IIIIB</td>
<td>CA-125 positive</td>
<td>1260</td>
<td>49</td>
</tr>
<tr>
<td>Ovary</td>
<td>Yes</td>
<td>Serous cystadenocarcinoma</td>
<td>Grade III</td>
<td>IIIIC</td>
<td>CA-125 &amp; TVU +</td>
<td>1556</td>
<td>65</td>
</tr>
<tr>
<td>Ovary</td>
<td>Yes</td>
<td>Serous cystadenocarcinoma</td>
<td>Grade III</td>
<td>IIIIC</td>
<td>CA-125 positive</td>
<td>2182</td>
<td>76</td>
</tr>
<tr>
<td>Ovary</td>
<td>Yes</td>
<td>Serous cystadenocarcinoma</td>
<td>Grade III</td>
<td>IIIIC</td>
<td>CA-125 &amp; TVU +</td>
<td>73</td>
<td>34</td>
</tr>
<tr>
<td>Ovary</td>
<td>Yes</td>
<td>Serous cystadenocarcinoma</td>
<td>Grade III</td>
<td>IIIIC</td>
<td>CA-125 positive</td>
<td>406</td>
<td>49</td>
</tr>
<tr>
<td>Ovary</td>
<td>Yes</td>
<td>Serous cystadenocarcinoma</td>
<td>Grade III</td>
<td>IIIIC</td>
<td>CA-125 positive</td>
<td>875</td>
<td>80</td>
</tr>
<tr>
<td>Ovary</td>
<td>Yes</td>
<td>Adenocarcinoma</td>
<td>Grade III</td>
<td>IIIIC</td>
<td>CA-125 &amp; TVU +</td>
<td>1143</td>
<td>104</td>
</tr>
<tr>
<td>Ovary</td>
<td>Yes</td>
<td>Mucin-producing adenocarcinoma</td>
<td>Grade III</td>
<td>IIIIC</td>
<td>CA-125 positive</td>
<td>78</td>
<td>44</td>
</tr>
<tr>
<td>Ovary</td>
<td>Yes</td>
<td>Papillary adenocarcinoma</td>
<td>Grade III</td>
<td>IIIIC</td>
<td>CA-125 &amp; TVU +</td>
<td>123</td>
<td>23</td>
</tr>
<tr>
<td>Ovary</td>
<td>Yes</td>
<td>Carcinoma, NOS</td>
<td>Grade III</td>
<td>IIIIC</td>
<td>CA-125 &amp; TVU +</td>
<td>410</td>
<td>59</td>
</tr>
<tr>
<td>Ovary</td>
<td>Yes</td>
<td>Serous cystadenocarcinoma</td>
<td>Grade III</td>
<td>IV</td>
<td>TVU positive</td>
<td>31</td>
<td>237</td>
</tr>
<tr>
<td>Fallopian Tube</td>
<td>Yes</td>
<td>Endometrioid adenocarcinoma</td>
<td>Grade III</td>
<td>IIIB</td>
<td>CA-125 &amp; TVU +</td>
<td>96</td>
<td>21</td>
</tr>
<tr>
<td>Fallopian Tube</td>
<td>Yes</td>
<td>Papillary adenocarcinoma, NOS</td>
<td>Grade III</td>
<td>IV</td>
<td>CA-125 positive</td>
<td>85</td>
<td>96</td>
</tr>
<tr>
<td>Peritoneum</td>
<td>Yes</td>
<td>Carcinoma, NOS</td>
<td>Grade III</td>
<td>IIIIC</td>
<td>CA-125 positive</td>
<td>69</td>
<td>83</td>
</tr>
</tbody>
</table>

* Low malignant potential.

† Grade determined from metastasis.
abnormal. In this table, all biopsies were performed surgically, either with laparoscopy or laparotomy, as indicated in the table legend.

### Follow-up diagnostic procedures

All participants with at least 1 abnormal screening result (n = 1706) were tracked to assess the diagnostic procedures performed as a consequence of an abnormal screening test result. Because the PLCO protocol did not mandate specific evaluation of a woman with positive screens, follow-up procedures up to and including surgery were at the discretion of the patient’s physician. Each subject and their physician received a letter that notified them of the positive screen and the recommendation to obtain follow-up, but some subjects (about 15%) did not receive further evaluation of a positive screen. Follow-up procedures are summarized in Table VI. Of importance, 570 women underwent a surgical procedure (325 laparotomy and 245 laparoscopy and/or vaginal approach), including 541 who proved not to have a neoplasm. Thus, 541 of 1706 subjects (31.7%) who had at least 1 positive screening test underwent surgery but did not have cancer.

### Comment

Screening to detect early-stage ovarian cancer is theoretically appealing because this malignancy is typically heralded by vague, nonspecific symptoms and is characterized by advanced stage at diagnosis. This report of the initial ovarian cancer screen in women aged 55 to 74 who volunteered to participate in the PLCO study demonstrates some of the practical difficulties inherent in screening for diseases with a low prevalence in the target population.

We screened 28,816 women and detected 29 neoplasms, or 1 neoplasm for every 994 subjects screened. Nineteen were invasive epithelial cancers, or 1 cancer for every 1517 subjects screened. The cancers were identified among 1338 women who had an abnormal TVU (4.7% of all women screened) and 402 who had an abnormal CA-125 (1.4% of all women screened). Diagnostic evaluation of these abnormalities included not only relatively benign (although expensive and anxiety-provoking) radiographic procedures, but also 570 surgical procedures, including 325 laparotomies. When used alone, the positive predictive value of TVU for invasive cancer was 1.0%, and for CA-125 it was 3.7%. The PPV of TVU in our study was lower than in previous studies, possibly because of the multicenter nature of the PLCO trial. TVU results from studies conducted with a single or a limited number of ultrasonographers may have a higher PPV. On the other hand, the PPV of CA-125 in the PLCO trial was higher than in other studies. PLCO subjects were all 55 years of age or older and, therefore, almost exclusively postmenopausal. Premenopausal women, who were included in most...
previous studies of screening, have a larger number of false positive CA-125 values than older women. Not surprisingly, an abnormal TVU resulted in a higher rate of surgeries (535 of 1338, or 40%) than an abnormal CA-125 (62 of 403, or 15.4%). Knowing that many other (nonmalignant) conditions may influence CA-125 levels, it is harder to justify performing surgery in an asymptomatic subject with an elevated CA-125 than in a similar patient with an adnexal mass found on TVU. Abnormalities in both tests were found in only 34 subjects, in whom the PPV for the 2 tests combined was 23.5%. However, if one chose to evaluate only subjects in whom both screening tests were abnormal, 20 of the 29 ovarian neoplasms (and 12 of the 20 invasive cancers) would have been missed. We identified 1 neoplasm for every 20 surgical procedures performed. If surgery was done to evaluate an abnormal CA-125, 1 neoplasm was identified per 3.9 surgeries (16 of 42), whereas 24 surgeries were required to identify 1 neoplasm based on an abnormal TVU (22 of 535). One neoplasm was found per 3 surgeries (9 of 27) if both TVU and CA-125 were abnormal.

Almost 2% of women who were screened at the time of entry into the PLCO trial (570 of 28,816) underwent abdominal surgery in the course of evaluating abnormalities detected by the PLCO ovarian cancer screening strategy. Several factors may have contributed to the high rate of pelvic surgery among the women in this series. 1) Many of the subjects may have had previously existing indications for elective pelvic surgery (bladder dysfunction, etc), with the finding of an abnormal TVU or CA-125 simply being the final impetus for performing the procedure. 2) Follow-up may have been particularly aggressive because subjects were part of a national cancer screening study, and physicians may have felt that their medical decisions would be closely scrutinized. However as noted above, 15% of subjects with a positive screen had no follow-up. 3) Ovarian screening with TVU and CA-125 is not done routinely, and the false-positive rate is not as well defined as that for other more common screening procedures. Because ovarian cancer is notoriously a highly lethal disease, physicians may not have been comfortable with a “watch and wait” approach to those with an abnormal screening test. 4) The definitive diagnostic test for an abnormal ovary seen on TVU is oophorectomy. Percutaneous biopsy is not utilized as routinely for diagnosis of ovarian masses as it is for many other organs, because of the risk of false-negative studies and the theoretical possibility of seeding the peritoneum with cancer cells if the biopsied lesion proves to be malignant. 5) Hysterectomy with or without oophorectomy has been performed very commonly in the past, particularly among women in the PLCO age range. By way of illustration, 35.6% of women enrolling in the PLCO Trial (13,918 of 39,115; see Table 1) had undergone hysterectomy and/or oophorectomy before joining the study. Additionally, oophorectomy can often be done by laparoscopy. Physicians therefore may have had a low threshold for recommending this common procedure.

In addition to the invasive tumors, 9 cystadenomas of low malignant potential (“borderline”) tumors were identified, representing 31% of the total malignant neoplasms found. This entity typically comprises only about 15% of all ovarian tumors as they present clinically in the general population. While the number of cancer cases in this report is small, this observation suggests that screening as performed in the PLCO Trial, particularly with TVU, may preferentially detect low-grade tumors. This “length” bias (ie, slowly growing tumors are more likely than rapidly growing tumors to be detected with screening) has the ultimate effect of finding neoplasms that might go undetected for the lifetime of the patient, and therefore, not contribute to cancer mortality (“overdiagnosis”). Using mortality as the end point of a screening trial (as will be done for the PLCO study when follow-up is complete) eliminates the length bias. The detection of ovarian tumors of low malignant potential is unlikely to affect ovarian cancer mortality, since even stage III borderline tumors have a 90% 10-year disease-free survival. Because only data from the initial (baseline) ovarian cancer screen have been analyzed thus far, the effect of repeated annual screens on detection rates and mortality, as is the protocol in the PLCO Trial, is currently unknown. The predictive value of these tests may be improved if post hoc analysis identifies a specific pattern of change over time in either TVU or CA-125 results, or if the combination of a specific TVU imaging pattern and a relatively high CA-125 correlates with a greater likelihood of having ovarian cancer. Fifteen of the 19 invasive neoplasms (79%) found

Table VI  Participant-based diagnostic procedures following a positive screen

<table>
<thead>
<tr>
<th>Diagnostic procedures</th>
<th>No neoplasm</th>
<th>Neoplasms</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA-125</td>
<td>377 (22.5%)</td>
<td>12 (41.4%)</td>
</tr>
<tr>
<td>Ultrasounds</td>
<td>721 (43.0%)</td>
<td>12 (41.4%)</td>
</tr>
<tr>
<td>Chest radiograph</td>
<td>68 (4.1%)</td>
<td>7 (24.1%)</td>
</tr>
<tr>
<td>Surgery*</td>
<td>541 (32.3%)</td>
<td>29 (100.0%)</td>
</tr>
<tr>
<td>CT scan/MRI</td>
<td>150 (8.9%)</td>
<td>12 (41.4%)</td>
</tr>
<tr>
<td>Needle aspiration, culdocentesis, or paracentesis</td>
<td>22 (1.3%)</td>
<td>1 (3.4%)</td>
</tr>
<tr>
<td>IVP</td>
<td>8 (0.5%)</td>
<td>-</td>
</tr>
<tr>
<td>Barium enema</td>
<td>7 (0.4%)</td>
<td>-</td>
</tr>
<tr>
<td>No diagnostic procedure recorded</td>
<td>260 (15.5%)</td>
<td>-</td>
</tr>
<tr>
<td>Total number of participants</td>
<td>1677 (100.0%)</td>
<td>29 (100.0%)</td>
</tr>
</tbody>
</table>

* Two hundred ninety-eight of the surgeries of the participants without neoplasms were a laparotomy; 27 of the surgeries of the participants with neoplasms were a laparotomy.
with the initial (prevalent) screen were stage III or IV cancers. Now that these high-stage cancers have been removed from the screened population, it is possible that subsequent screens will identify a group of early stage ovarian cancers with improved survival.

Further follow-up from the PLCO Trial will provide valuable data regarding the effect of annual screening on ovarian cancer mortality. At the present time, nothing in the findings reported here suggests a need to revise the present (1996) ovarian cancer screening guidelines of the US Preventive Services Task Force,21 which state “routine screening for ovarian cancer by ultrasound, the measurement of serum tumor markers, or pelvic examination is not recommended.”

References

Endometrial cancer in women 45 years of age or younger: A clinicopathological analysis

Gilbert P. Pellerin, MD, Michael A. Finan, MD

Department of Obstetrics and Gynecology, Ochsner Clinic Foundation, New Orleans, LA

Received for publication November 4, 2004; revised April 6, 2005; accepted May 2, 2005

Objective: The purpose of this study was to evaluate the experience with endometrial carcinoma in women ≤45 years of age at Ochsner Clinic Foundation, New Orleans, La.

Study design: We evaluated the clinical history, treatment, and follow-up of 38 women ≤45 years of age diagnosed with endometrial cancer.

Results: Thirty-eight patients received primary treatment for endometrial cancer: stage I, 32 (84.2%); stage II, 1 (2.6%); stage III, 4 (10.5%); stage IV, 1 (2.6%). Tumors were well differentiated in 20 (52.6%), moderately differentiated in 10 (26.3%), and poorly differentiated in 8 (21.1%). At end of study period 32 women (84.2%) were alive with no evidence of disease, 5 had died of recurrent disease, and 1 died of metastatic breast cancer.

Conclusion: Patients ≤45 years of age had lower incidence of advanced stage disease, higher degree of tumor differentiation, and better prognosis compared to patients older than 45 years.

Endometrial carcinoma is the most common malignancy of the lower female genital tract in the US. Approximately 40,880 new cases develop in the US each year, according to the year 2005 figures from the American Cancer Society. This is about 1.3 times the frequency of ovarian cancer and approximately twice the number of new cases of cervical cancer. Overall, about 1 woman in 50 in the US will develop endometrial carcinoma during her life.

Endometrial adenocarcinoma occurs during the reproductive and menopausal years. The median age for adenocarcinoma of the uterine corpus is 61 years; most patients are between the ages of 50 and 59 years. Approximately 5% of women will have adenocarcinoma before age 40 years, and 20% to 25% will be diagnosed before menopause. Multiple factors increase the risk of developing endometrial carcinoma, including obesity, nulliparity, unopposed estrogen, and late menopause. Most young patients who develop endometrial cancer are obese, in many instances massively overweight, often with anovulatory menstrual cycles. Combination oral contraceptive use decreases the risk for development of endometrial cancer.

Postmenopausal bleeding and abnormal premenopausal bleeding are the primary symptoms of endometrial carcinoma. All postmenopausal women with uterine bleeding should be evaluated for endometrial cancer. As a patient’s age increases after menopause, the probability that uterine bleeding is caused by endometrial cancer progressively increases. According to Feldman et al, a woman > 70 years has about a 50% chance of having cancer when vaginal bleeding is present.
high index of suspicion must be maintained if the diagnosis of endometrial cancer is to be made in the young patient. Prolonged and heavy menstrual periods, and intermenstrual spotting, may indicate cancer, and endometrial sampling is advised. Most young patients who develop endometrial cancer are obese, in many instances massively overweight, often with anovulatory menstrual cycles.

Adenocarcinoma, the most common histologic type of endometrial cancer, is sometimes preceded by a predisposing lesion: atypical endometrial hyperplasia. Patients with atypical endometrial hyperplasia have a 30% to 50% risk of simultaneous invasive endometrial cancer. There are various types of endometrial carcinoma.

Endometrial cancer is staged using the surgicopathologic staging system adopted by the International Federation of Gynecology and Obstetrics (FIGO) in 1988. Tumor stage is a well-recognized prognostic factor for endometrial carcinoma; higher stage of disease indicates a worse prognosis.

Another major determinant of prognosis is the histologic grade of the tumor. As the tumor loses its differentiation, survival decreases. A better prognosis for endometrial carcinoma is associated with endometrioid adenocarcinoma, as well as better differentiation of the tumor with or without squamous elements, and secretory carcinomas. Approximately 80% of all endometrial carcinomas fall into the favorable category. Poor prognostic histologic types are papillary serous carcinomas, clear cell carcinomas, and poorly differentiated carcinoma with or without squamous elements.

The incidence of endometrial adenocarcinoma has increased in recent years, but the disease remains uncommon in premenopausal women. Depending on the age cut-off used for the study, proportions between 2% and 14% have been reported. A literature review suggests that endometrial adenocarcinoma in young women is often associated with early stage disease, high differentiation of the tumor, and a good prognosis. Endometrial cancer is frequently unsuspected clinically or pathologically in women under the age of 45 years because greater than 75% of cases occur in patients over 50. The presence of obesity and abnormal vaginal bleeding should encourage a greater practice of endometrial sampling.

The purpose of this retrospective study was to evaluate the risk factors and outcomes of 38 women aged 45 years of age with endometrial cancer.

Material and methods

From 1982 through 2002, 463 women underwent therapy at Ochsner Clinic Foundation and Hospital, New Orleans, LA, for adenocarcinoma of the endometrium. Patients with other types of malignancies (leiomyosarcoma, fibrosarcoma, carcinosarcoma, or choriocarcinoma) were excluded. Thirty-eight patients were ≤ 45 years of age, and 425 patients were > 45 years of age.

From the hospital records of the 38 patients, data pertaining to age, parity, body mass index (BMI), prediagnostic hormone treatment, result of the last Papanicolaou smear, preceding symptoms, smoking history, past medical history, family history, and postoperative hormone replacement therapy were tabulated. International Federation of Gynecology and Obstetrics (FIGO) staging and grading were investigated.

Tumors were staged according to the FIGO guidelines of 1988. The BMI was calculated as body weight divided by the square of height (kg/m²). A BMI of 25 indicated the patient was overweight and above 30 as obese.

Statistical analysis consisted of the use of Pearson chi-square or Fisher exact test where appropriate. Survival curves were estimated using the Life-table analysis method. Statistical significance was defined as P < .05. SPSS 6.1 statistical analysis software was used (Chicago, IL).

Results

Of the 38 patients in the study group, the mean age at diagnosis was 39 years (range 28-45). The clinical details of each patient are summarized in Table I. Clinical details for the 425 patients > 45 years of age are limited to stage and survival data. Three (7.9%) patients had a previous history of colon cancer, 1 (2.6%) had a history of breast cancer, and 1 (2.6%) had a history of thyroid cancer. Thirteen (34.2%) women had 1 first-degree relative with a history of malignant disease, which involved breast, ovarian, endometrial, or colon cancer. Three (7.9%) women had more than 1 first-degree relative with a history of malignant disease.

The presenting symptom was abnormal uterine bleeding in 35 (92.1%) cases and pelvic pain related to a pelvic mass in 1 (2.6%). One patient was diagnosed by dilatation and curettage (D&C) secondary to the evaluation of a high grade squamous intraepithelial lesion pap, and another patient had an endocervical biopsy that revealed atypical hyperplasia, prompting a D&C that resulted in the diagnosis of endometrial cancer. The diagnosis of endometrial cancer was made by endometrial biopsy in 14 (36.8%) patients and by D&C in 18 (47.4%). Six (15.8%) patients failed to have carcinoma diagnosed before hysterectomy.

All women underwent total abdominal hysterectomy and bilateral salpingo-oophorectomy. In 25 (65.8%) instances, pelvic lymph node dissections were also performed. Tumors were classified as endometrioid adenocarcinoma in 34 (89.5%), adenosquamous carcinoma in 3 (7.9%), and adenoacanthoma in 1 (2.6%). Stage I disease was found in 32 (84.2%), stage II in...
1 (2.6%), stage III in 4 (10.5%), and stage IV in 1 (2.6%). Tumors were classified as well differentiated in 20 (52.6%), moderately differentiated in 10 (26.3%), and poorly differentiated in 8 (21.1%). Eleven patients had postoperative radiation therapy.

Patient follow-up after surgery ranged from 1 month to 260 months. The follow-up period was >3 years for 13 patients, between 3 and 5 years for 10, and >5 years for 15. During the follow-up period, 5 patients died of progression or recurrence of endometrial carcinoma, and 1 patient who had no sign of recurrence of endometrial cancer died of widely metastatic breast cancer. The Figure shows the Life-table analysis curves for patients 45 years of age or younger (n = 38) and for those older than 45 years (n = 425). The estimated 5-year survival rate for all patients diagnosed with endometrial cancer during the study period was 71%. The estimated 5-year survival rates of women age 45 years or younger and women over the age of 45 were 82% and 70%, respectively. The 5-year survival for patients 45 years of age or younger and patients greater than 45 years of age are shown in Table II. There were not enough patients with stage II or stage IV disease to calculate a survival curve.

### Table I  Clinical details of all patients

<table>
<thead>
<tr>
<th>Age (28–45 y)</th>
<th>No. of patients</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;40</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>41–45</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Body mass index (18–63)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (&lt;20)</td>
<td>2</td>
<td>5.3%</td>
</tr>
<tr>
<td>Normal (20–25)</td>
<td>4</td>
<td>10.5%</td>
</tr>
<tr>
<td>Overweight (25–30)</td>
<td>2</td>
<td>13.2%</td>
</tr>
<tr>
<td>Obese (&gt;30)</td>
<td>27</td>
<td>71.0%</td>
</tr>
<tr>
<td>Parity (0–3)</td>
<td>No. of patients</td>
<td>Percent</td>
</tr>
<tr>
<td>Para 0</td>
<td>23</td>
<td>60.5%</td>
</tr>
<tr>
<td>Para 1</td>
<td>4</td>
<td>10.5%</td>
</tr>
<tr>
<td>Para 2</td>
<td>8</td>
<td>21.1%</td>
</tr>
<tr>
<td>Para 3</td>
<td>3</td>
<td>7.9%</td>
</tr>
<tr>
<td>Stage</td>
<td>No. of patients</td>
<td>Percent</td>
</tr>
<tr>
<td>Stage Ia</td>
<td>12</td>
<td>31.5%</td>
</tr>
<tr>
<td>Stage Ib</td>
<td>20</td>
<td>52.6%</td>
</tr>
<tr>
<td>Stage II</td>
<td>1</td>
<td>2.7%</td>
</tr>
<tr>
<td>Stage III</td>
<td>4</td>
<td>10.5%</td>
</tr>
<tr>
<td>Stage IV</td>
<td>1</td>
<td>2.7%</td>
</tr>
<tr>
<td>Histology</td>
<td>No. of patients</td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>20</td>
<td>52.6%</td>
</tr>
<tr>
<td>G2</td>
<td>10</td>
<td>26.3%</td>
</tr>
<tr>
<td>G3</td>
<td>8</td>
<td>21.1%</td>
</tr>
<tr>
<td>Follow-up</td>
<td>No. of patients</td>
<td>Percent</td>
</tr>
<tr>
<td>&lt;3 years</td>
<td>13</td>
<td>34.2%</td>
</tr>
<tr>
<td>3–5 years</td>
<td>10</td>
<td>26.3%</td>
</tr>
<tr>
<td>&gt;5 years</td>
<td>15</td>
<td>39.5%</td>
</tr>
<tr>
<td>Medical history</td>
<td>No. of patients</td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>9</td>
<td>23.7%</td>
</tr>
<tr>
<td>Hypertension</td>
<td>8</td>
<td>21.1%</td>
</tr>
<tr>
<td>Cervical cytology</td>
<td>No. of patients</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>19</td>
<td>79.2%</td>
</tr>
<tr>
<td>Squamous abnormal</td>
<td>3</td>
<td>12.5%</td>
</tr>
<tr>
<td>Glandular abnormal</td>
<td>2</td>
<td>8.3%</td>
</tr>
<tr>
<td>Smoking history</td>
<td>No. of patients</td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>8</td>
<td>22.2%</td>
</tr>
<tr>
<td>Never used</td>
<td>28</td>
<td>77.7%</td>
</tr>
<tr>
<td>Oral contraceptive use</td>
<td>No. of patients</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>8</td>
<td>21.0%</td>
</tr>
<tr>
<td>No</td>
<td>30</td>
<td>79.0%</td>
</tr>
</tbody>
</table>

### Table II  Five-year survival by stage for patients less than or equal to 45 years of age versus patients greater than 45 years of age

<table>
<thead>
<tr>
<th>Stage</th>
<th>≤45 years</th>
<th>≥46 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>93.1</td>
<td>79.9</td>
</tr>
<tr>
<td>2</td>
<td>0.0</td>
<td>68.7</td>
</tr>
<tr>
<td>3</td>
<td>66.7</td>
<td>42.3</td>
</tr>
<tr>
<td>4</td>
<td>0.0</td>
<td>13.2</td>
</tr>
</tbody>
</table>

### Figure  Survival curve (Life-table analysis curves). Age less than or equal to 45 years and combined stage versus age greater than 45 years and combined stage.

### Comment

Endometrial cancer in young women can be difficult to diagnose. It is uncommon and usually presents as abnormal uterine bleeding at an age when dysfunctional bleeding is much more common. Obesity seems to be strongly associated with the development of endometrial cancer in young women.7,12,20-22 This theory is supported in the present review by the finding that 71% of the patients studied were classified as obese with a BMI greater than 30.

In our study, 92.1% of patients presented to their physicians with abnormal vaginal bleeding. The most common physical abnormality was patient obesity. The combination of abnormal vaginal bleeding and obesity should alert the physician to the possibility of underlying endometrial abnormalities, yet the diagnosis of adenocarcinoma was first made after hysterectomy in 15.8% of cases.

No statistically significant correlation was found when age, reproductive history, smoking history, race, past medical history, or family history was compared with histologic grade or surgical stage of the tumor. The incidence of endometrial carcinoma in our patients who were 45 years of age or younger was 8.1%, similar to the findings in other reports.7,11,12,20,21,23
Young patients with endometrial carcinoma tend to have prognostically favorable histologic types, such as well-differentiated adenocarcinoma (G1) and adenosquamous carcinoma. In addition, young patients with endometrial carcinoma were more likely to present with stage I disease. Quinn et al reviewed the literature and their own series of endometrial cancer in premenopausal women and reported that 154 of 231 patients (67%) had a well-differentiated tumor, but on the other hand, 11 of 231 patients (4.8%) exhibited adenosquamous carcinoma, a relatively unfavorable histologic type. Of 106 premenopausal patients investigated by Quinn et al, 96 (91%) of patients were diagnosed with stage I disease. In the series by Crissman et al, all 32 patients had well-differentiated tumors and 12 of the 32 were adenoacanthoma. In the series by Yasumizu et al, 30 of the 76 patients (39.5%) had a well-differentiated tumor with 7 patients (9.2%) having adenosquamous carcinoma. In addition, 40 out of 76 patients (52.6%) had stage I disease.

In our study, the association between cancer cell type and stage of disease was statistically significant. Patients with endometrial adenocarcinoma were more likely to present with stage I disease, whereas patients with adenosquamous carcinoma were more likely to have a higher stage of disease. There was no statistical significance in relation to cell type versus grade of tumor.

In our study, 84.2% of patients presented with stage I disease. The incidence of early stage disease and well-differentiated tumors in our series is more consistent with the results obtained by Quinn et al and slightly higher than the results obtained by Yasumizu et al.

In general, endometrial carcinoma in young women has a favorable histology (well-differentiated tumor) and early stage disease (stage I). In our study, 32 of 37 patients were alive and without evidence of tumor recurrence, with 15 patients (39.5%) having a follow-up time greater than 60 months. Of the remaining patients, 5 died of disease and 1 died of widely metastatic breast cancer 4 years after the diagnosis of endometrial cancer.

Statistical analysis revealed a positive association between both recurrence and survival compared with stage of disease and histologic grade of tumor. In terms of recurrence, there was a statistically significant correlation between tumor recurrence and both stage of disease and histologic grade, \( P = 0.01 \) and \( P = 0.02 \), respectively. Patients with early stage disease or a high degree of tumor differentiation (G1) had a lower rate of recurrence than patients with advanced disease or poorly differentiated tumor (G2 or G3).

Analysis of survival curves demonstrates that patients with endometrial carcinoma who are 45 years of age or younger have a higher 5-year survival rate compared with those older than 45, 82% versus 70%, respectively. Patients 45 years of age or younger with endometrial cancer also had a higher incidence of stage I disease than those older than 45, 84% versus 72%, respectively. The 5-year survival for patients with stage I disease was 92% for those 45 years of age or younger compared with 79% for those older than 45. Stage for stage, patients with endometrial cancer 45 years of age or younger had a better 5-year survival than those older than 45. The reason for improved survival in younger patients compared with older patients cannot be determined from our small sample size. However, likely reasons for improved survival would include a higher ratio of grade I tumor and stage I disease in our patient population. Endometrial carcinoma related to obesity generally has an improved prognosis in younger patients than endometrial carcinoma in older patients with a low BMI. We have found this to be true in our practice as well (Table I). Patients who are younger, obese, and who present with abnormal bleeding should undergo endometrial biopsy. The clinician’s index of suspicion of endometrial cancer should be raised in those patients who are at an increased risk.

The clinicopathologic comparison between the patients 45 years or younger and those older than 45 demonstrated that the ratio of well-differentiated tumors was higher, the incidence of advanced stage tumors significantly lower, and the prognosis of endometrial carcinoma significantly better in those 45 or younger.

In conclusion, patients with endometrial carcinoma 45 years of age or younger were more likely to be obese. Histologically, the tumors of these patients had prognostically favorable histologic findings, such as well-differentiated tumors (G1) and early stage disease (stage I). Thus, the survival rate of patients 45 years of age or younger was significantly better that of patients older than 45 in our series.

Acknowledgments

Special thanks to Liz Adams, Director, Cancer Registry, Ochsner Clinic Foundation, New Orleans.

References

Laparoscopic staging in patients with incompletely staged cancers of the uterus, ovary, fallopian tube, and primary peritoneum: A Gynecologic Oncology Group (GOG) study

Nick M. Spirtos, MD,a,* Scott M. Eisekop, MD,b Guy Boike, MD,c John B. Schlaerth, MD,d James O. Cappellari, MD,e

Women's Cancer Center,a Los Gatos, CA; Women’s Cancer Center at Encino-Tarzana Regional Medical Center,b Tarzana, CA; Gynecologic Oncology, Synergy Medical Alliance, Saginaw, MI; Department of Obstetrics and Gynecology, Michigan State University,c East Lansing, MI; Women’s Cancer Center at Pasadena,d Pasadena, CA; Department of Pathology, Wake Forest University School of Medicine,e Winston-Salem, NC

Received for publication January 16, 2004; revised April 1, 2005; accepted May 2, 2005

KEY WORDS
Laparoscopy
Interval staging
Gynecologic malignancy

Objective: The purpose of this study was to determine the feasibility of laparoscopically staging patients with incompletely staged cancers of the uterus, ovary, fallopian tube, and primary peritoneum, and to evaluate related effects.

Study design: Patients without evidence of metastatic disease had laparoscopic bilateral para-aortic and pelvic lymph node dissection. Other procedures were individualized based on extent of the primary surgery; laparotomy was undertaken for identified resectable disease.

Results: Ninety-five eligible patients were entered on 2 Gynecologic Oncology Group (GOG) protocols. Eleven were excluded. Fifty-eight patients (69%) underwent complete endoscopic staging with photographic documentation. Nine others (10%) were incompletely staged. Seventeen patients (20%) had laparotomy. In patients undergoing laparoscopy, 6% had bowel complications; 11% were found to have more advanced disease. Hospital stay was significantly shorter with laparoscopy alone (3 vs 6 days, \( P = .04 \)).

Conclusion: Interval laparoscopic staging of gynecologic malignancies can be successfully undertaken in selected patients, but laparotomy for adhesions or metastatic disease and risk of visceral injury may be anticipated.

© 2005 Mosby, Inc. All rights reserved.
Except in cervical cancer, surgical staging has surpassed clinical staging as the recognized standard for all gynecologic malignancies. Until recently, patients with incompletely staged gynecologic cancers faced either exploratory laparotomy to perform staging, or potential under- or overtreatment. The introduction of minimally invasive surgical techniques has led to the evaluation of their utility in the management of patients with incompletely staged gynecologic malignancies. Case reports and small series of patients describe the successful use of minimally invasive techniques to stage patients with incompletely staged cancers of the reproductive organs. Claims of decreased morbidity, shortened hospital stays, and early return to normal activity gave rise to evaluating these techniques on a prospective basis.

In July of 1992, the Gynecologic Oncology Group (GOG) initiated protocols #9302 and #9402 to determine the feasibility of completing surgical staging in patients with incompletely staged ovarian, fallopian tube, endometrial, and primary peritoneal cancers, and to assess the associated adverse effects.

Materials and methods

In January of 1993, a subcommittee of the GOG Gynecologic Oncology Management Committee met and, with input from the GOG Statistical and Data Center, agreed that it would require approximately 50 fully evaluable patients to assess morbidity and mortality associated with minimally invasive surgery in this setting, as well as to assess the feasibility of successfully completing surgical staging.

Eligibility required confirmation of primary ovarian, uterine, fallopian tube, or peritoneal carcinoma in which, at minimum, a retroperitoneal lymph node dissection had not been undertaken at initial surgery. A GOG performance status of 0 or 1, and normal bone marrow, renal, and hepatic functions were also required. The interval between surgical procedures must have been less than 10 weeks, and patients must have had a Quetelet index of less than 35. Further, patients without nodal tissue identified at the time of pathology review were considered incompletely staged.

Variables evaluated included age, Quetelet index, operative times, estimated blood loss, lymph node counts, and hospital stay for patients undergoing laparoscopy and laparotomy.

Postoperative morbidity and mortality was graded according to GOG criteria and limited to occurrence within 30 days of the second surgical procedure.

Comparison of means within groups of patients was accomplished with the independent samples t test. Analysis of discrete and binomial data utilized the chi-square test. Statistical analysis was accomplished using the SPSS statistical package, version 10.0 (SPSS Incorporated, Chicago, IL). Comparison of the results from this investigation with other published outcomes was accomplished with a literature review using the Unabridged Medline as accessed through version 4.27 of Knowledge Finder software (Aries Systems Corporation, North Andover, MD).

Results

Ninety-five patients were entered from nine institutions over a period of 7 years: 73 patients with ovarian, fallopian tube, or primary peritoneal cancer, and 22 patients with uterine cancer. Eleven patients were excluded based on pathology review, progression of disease, incomplete video or photographic documentation, or refusal to participate. Nine patients were incompletely staged laparoscopically, lacking peritoneal biopsies, bilateral lymph nodes, or cytology. Seventeen patients required laparotomy, 13 resulting from poor exposure as a result of adhesions, 3 caused by complications, and 1 caused by metastatic macroscopic disease, leaving 58 evaluable patients, all with a performance status = 0, having had complete surgical staging using minimally invasive techniques. A comparison of variables evaluated including age, Quetelet index, operative times, estimated blood loss, lymph node counts, and hospital stay is presented for patients undergoing laparoscopy and laparotomy (Table I). These results include data for 75 patients, the 58 evaluable and 17 patients undergoing laparotomy. Those patients undergoing
of patients, respectively. Of 84 eligible patients, 17 (20%) required immediate laparotomy regardless of the cause differed in some respects from those having surgical staging completed using minimally invasive techniques (Table I).

Excluding the 11 ineligible or inevaluable patients, complications in the remaining 84 cases associated with laparoscopy included: bowel injury (5); cystotomy (1); small bowel obstruction (1); fever (1); venotomy (1); and excess blood loss requiring transfusion (2). One bowel injury, the cystotomy, and the venotomy were recognized immediately, and laparotomy undertaken. The other bowel injuries were recognized in the postoperative period before discharge, and the patients underwent corrective surgery without deleterious long-term effect. There were no intra- or postoperative deaths. One patient died from progressive disease before surgery.

Six patients with ovarian cancer and 2 with endometrial cancer were found to have advanced (stage II or III) disease. Two patients were found to have disease involving the opposite ovary, unrecognized at primary surgery. Details of histopathology and staging changes are seen in Table II.

**Comment**

These phase II studies were undertaken to determine if complete surgical staging could be successfully undertaken laparoscopically in patients having recently (within 10 weeks) undergone incomplete surgical staging. Additionally, adverse effects related to laparoscopy in this setting were to be evaluated. Both objectives were met.

Of 84 eligible patients, 17 (20%) required immediate laparotomy because of either poor exposure or surgical complications. Fifty-eight patients met all protocol requirements, including photographic documentation of the surgical procedure. Nine patients had incomplete staging undertaken according to the protocol; some, but not all, of the required peritoneal biopsies were taken. Only unilateral lymph node dissection was undertaken; in 1 case, no lymph nodes were identified in the specimen from a single site of dissection. Six patients had no photographic documentation and were inevaluable. Over 70% of the patients had complete surgical staging using minimally invasive techniques and, allowing for individualization of patient care, it could be argued that 6 additional patients had adequate surgery, for a total of approximately 75%.

Need for immediate laparotomy in 17 (20%) patients, 14 caused by adhesions and 3 from complications, helps establish a baseline in this patient population. In 2 other series describing similar groups of patients, laparotomy was required in 6% and 10% of the patients, and complications reported in 14% and 20% of those same patients, respectively. The relatively high conversion rate in this multi-institutional trial might be explained if consideration was given to the experience of the surgeons in the referenced series compared with those contributing patients to these trials. Even though the number of patients accrued by institution is identifiable, it is not possible to quantify the experience of any individual surgeon. Moreover, there is no method for a direct comparison of surgical experience between studies. As adhesions resulted in the need for laparotomy in most patients, perhaps the timing of the procedure contributed to the rate of conversion to laparotomy. This supposition is not supportable, as the interval between surgical procedures in other comparable reports is similar to the standard (<10 weeks) used for this study.

Of much more concern are the 5 patients with unrecognized bowel injury. In reports describing a similar group of patients undergoing laparotomy, Soper et al reported an incidence of 20% for both vascular and bowel injury. Stier et al reported no serious intraoperative bowel or vascular injury in 47 patients. Specific concerns relating to patients undergoing laparoscopy stem from the fact that electrosurgical instrumentation with shafts extending beyond visual fields are used; thus, injury to the small bowel resulting from inadvertent contact or poor insulation is possible. These

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Laparoscopy: completely evaluable (58)</th>
<th>Laparotomy: caused by adhesions (14) or complications (3)</th>
<th>P value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>43.9 ± 13.8</td>
<td>47.1 ± 1.98</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Quetelet index</td>
<td>24.6 ± 4.6</td>
<td>29.2 ± 3.8</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>EBL</td>
<td>171.9 ± 128.0</td>
<td>352.8 ± 415.0</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>OR time</td>
<td>187.9 ± 59.8</td>
<td>218 ± 72.8</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>No. RPLN</td>
<td>9.6 ± 5.2</td>
<td>9.14 ± 9.9</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>No. RALN</td>
<td>9.0 ± 6.0</td>
<td>8.2 ± 5.0</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>No. LPLN</td>
<td>5.0 ± 3.4</td>
<td>4.4 ± 2.7</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>No. LALN</td>
<td>5.3 ± 3.3</td>
<td>4.6 ± 2.6</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Hospital stay</td>
<td>3.35 ± 5.10</td>
<td>7.31 ± 9.3</td>
<td>&gt;.04</td>
</tr>
</tbody>
</table>

* No significant difference.

<table>
<thead>
<tr>
<th>Primary site and number of patients</th>
<th>Initial stage</th>
<th>Final stage Positive site</th>
<th>Initial/final grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uterus/endo (1)</td>
<td>Ib</td>
<td>IIA (tube)</td>
<td>2/2</td>
</tr>
<tr>
<td>Uterus/endo (1)</td>
<td>Ib</td>
<td>IIIC (nodes)</td>
<td>2/3</td>
</tr>
<tr>
<td>Fallopian tube (1)</td>
<td>Ia</td>
<td>IIA (peritoneum)</td>
<td>2/2</td>
</tr>
<tr>
<td>Ovary (2)</td>
<td>Ia</td>
<td>Ib</td>
<td>3/3; 0/0</td>
</tr>
<tr>
<td>Ovary (3)</td>
<td>Ia</td>
<td>IIIC (LN-2; omentum-1)</td>
<td>3/3; 2/3; 3/3</td>
</tr>
<tr>
<td>Ovary (3)</td>
<td>Ia</td>
<td>IIa (uterus-1; fallopian tube-1; peritoneum-1)</td>
<td>3/3; 2/2; 2/2</td>
</tr>
</tbody>
</table>

Table I Patient characteristics (n = 75)

Table II Changes in stage
results confirm that this risk is not theoretical, and may be increased in the acute postoperative setting (<10 weeks) because of adhesions. The adhesions can obscure the surgeon’s view of the operative field. The presence of adhesions often results in dissectional enterotomies or cystotomies in this setting, and should not be considered unexpected. If identified, these injuries are not difficult to repair, nor associated with a poor surgical outcome. In most, repair can be accomplished laparoscopically. Those patients in whom the problem was immediately identified did not have particularly long hospitalizations or long-term morbidity. The mean hospital stay for patients undergoing laparoscopy only was 3.35 days (standard deviation [sd] 5.10; standard error [se] .68), while those undergoing immediate laparotomy had a mean hospital stay of 7.31 days (sd 9.33; se 2.59). However, in those with bowel complications not immediately identified, mean hospital stay was 13.5 days (sd 11.12; se 5.56). These data suggest that great care must be given to evaluating intraabdominal organs for signs of injury before terminating the surgical procedure. Any compromise of visual inspection of the operative field in this setting may be more problematic than in those patients undergoing laparotomy because surgeons performing laparoscopic procedures are unable to palpate tissues and, therefore, may not appreciate organ injury using only visual signs.

It is of paramount importance that patients be aware of these risks, in particular, and that they be instructed to seek medical attention immediately should symptoms and signs of abdominal organ injury develop. Morbidity associated with visceral injury after a laparoscopic procedure might not become apparent until after discharge after a seemingly successful procedure.

Only 2 patients required transfusion. One required 2 units of packed red blood cells after a venotomy that occurred during laparoscopy and resulted in the patient undergoing laparotomy. Febrile morbidity was noted in 3 patients. These results were comparable to other series whether surgery was performed using traditional, or minimally invasive, surgical techniques.

Eight of 74 patients with ovarian, fallopian tube, or primary peritoneal cancers were found to have more advanced disease, while 2 of 21 patients with uterine cancers were found to have advanced disease that otherwise would have gone undetected.

Two patients initially thought to have ovarian cancer of low malignant potential were reclassified as invasive cancer based on complete histologic review.

Other series have reported a larger percentage of patients with advanced stage disease identified at the time of surgical restaging. Similar results have been reported by Stier et al. This difference might be caused by the high percentage of grade 1 endometrial cancers included in this study (48%) and the study requirement that preoperative radiologic evaluation be negative for metastatic disease. In other series, patients with disease identifiable radiologically would not have necessarily been excluded.

Not surprisingly, the Quetelet index was significantly different when comparing those patients who required laparotomy with those undergoing laparoscopic staging. So, too, was the estimated blood loss and hospital stay (Table I). In this study, the completeness of the lymph node dissection achieved using minimally invasive techniques is supported by the lymph node counts and the photographic documentation required. It should be noted that, to our knowledge, no other surgical technique has been subjected to such scrutiny by a cooperative group or other body. Objective evidence of equivalency between new techniques and an established standard of care should be elucidated before using them interchangeably. This is especially important in the clinical research setting.

Minimally invasive surgical techniques can be used to complete surgical staging in selected patients with most gynecologic malignancies. This study confirms that patients not completely staged are at risk for undertreatment, as 10 in 90 patients (11%) were found to have more advanced disease than was apparent at the time of the first surgical evaluation. Although approximately two thirds of the study population had successful and complete staging performed laparoscopically, the presence of adhesions was the most common finding, resulting in conversion from laparoscopy to laparotomy. The estimated blood loss, hospital stay, and Quetelet index was also greater in patients requiring laparotomy compared with those undergoing laparoscopy. The number of aortic and pelvic lymph nodes resected was not affected by the surgical technique used. Complications appear to be manageable, but special care must be taken in evaluating the gastrointestinal tract or other viscera for injury before terminating the procedure.

Feasibility studies such as this serve as the foundation for randomized prospective studies, such as a trial currently being conducted by the GOG comparing laparoscopy to laparotomy in patients with endometrial cancer.

References


Uterine innervation after hysterectomy for chronic pelvic pain with, and without, endometriosis

Gurprit Atwal, MB ChB,a Daniel du Plessis, MD,a Gordon Armstrong, MD,a Richard Slade, MD,b Martin Quinn, MB ChB, MDb,*

Departments of Histopathologya and Gynaecology, b Hope Hospital, Manchester, UK

Received for publication March 23, 2005; accepted May 6, 2005

KEY WORDS
Chronic pelvic pain
Endometriosis
Reinnervation

Objective: Chronic pelvic pain is associated with a wide range of clinical conditions that include endometriosis. The precise cause, mechanisms of pain, and natural history are imprecise. Patterns of uterine innervation have been studied after hysterectomy for chronic pelvic pain with and without endometriosis.

Study design: Tissue blocks were taken from the lower one half of the uterus after hysterectomy for advanced endometriosis (n = 16 specimens; group 1) and for chronic pelvic pain without endometriosis (n = 15 specimens; group 2). The control group consisted of uteri that were removed for painless gynecologic conditions (n = 25 specimens; group 3). Tissue sections from the lower one half of the uterus were stained with anti-S100 to demonstrate patterns of innervation, and nerve fiber profiles were counted by standardized techniques; qualitative differences were also recorded.

Results: In uteri from women with advanced endometriosis, there were increased numbers of nerve fiber profiles compared with control specimens (group 1 vs group 3; P = .0013, Mann Whitney U test). There were also increased numbers of nerve fiber profiles in uteri that were associated with chronic pelvic pain without endometriosis (group 2 vs group 3; P = .04, Mann Whitney U test). There were no differences in nerve fiber count in uteri from groups 1 and 2 (P = .35, Mann Whitney U test). Comparing both groups of uteri with controls (groups 1 and 2 vs 3) demonstrated marked differences in nerve fiber counts (P = .002, Mann Whitney U test). Two distinctive patterns of reinnervation that were observed: disruption of nerve bundles (collateral sprouting with microneuroma formation) and ingrowth around blood vessels (perivascular nerve fiber proliferation). There were increased numbers of microneuromas (groups 1 and 2 vs 3; P = .001, chi-squared test with Yates correction) and perivascular nerve fiber proliferation (groups 1 and 2 vs 3; P = .008, chi-squared test with Yates correction) in the myometrium in chronic pelvic pain with, and without, endometriosis compared with the control group.

Conclusion: Nerve fiber proliferation and other features of reinnervation have been observed in the isthmic regions of uteri that were removed at hysterectomy for chronic pelvic pain with and without endometriosis. There were no quantitative differences between the groups with chronic pelvic pain and endometriosis. These observations provide an alternative explanation for the source of pain and other clinical symptoms in these clinical settings.

© 2005 Mosby, Inc. All rights reserved.

* Reprint requests: M. J. Quinn, MD, Department of Obstetrics & Gynaecology, Hope Hospital, Stott Lane, Salford, Manchester M6 8HD, UK. E-mail: quinnobgyn@aol.com

0002-9378/$ - see front matter © 2005 Mosby, Inc. All rights reserved. doi:10.1016/j.ajog.2005.05.035
Epidemiologic studies have found an annual prevalence of chronic pelvic pain of 38 per 1000 women that frequently is associated with other sensory pelvic problems including cystitis and irritable bowel syndrome.\textsuperscript{1-4} Endometriosis is found in association with chronic pelvic pain and may be present in as many as 10% of women aged 15 to 45 years.\textsuperscript{1} Laparoscopy demonstrates variable appearances on peritoneal surfaces in the pouch of Douglas in women with the typical symptoms of chronic pelvic pain, dysmenorrhea, dyspareunia, and subfertility; severe dysmenorrhea is the most consistent symptom in the advanced stages of the condition.\textsuperscript{5} The enigma of endometriosis remains the variable relationship between symptoms and the extent of ectopic endometrium at laparoscopy, when severe pain may be associated with minor endometrial deposits; however, extensive ovarian endometriomas may have no associated symptoms. Many women are found to have chronic pelvic pain without ectopic endometrium being detected at laparoscopy; convincing evidence for the attribution of pelvic pain to the ectopic endometrial deposits is limited. Treatment with gonadotropin-releasing hormone agonists provides symptomatic relief for women with chronic pelvic pain with or without endometriosis.\textsuperscript{6} One randomized, controlled trial of gonadotropin-releasing hormone agonists in women with chronic pelvic pain found that almost all women had improvements in pain scores with gonadotropin-releasing hormone agonists compared with placebo, although not all the women had endometriosis when laparoscopy was performed after the course of treatment.\textsuperscript{6,7} Chronic pelvic pain with or without endometriosis may require hysterectomy with bilateral salpingo-oophorectomy for effective control of symptoms; however, recurrence of pain after surgery is not uncommon.\textsuperscript{8}

Sampson originally proposed the theory of retrograde menstruation, where endometrium passes into the peritoneal cavity in a reverse direction along the Fallopian tubes.\textsuperscript{9} This process may be observed in asymptomatic women with normal menstrual cycles; an explanation of the reason that some women experience symptoms of endometriosis and other women do not has not been resolved.\textsuperscript{10} Defects in both cell-mediated and humoral immunity have been demonstrated, though they do not explain the rate of tissue deposition if endometrium were behaving as a transplanted graft.\textsuperscript{11} Inoculation of endometrial cells in the cornea has achieved successful implantation, and primary tissue injury along the needle track may have contributed to these results.\textsuperscript{12} Coelomic metaplasia, vascular spread, and genetic theories have also been proposed to explain endometriosis, although none have been substantiated fully.\textsuperscript{13-15}

We have proposed that processes of denervation succeeded by reinnervation cause chronic pelvic pain, dysmenorrhea, menorrhagia, dyspareunia, and subfertility in severe endometriosis.\textsuperscript{16} The nerve supply is delivered with the uterine arteries in the parametrial tissues at the level of the uterine isthmus and secondarily within the fascial supports, which include the uterosacral-cardinal ligament complex.\textsuperscript{17} These ligaments suspend the lower genital tract through variable insertions into the cervix and upper vagina that are demonstrated readily at laparoscopy.\textsuperscript{18} Asymmetric damage to the uterine supports is frequently observed in parous women with chronic pelvic pain that is associated with difficult intrapartum episodes and typical features of reinnervation.\textsuperscript{19,20} This study investigates patterns of uterine innervation after hysterectomy for chronic pelvic pain with, and without, endometriosis.

**Methods**

**Patients**

Samples from the lower one half of the uterus of 16 consecutive patients with advanced endometriosis (revised American Fertility Society, grades III-IV) who underwent hysterectomy for endometriosis (group 1; mean age, 41.5 years [range, 27-53 years]; nulliparous, 8/16 samples; multiparous, 8/16 samples) were collected. The diagnosis of endometriosis was confirmed by histologic examination in 14 of 16 specimens. Eight uteri had incidental fibroid tumors, and 2 uteri had mild adenomyosis. Samples were also collected from 15 uteri that were removed at hysterectomy for chronic pelvic pain without endometriosis (group 2; mean age, 39.5 years [range, 31-46 years]; mean parity, 1.6 [range, 0-3]). Seven uteri had associated small, fibroid tumors (<3 cm), and 2 uteri had mild adenomyosis.

Control samples (group 3) were obtained from 2 sources: 17 parous uteri that had been removed for painless gynecologic indications (mean age, 42.0 years [range, 34-47 years]; mean parity, 2.4 [range, 1-4]; 4 for ovarian cysts, 4 during reconstructive surgery, 6 for menstrual dysfunction, 3 for miscellaneous indications) and 8 nulliparous uteri (mean age, 40 years [range, 30-52 years]; removed during surgery for single fundal fibroid tumors (4/8 samples) and for incidental reasons (4/8 samples) from the tissue archive. Tissue blocks were collected in the sagittal plane from the lower one half of the uterus to include both the endometrium and the serosal margin. All tissue samples were fixed in 10% phosphate buffered formalin then processed into paraffin wax. Ethics approval was obtained from the Local Research Ethics Committee, and each woman gave consent for the tissue samples to be studied.

**Immunohistochemistry**

Three-micron sections were cut from the tissue blocks and stained with hematoxylin and eosin and anti-S100...
protein. Deparaffinized sections were pretreated with trypsin (Difco 215230; DakoCytomation A/S, Glostrup, Denmark) before polyclonal anti-S100 protein (Dako Z0311; DakoCytomation A/S) was applied. Endogenous peroxidase activity was eliminated by the application of a commercial peroxidase blocking solution (Dako S2023). We stained the sections for S100 to detect nerves along with 2 positive controls on each slide. The sections were stained on an automated immunohistochemical stainer with detection kit (Dako K5001; ChemMate TM; Dako-Cytomation A/S). The nuclei were counterstained with hematoxylin.

Tissue analysis

Tissue sections were examined on a microscope with both semiquantitative and qualitative techniques. Immunoreactive nerve profiles were counted in 6 random grids that measured 1 mm² each, at least 2 mm apart. To ensure representative sampling across the section, myometrial sections were divided into inner and outer halves, and a total of 6 grids were counted, with 3 grids in each half of the section. The observers were blind to the source of the material and counted representative fields rather than areas of high nerve fiber density. Areas of artifact, or those areas that were associated with high vessel density, were avoided along with foci of adenomyosis or leiomyomas. To ensure consistency, similar random fields from the same section were re-counted by a second observer. Further blocks were taken from the upper half of the uterus in a subset of cases to exclude reinnervation at this site.

Results

Increased numbers of nerve profiles were observed in the myometrium of the lower half of the uterus in endometriosis (group 1 vs group 3;  \( P = .0013 \), Mann Whitney \( U \) test) and chronic pelvic pain (group 2 vs group 3;  \( P = .04 \) ) compared with controls (Table; Figures 1-4). There were no significant differences in the nerve counts between these 2 groups (group 1 vs group 2;  \( P = .35 \), Mann Whitney \( U \) test).

A comparison of uteri from women with chronic pelvic pain with and without endometriosis (groups 1 and 2) with control uteri (group 3) demonstrated clear differences in nerve fiber counts ( \( P = .002 \), Mann Whitney \( U \) test; Table). In addition, increased numbers of microneuromas (Figures 3 and 4) were observed in the combined group (groups 1 and 2), compared with the control group (group 3; 16/31 vs 2/25;  \( P = .001 \), chi-squared test with Yates correction). Perivascular nerve
fiber proliferation (Figures 3 and 4) was also more frequent in chronic pelvic pain with or without endometriosis (22/31 vs 8/25; \( P = .008 \), Fisher's exact test).

Observations in the uterine cervix showed some features of reinnervation, although there were no differences in nerve counts among the 3 groups. There were no signs of abnormal innervation in the uterine fundus.

**Comment**

Nerve fiber proliferation in the lower half of the uterus has been observed in women with chronic pelvic pain with and without endometriosis that may contribute to clinical symptoms in both groups. There were no quantitative differences in nerve counts among the 3 groups. There were no signs of abnormal innervation in the uterine fundus.

Nerve fiber proliferation (Figures 3 and 4) was also more frequent in chronic pelvic pain with or without endometriosis (22/31 vs 8/25; \( P = .008 \), Fisher’s exact test).

Observations in the uterine cervix showed some features of reinnervation, although there were no differences in nerve counts among the 3 groups. There were no signs of abnormal innervation in the uterine fundus.

**Table** Differences in myometrial innervation in the lower one half of the uterus in endometriosis (group 1), chronic pelvic pain without endometriosis (group 2), and control subjects (group 3)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group 1 (n = 16)</th>
<th>Group 2 (n = 15)</th>
<th>Group 3 (n = 25)</th>
<th>( P ) value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nerve fiber profiles per 0.5 mm(^2)</td>
<td>31.5 (17-53)</td>
<td>22 (14-66)</td>
<td>12 (6-29)</td>
<td>.002, Mann Whitney</td>
</tr>
<tr>
<td>Microneuromas (n)</td>
<td>9</td>
<td>7</td>
<td>2</td>
<td>.001, chi-squared, Yates correction</td>
</tr>
<tr>
<td>Perivascular nerve fiber proliferation (n)</td>
<td>13</td>
<td>9</td>
<td>8</td>
<td>.008, Fisher’s exact test</td>
</tr>
</tbody>
</table>

\( ^* \) Groups 1 and 2 versus 3.

\(^1\) Data are given as median (interquartile range).

Perivascular nerve fiber proliferation (Figures 3 and 4) and perivascular nerve fiber proliferation (Figures 3 and 4), both of which were more common in the groups with chronic pelvic pain.

Traumatic disruption of nerve bundles initiates processes of axonal regeneration along the line of the nerve bundle results in the appearance of collateral sprouting, which has also been termed microneuroma formation.\(^{21,22}\) Varying degrees of collateral sprouting have been reported in the uterus and attributed to the traumatic effects of parturition.\(^{21}\) In this study, we found microneuromas in the myometrium that were typical of the appearances after traumatic injury.\(^{23}\) Nerve fiber proliferation was seen to varying degrees around the circumference of the vessel and, in severe cases, in concentric layers. This observation has been reported previously in the vulva\(^{24}\) and may contribute to cyclic, gynecologic symptoms in which increases in blood flow in the second half of the menstrual cycle are associated with the severe dysmenorrhea that may respond to treatment that reduces pelvic blood flow.\(^{6,8}\)

Perivascular nerve fiber proliferation may represent the consequences of injury to branches of the uterine
Patterns of reinnervation in a retrospective survey of stored uteri have been reported, though there are no other quantitative studies for comparison. It is important to note that the quantitative analysis in this report refers to the mean nerve count in random fields by blinded observers; peak nerve counts may have emphasized the differences although the precise orientation of some of the tissue blocks was unknown (Figures 3 and 4). Both adenomyosis and fibroid tumors were recorded in the study groups (groups 1 and 2); for the most part, these were incidental pathologic findings that were avoided during the quantitative analysis and did not contribute to differences in myometrial nerve counts. The selection of nulliparous control samples in this series was limited because it is relatively unusual to remove nulliparous uteri. Preliminary studies within our group confirm that uteri that have been removed for single large fibroid tumors (4/8 nulliparous control samples) have normal innervatory patterns; however, 2 nulliparous uteri that had been removed for menstrual disorders showed minor increases in nerve fiber counts. There were several uteri with increased nerve fiber counts in the parous control samples that may have been caused by previous vaginal delivery. In the endometriosis group, 2 parous uteri were removed in association with endometriotic cysts. Myometrial nerve counts were similar to parous control samples, which suggests the possibility of a different cause for ovarian endometriomas in these circumstances. Each of these 3 features would serve to reduce the quantitative and qualitative differences between the control and study groups.

This study was not large enough to differentiate the effects of parity, although these preliminary observations suggest that perivascular nerve fiber proliferation was more frequent in nulliparous uteri and that microneuromas were more frequent in parous uteri from women with chronic pelvic pain. Several nulliparous women in this study gave a history of prolonged constipation although this observation was not examined formally in this study; most of the parous subjects had difficult intrapartum episodes in their first labor. That these histologic findings were present occasionally in the control group suggests a universal pattern of this kind of injury, such as excessive abdominal straining that may occur during daily activities. Injury to the parametrial tissues and uterosacral-cardinal ligaments during the second stage of labor may account for injuries to nerve bundles in parous women. Prolonged maternal voluntary efforts that complicated malpositions, big babies, and operative vaginal delivery may be significant in this respect.

Successful implantation of endometrium has followed needle inoculation in the cornea, and endometriosis is also found in abdominal and perineal incisions after cesarean delivery and episiotomy, respectively. Adherence of ectopic endometrium may reflect its availability at the time of tissue injury. The reason that some women with chronic pelvic pain have endometriosis and other women do not, may be attributable to the availability of ectopic endometrium at the time of the injury. Avulsion of the uterine supports during single intrapartum episodes or recurrent abrasions and petechial hemorrhages to the uterosacral ligaments that were caused by prolonged constipation may provide injured tissue surfaces. The choice of breast or bottle feeding may influence the availability of endometrium during tissue repair in the early puerperium. These observations may explain the positive response to treatment with gonadotropin-releasing agonists in women with chronic pelvic pain with or without endometriosis. Disruption of uterine innervation also interrupts normal fundocervical polarity of uterine contractility and may promote retrograde menstruation. The spectrum of intrapelvic manifestations of the condition may be further extended by pathologic processes that cause tissue injury without accompanying reinnervation, such as pelvic infection or serosal damage to small bowel in Crohn’s disease.

Nerve fiber proliferation has been reported in all female pelvic viscera and may account for common clinical presentations that include pelvic pain, dysmenorrhea, dyspareunia, vulval pain, and urinary and bowel urgency. Denervation that is caused by injuries to uterine neurovascular bundles and myofascial supports is succeeded by reinnervation that may provide an explanation for some forms of chronic pelvic pain that is associated with endometriosis. Retrograde delivery of ectopic endometrium to injured peritoneal surfaces may determine the varying laparoscopic appearances of different stages of the condition; processes of denervation-reinnervation in the uterine isthmus and myofascial supports may account for some clinical symptoms. If neurologic processes such as those described in this article contribute to the cause of the condition, then it may explain the response to treatment that reduces pelvic blood flow in the second half of the menstrual cycle. It may also have some bearing on the recurrence of pain after hysterectomy with bilateral oophorectomy or other surgery that is directed at the interruption of nerve pathways. In many women, ectopic endometrium may represent an epiphenomenon to underlying processes of denervation-reinnervation that may be responsible for persistent dysmenorrhea and chronic pelvic pain.

Acknowledgment

We thank Peter Clark for the immunohistochemical studies and Andy Vail for the statistical analysis.
References

Maternal complications with vaginal birth after cesarean delivery: A multicenter study

George A. Macones, MD, MSCE,a,b,c,* Jeffrey Peipert, MD, MPH,d Deborah B. Nelson, PhD,a,b Anthony Odibo, MD,a,b Erika J. Stevens, MA,a,b David M. Stamilio, MD, MSCE,a,b Emmanuelle Pare, MD,a,b Michal Elovitz, MD,a Anthony Sciscione, DO,e Mary D. Sammel, ScD,b Sarah J. Ratcliffe, PhDb

Departments of Obstetrics and Gynecology,a Biostatistics and Epidemiology,b and Leonard Davis Institute for Health Economics,c University of Pennsylvania, Philadelphia, PA; Department of Obstetrics and Gynecology, Women and Infants Hospital, Brown Medical School,d Providence, RI; Drexel University School of Medicine,e Philadelphia, PA

Received for publication January 5, 2005; revised March 1, 2005; accepted April 1, 2005

Objective: This study was undertaken to determine incidence and risk factors for uterine rupture in women attempting vaginal birth after cesarean delivery (VBAC) in a wide range of hospital settings.

Study design: We performed a case-control study nested within a cohort of women who have had a prior cesarean to determine the incidence and risk factors for uterine rupture in women attempting VBAC.

Results: The incidence rate of uterine rupture in those who attempt VBAC was 9.8 per 1000. A prior vaginal delivery was associated with a lower risk of uterine rupture (adjusted odds ratio [OR] = 0.40, 95% CI 0.20-0.81). Although prostaglandins alone were not associated with uterine rupture, sequential use of prostaglandin and pitocin was associated with uterine rupture (adjusted OR = 3.07, 95% CI 0.98-9.88).

Conclusion: Women with a prior cesarean should be offered VBAC, and women with a prior cesarean and prior vaginal delivery should be encouraged to VBAC. Although other studies have suggested that prostaglandins should be avoided, we suggest that inductions requiring sequential agents be avoided.

Key words: Vaginal birth after cesarean delivery, Uterine rupture, Tertiary care

A goal of Healthy People 2010 is to reduce both the primary and repeat cesarean delivery rates.1 One way to accomplish the latter would be to increase the proportion of women who attempt a vaginal birth after a prior cesarean. Unfortunately, the rate of vaginal birth after cesarean (VBAC) continues to fall in the United States, primarily because of concerns about complications such as uterine rupture.2-4
Prior studies outlining the incidence and risk factors for uterine rupture among a cohort of women attempting VBAC have come mainly from university/tertiary care centers with most studies reporting relatively few cases of uterine rupture. Thus, questions remain concerning the adequacy of the sample size and the generalizability of the study results. The goal of this study was to improve upon prior work and to assess the incidence and risk factors of uterine rupture in a cohort of who attempt VBAC, both in community and tertiary care hospital settings. We report the largest series of validated cases of uterine rupture in women who attempt VBAC, and focus on whether there are antepartum or early intrapartum predictors that can help guide patient counseling and clinical management.

Methods

We performed a multicenter, case-control study within a retrospective cohort (1996-2000) to assess maternal outcomes among women with a prior cesarean delivery. There were 17 participating hospitals in this study, 16 of which were in a defined geographic area of Southeastern Pennsylvania; the other was a large teaching hospital in Rhode Island. Given that a major goal of this study was to assess clinical outcomes in a mix of hospitals, we included both tertiary care hospitals and community hospitals (with and without obstetric/gynecology residency programs). Six of the included sites are tertiary, university hospitals and 5 of sites did not have a residency program in obstetrics and gynecology. Institutional Review Board approval was obtained from each hospital before the conduct of this study.

The participants included in the retrospective cohort portion of the study were delivered women with a history of a prior cesarean identified by an inclusive International Classification of Disease (ICD) code-based search at each of the participating hospitals. The ICD code search included the term “prior cesarean delivery, delivered” which, necessarily included both women who had an attempt at VBAC as well as those women who underwent an elective repeat cesarean delivery. The sensitivity of this ICD code-based search had been validated in several pilot studies that predated the start of this study. The medical records from this ICD-based search were requested from each of the participating institutions.

A team of trained nurse abstractors reviewed each medical record for this cohort, using standardized, closed-ended data collection forms. Approximately 3% of records requested were never found (despite multiple requests). During this initial review, each medical record was reviewed briefly (approximately 20 minutes) and information concerning the demographic, obstetric/medical history, and clinical outcomes (VBAC attempt/elective repeat cesarean, success/failure, rupture, major complications) for each woman was abstracted. The purpose of this brief review was to identify the subset of women who attempted VBAC, from which cases of uterine rupture and controls would be identified (Figure). We excluded subjects with either an unknown prior cesarean scar as well as women with a prior classical uterine incision. At the start of the study, and at several points during the review period, the abstractors underwent training to further ensure data validity. Using this strategy, we identified cases of uterine rupture among women with a prior cesarean attempting VBAC from each of the 17 participating hospitals. Because uterine rupture, the outcome of interest, can be confused with an asymptomatic dehiscence of the prior scar, we defined uterine rupture a priori as separation of the uterine scar (determined at laparotomy); immediately preceded by either a nonreassuring fetal heart rate pattern (determined by the treating obstetrician) or by signs/symptoms of acute maternal bleeding (SBP <70 mm Hg, DBP <40 mm Hg, HR >120) or by the presence of blood in the maternal abdomen at the time of laparotomy. All possible cases of uterine rupture were reviewed by the principal investigator (G.M.) to be certain that the classification was correct. Controls were randomly selected from the set of women who attempted VBAC but did not experience a uterine rupture. This random selection was accomplished by using a random numbers generated sequence applied to subjects who attempted VBAC but did not meet the case definition.

Among the subset of cases and selected controls, the inpatient records were reabstracted in detail by research nurses trained specifically for this more complex data collection. We were interested in a variety of types of potential predictors for uterine rupture, including patient demographics, obstetric history, medical history, and social history. We were also specifically interested in examining several pregnancy complications as potential risk factors for uterine rupture, including gestational diabetes and preeclampsia. Detailed information on the process of labor was also collected, with specific interest in whether labor occurred spontaneously, was induced, or augmented. We also collected information on medications for cervical ripening and induction/augmentation of labor, such as pitocin, prostaglandins, or Foley bulbs.

The data from the case and control records were then entered into a relational database, with frequent quality assurance procedures implemented to ensure quality data entry. Data analysis was performed in several steps. For descriptive purposes, comparisons of demographic and historical factors between women with a prior cesarean who attempted VBAC and women with a prior cesarean who opted for elective repeat cesarean were performed with standard bivariate techniques. Major and minor complications between these 2 groups were expressed as relative risks (unadjusted and 95%
For this study, the primary analysis was between the group of women with a prior cesarean attempting VBAC with a uterine rupture (cases) and the group of women with a prior cesarean attempting VBAC without a uterine rupture (controls). First, descriptive statistics were conducted to explore the risk factors for uterine rupture. Second, baseline characteristics of cases of uterine rupture and controls were compared, by using unpaired \( t \) tests (for normally distributed continuous variables), Mann-Whitney \( U \) tests (for non-normally distributed variables), and \( \chi^2 \)/Fisher exact for categorical variables. The results from these bivariate analyses were used to select variables for our multivariable logistic model for uterine rupture. Backward selection was used to reduce the number of variables, provided that removing the variable did not greatly affect any of the remaining estimates. Potential confounders were included based on their known, or suspected relationship, with uterine rupture. These variables were included regardless of their statistical significance. Indicator variables for the study sites were included in all comparisons.

We had specific interest in induction/augmentation of labor (and specific medications for this purpose) and uterine rupture. Because of the complex relationship between the type of labor and medications used, 3 models were developed. In 1 model, labor was coded as “spontaneous, induced, augmented,” which ignores the specific medication used. In the second model, we assessed the specific medication used (none, prostaglandin, pitocin, both prostaglandin and pitocin)—this does not account for whether labor was induced or augmented (ie, some patients may have labor induced with pitocin alone). The third model accounts for both the specific agent and whether labor was induced/augmented/spontaneous.

A priori, we performed a sample size calculation to determine the number of subjects required for the nested case-control portion of the study. We made the following assumptions: 2-sided alpha level (type I error) of .05, a beta level (type II error) of .2 (power of 80%), and a control/case ratio of 5:1. To detect an odds ratio (OR) of 2.0 or greater for risk factors with a prevalence of greater than 15%, 134 cases of uterine rupture will be required.

### Results

We reviewed the records of 25,005 women with a prior cesarean section, of which 13,706 (53.7%) underwent a VBAC attempt and 11,299 (44.3%) underwent an elective repeat cesarean (Figure). Fifty-nine percent of subjects were delivered at nonuniversity hospitals and 41% at university hospitals.

Patients with a prior cesarean section who attempt VBAC differ from those who opted for an elective repeat cesarean (Table I). Women who attempt VBAC tended to have fewer prior cesarean sections, fewer prepregnancy...
medical problems, and fewer antepartum complications. There was a statistically significant, but clinically unimportant, difference in the mean gestational age at delivery and mean birth weight between those who attempt VBAC and those who received repeat elective cesareans. Among those women who attempted VBAC (Figure), the vaginal delivery rate was 75.5%, and this was similar among the group of women attempting VBAC with a single prior cesarean (75.5%) and the group of women attempting VBAC with 2 or more prior cesareans (75.0%). These success rates are consistent with prior work.6,7

Major complications, defined as uterine rupture, bladder injury, or other major operative complications (bowel injury, uterine artery laceration)8 were more common among the women who attempt VBAC, whereas minor complications (blood transfusion, postpartum fever) were more common among the group of women who underwent elective repeat cesarean (Table II). We identified 134 cases of uterine rupture among the group of women who attempted VBAC, which yields a cumulative incidence of 9.8 per 1000 (95% CI: 8.1-11.4 per 1000). Uterine rupture was more common among women with 2 or more prior cesareans (200/1000) compared with women with only a single prior cesarean (87/1000).

The bivariate analysis of historical factors related to uterine rupture among women attempting VBAC suggested that nonwhite race was protective for uterine rupture (Table III). A prior vaginal delivery was strongly protective, whereas other obstetric historical factors were largely unrelated to uterine rupture among this group. There was a trend toward a positive association between the number of prior cesareans and the risk of uterine rupture. We found no maternal medical factors or social factors (smoking/drug use) associated with uterine rupture among women attempting VBAC. Delivery beyond 37 weeks in the index pregnancy was associated with a small increase in the rate of uterine rupture, though birth weight (categorized as <4000 g) was not associated with uterine rupture. Compared with the group of patients attempting VBAC who had spontaneous labor, both women who had either induced or augmented labors were more than 3 times more likely to experience a uterine rupture. However, medications used for labor stimulation (pitocin/prostaglandin) were not individually associated with uterine rupture, though when sequential prostaglandin-pitocin was used, there was an increase in risk. Importantly, only PGE-2 prostaglandins were used in subjects in this study.

The bivariate results were used to select variable for inclusion in the multivariable model to explore the independent factors related to uterine rupture among

| Table I Characteristics of women who attempt VBAC compared with those who undergo elective repeat cesarean section |

<table>
<thead>
<tr>
<th></th>
<th>VBAC attempt (n = 13,706)</th>
<th>Elective repeat cesarean section (n = 11,299)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal age (y)</td>
<td>30.2 (5.5)</td>
<td>31.5 (5.1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>56.8% (7785)</td>
<td>67.1% (7581)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Black</td>
<td>30.7% (4208)</td>
<td>22.7% (2565)</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>5.5% (754)</td>
<td>4.8% (543)</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>2.3% (315)</td>
<td>1.7% (192)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>4.7% (644)</td>
<td>3.7% (418)</td>
<td></td>
</tr>
<tr>
<td>Delivered at hospital</td>
<td>55% (7538)</td>
<td>65% (7344)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Prior spontaneous abortion</td>
<td>27.3% (3742)</td>
<td>29.7% (3356)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Prior elective abortion</td>
<td>21.2% (2906)</td>
<td>19.6% (2215)</td>
<td>.003</td>
</tr>
<tr>
<td>Prior vaginal delivery</td>
<td>36.3% (4975)</td>
<td>13.9% (1570)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2 or more prior cesarean deliveries</td>
<td>8.5% (1165)</td>
<td>32.3% (3650)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Chronic hypertension in index pregnancy</td>
<td>2.8% (384)</td>
<td>4.0% (452)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Gestational diabetes in index pregnancy</td>
<td>4.4% (603)</td>
<td>7.6% (859)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Preeclampsia in index pregnancy</td>
<td>2.6% (356)</td>
<td>3.1% (350)</td>
<td>.009</td>
</tr>
<tr>
<td>Asthma in index pregnancy</td>
<td>8.3% (1138)</td>
<td>8.4% (949)</td>
<td>.76</td>
</tr>
<tr>
<td>Preexisting diabetes</td>
<td>1.0% (137)</td>
<td>2.1% (237)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Self-reported tobacco use</td>
<td>18.7% (2563)</td>
<td>17.1% (1932)</td>
<td>.001</td>
</tr>
<tr>
<td>Self-reported cocaine use</td>
<td>3.8% (521)</td>
<td>2.4% (271)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Birth weight (g)</td>
<td>3334 (SD = 672)</td>
<td>3358 (SD = 706)</td>
<td>.009</td>
</tr>
<tr>
<td>Gestational age at delivery (wks)</td>
<td>38.6 (SD = .8)</td>
<td>37.9 (SD = 2.3)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

| Table II Major and minor morbidities comparing women attempting VBAC and women with an elective cesarean section |

<table>
<thead>
<tr>
<th></th>
<th>VBAC attempt</th>
<th>Elective repeat cesarean section</th>
<th>RR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major morbidity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uterine rupture</td>
<td>0.9%</td>
<td>0.004%</td>
<td>21.1 (8.6-51.5)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Bladder injury</td>
<td>0.4%</td>
<td>0.4%</td>
<td>1.05 (0.71-1.51)</td>
<td>.79</td>
</tr>
<tr>
<td>Other major operative injury</td>
<td>0.9%</td>
<td>0.6%</td>
<td>1.52 (1.14-2.02)</td>
<td>.003</td>
</tr>
<tr>
<td>Minor morbidity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blood transfusion</td>
<td>0.7%</td>
<td>1.2%</td>
<td>0.58 (0.45-0.75)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Postpartum fever</td>
<td>9.4%</td>
<td>13.0%</td>
<td>0.73 (0.68-0.78)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
women with a prior cesarean section attempting VBAC. After adjustment for confounding, the only historical factor significantly associated with uterine rupture was a prior vaginal delivery, which reduced the odds of rupture by 60% (Table IV).

Neither induction nor augmentation of labor was associated with uterine rupture, compared to women who labor spontaneously (model 1) (Table V). However, the analysis of labor stimulating agents (models 2 and 3) suggested that the risk of uterine rupture was increased only when both pitocin and prostaglandins were used for labor induction. In addition, we did not find any evidence of effect modification, when considering the relationship between induction/augmentation and uterine rupture, stratified by gestational age. The analysis was also unchanged after restricting the analysis to those cases and controls with only a single prior cesarean section.

### Table III  Bivariate analysis comparing women attempting VBAC with a uterine rupture and women attempting VBAC without a uterine rupture: Results from nested case-control study

<table>
<thead>
<tr>
<th>Tables risk factor (ref. group)</th>
<th>Case (n = 134) no. (%) or mean (median)</th>
<th>Control (n = 665) no. (%) or mean (median)</th>
<th>Odds of rupture</th>
<th>95% CI</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (y)</td>
<td>31.95 (33)</td>
<td>30.61 (31)</td>
<td>1.05</td>
<td>1.01-1.09</td>
<td>.010</td>
</tr>
<tr>
<td>Nonwhite</td>
<td>37 (27.8)</td>
<td>293 (44.0)</td>
<td>0.49</td>
<td>0.32-0.75</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Married</td>
<td>95 (71.9)</td>
<td>445 (69.9)</td>
<td>1.10</td>
<td>0.73-1.67</td>
<td>.647</td>
</tr>
<tr>
<td>Private/HMO</td>
<td>50 (37.5)</td>
<td>281 (42.2)</td>
<td>0.80</td>
<td>0.54-1.20</td>
<td>.225</td>
</tr>
<tr>
<td>Nonuniversity hospital</td>
<td>90 (67.1)</td>
<td>421 (63.3)</td>
<td>1.19</td>
<td>0.80-1.76</td>
<td>.397</td>
</tr>
<tr>
<td>Obstetric history</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 or more prior cesarean sections</td>
<td>22 (16.4)</td>
<td>79 (11.8)</td>
<td>1.46</td>
<td>0.87-2.44</td>
<td>.151</td>
</tr>
<tr>
<td>Prior term pregnancy</td>
<td>123 (92.4)</td>
<td>634 (95.3)</td>
<td>0.60</td>
<td>0.27-1.35</td>
<td>.187</td>
</tr>
<tr>
<td>Prior vaginal delivery</td>
<td>25 (18.7)</td>
<td>251 (37.7)</td>
<td>0.38</td>
<td>0.23-0.62</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Prior abortion</td>
<td>49 (36.8)</td>
<td>277 (41.6)</td>
<td>0.82</td>
<td>0.55-1.22</td>
<td>.352</td>
</tr>
<tr>
<td>Prior cesarean section term</td>
<td>59 (56.1)</td>
<td>115 (49.3)</td>
<td>1.32</td>
<td>0.83-2.09</td>
<td>.245</td>
</tr>
<tr>
<td>Prior cesarean section birth weight ≤4000 g</td>
<td>105 (81.4)</td>
<td>405 (71.9)</td>
<td>1.71</td>
<td>1.06-2.76</td>
<td>.029</td>
</tr>
<tr>
<td>Medical history</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chronic hypertension</td>
<td>5 (3.7)</td>
<td>50 (7.5)</td>
<td>0.48</td>
<td>0.15-1.22</td>
<td>.12</td>
</tr>
<tr>
<td>Diabetes</td>
<td>3 (2.2)</td>
<td>36 (5.4)</td>
<td>0.41</td>
<td>0.08-1.31</td>
<td>.13</td>
</tr>
<tr>
<td>Autoimmune disease</td>
<td>2 (1.5)</td>
<td>6 (0.9)</td>
<td>1.67</td>
<td>0.16-9.49</td>
<td>.53</td>
</tr>
<tr>
<td>Asthma</td>
<td>13 (9.7)</td>
<td>61 (9.1)</td>
<td>1.07</td>
<td>0.52-2.05</td>
<td>.83</td>
</tr>
<tr>
<td>Gestational diabetes</td>
<td>5 (3.7)</td>
<td>30 (4.5)</td>
<td>0.82</td>
<td>0.24-2.19</td>
<td>.69</td>
</tr>
<tr>
<td>Social Factors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking</td>
<td>22 (16.5)</td>
<td>140 (21.0)</td>
<td>0.74</td>
<td>0.43-1.23</td>
<td>.24</td>
</tr>
<tr>
<td>Cocaine use</td>
<td>2 (0.01)</td>
<td>21 (0.03)</td>
<td>0.47</td>
<td>0.05-1.98</td>
<td>.30</td>
</tr>
<tr>
<td>Alcohol use</td>
<td>7 (0.05)</td>
<td>36 (0.05)</td>
<td>0.98</td>
<td>0.36-2.31</td>
<td>.96</td>
</tr>
<tr>
<td>Current obstetric data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term delivery</td>
<td>61 (45.8)</td>
<td>240 (36.7)</td>
<td>1.46</td>
<td>1.00-2.13</td>
<td>.048</td>
</tr>
<tr>
<td>Gestational age at delivery (wks)</td>
<td>39.11 (39)</td>
<td>38.39 (39)</td>
<td>1.14</td>
<td>1.03-1.27</td>
<td>.011</td>
</tr>
<tr>
<td>Birth weight &gt; 4000 g</td>
<td>21 (17.3)</td>
<td>78 (12.9)</td>
<td>1.41</td>
<td>0.83-2.39</td>
<td>.203</td>
</tr>
<tr>
<td>Birth weight (kg)</td>
<td>3.52 (3.51)</td>
<td>3.35 (3.41)</td>
<td>1.46</td>
<td>1.07-1.99</td>
<td>.017</td>
</tr>
<tr>
<td>Intrapartum factors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labor type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spontaneous</td>
<td>22 (16.4)</td>
<td>273 (41.0)</td>
<td>1.0 (reference)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Induced</td>
<td>69 (51.4)</td>
<td>230 (34.7)</td>
<td>3.68</td>
<td>2.21-6.14</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Augmented</td>
<td>43 (32.0)</td>
<td>162 (24.4)</td>
<td>3.26</td>
<td>1.88-5.64</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Labor medications (none)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>68 (50.7)</td>
<td>430 (64.6)</td>
<td>1.0 (reference)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pitocin only</td>
<td>37 (27.6)</td>
<td>188 (28.2)</td>
<td>1.25</td>
<td>0.81-1.92</td>
<td>.325</td>
</tr>
<tr>
<td>Prostaglandin only</td>
<td>3 (2.24)</td>
<td>25 (3.7)</td>
<td>0.76</td>
<td>0.22-2.58</td>
<td>.659</td>
</tr>
<tr>
<td>Both pitocin and prostaglandin</td>
<td>26 (19.4)</td>
<td>22 (3.3)</td>
<td>7.47</td>
<td>4.01-13.93</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

**Comment**

The results of this study support several conclusions. First, women with a prior cesarean section who choose to attempt VBAC differ from those who opt for elective repeat cesarean. In general, women who attempt VBAC...
tend to have fewer preexisting medical problems and current pregnancy complications. Second, rates of major complications among those women who attempt VBAC are low, though higher than the women who opt for elective repeat cesarean. As reported previously and as confirmed in this larger study, found in this study, minor complications are more common among the group of women who attempt VBAC but are low, though higher than the women who opt for an elective repeat cesarean.8,9 We report the incidence of uterine rupture among women with a prior cesarean attempting VBAC was less than 1%, which is of importance because some have suggested that the occurrence of uterine rupture is on the rise in the United States. Third, the case-control portion of study aimed to identify predictors of uterine rupture among women with a prior cesarean section attempting VBAC. Unfortunately, this study demonstrated that there are few reliable predictors of this catastrophic event.

In a prior cohort study that used the Washington State Birth Events Database, Lydon-Rochelle et al10 found that induction of labor with prostaglandins increased the risk of uterine rupture more than 15-fold (RR [relative risk] = 15.6, 95% CI 8.1-30.0). There was no information on type of prostaglandin used in that study, and uterine rupture was identified from hospital discharge codes (as was prostaglandin use). Both of these are prone to misclassification. A recent American College of Obstetricians and Gynecologists Committee Opinion and Practice Bulletin discourages the use of prostaglandins based largely on these data.11,12 Our study, in which all subjects received intravaginal prostaglandin (not misoprostol), does not support the strong association between these agents and uterine rupture as suggested reported by Lydon-Rochelle et al. In fact, only the sequential use of prostaglandin and pitocin was associated with an increased odds of uterine rupture. Even with prostaglandins and pitocin, the odds of rupture was increased only 3-fold compared with those who labor spontaneously. Although somewhat limited by the possibility of a type II error (and a small number of cases of rupture in those induced with prostaglandins alone), our data suggest that, inductions requiring sequential prostaglandin-pitocin may be associated with an increase in risk. Thus, the previously reported association between the use of prostaglandins and uterine rupture may be a by-product of confounding by indication rather than a true relationship.

Consistent with other work, a vaginal delivery preceding a VBAC attempt protected against uterine rupture.13-15 We found a prior vaginal delivery was associated with a 60% reduction in the odds of rupture. Thus, it would seem reasonable to encourage women with a prior vaginal delivery to consider a VBAC attempt. Unfortunately, no other obstetric or historical factors were accurate predictors of uterine rupture.

Prior work on VBAC safety has come from mainly tertiary care institutions,14,16-20 An advantage of our study is that we have included tertiary and community hospitals, and those with and without obstetrics/gynecology residency programs, making our results more generalizable to a wider spectrum of obstetric patients. Another advantage of our analysis, compared with others on VBAC safety, is that all inpatient medical records were reviewed, rather than relying on ICD codes or birth certificates for both exposure and outcome information. Lastly, our study represents the largest series to date on uterine rupture, in which both exposure and outcome information were validated from records. Still, despite these strengths, our study has several limitations. First, given that all information was obtained from the inpatient record, some data are subject to misclassification. For example, information on substance abuse is based solely on patient report. We believe that such misclassification is likely to be non-differential, and would likely bias the results toward the null. Second, although we report 1 of the largest studies on VBAC safety, we still have limited power to assess

### Table IV Multivariate analysis: Historical risk factors for uterine rupture

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>Adjusted OR</th>
<th>95% CI</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal age (continuous)</td>
<td>1.09</td>
<td>1.03-1.15</td>
<td>.003</td>
</tr>
<tr>
<td>Nonwhite race</td>
<td>0.78</td>
<td>0.38-1.57</td>
<td>.48</td>
</tr>
<tr>
<td>Nonuniversity hospital</td>
<td>0.71</td>
<td>0.38-1.33</td>
<td>.28</td>
</tr>
<tr>
<td>Private/HMO insurance</td>
<td>1.08</td>
<td>0.59-1.99</td>
<td>.80</td>
</tr>
<tr>
<td>Prior vaginal delivery</td>
<td>0.40</td>
<td>0.20-0.81</td>
<td>.01</td>
</tr>
<tr>
<td>2 or more cesarean sections</td>
<td>1.45</td>
<td>0.64-3.27</td>
<td>.36</td>
</tr>
<tr>
<td>Delivery gestational age</td>
<td>1.13</td>
<td>0.97-1.30</td>
<td>.11</td>
</tr>
</tbody>
</table>

### Table V Multivariate analysis of labor type/medications and uterine rupture

<table>
<thead>
<tr>
<th>Model</th>
<th>Adjusted OR</th>
<th>95% CI</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spontaneous labor</td>
<td>1.0 (reference)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labor induction</td>
<td>1.01</td>
<td>0.43-2.34</td>
<td>.97</td>
</tr>
<tr>
<td>Labor augmentation</td>
<td>1.72</td>
<td>0.80-3.64</td>
<td>.32</td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spontaneous labor</td>
<td>1.0 (reference)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pitocin</td>
<td>0.77</td>
<td>0.32-1.83</td>
<td>.56</td>
</tr>
<tr>
<td>Prostaglandin</td>
<td>1.41</td>
<td>0.24-8.23</td>
<td>.70</td>
</tr>
<tr>
<td>Prostaglandin + pitocin</td>
<td>3.07</td>
<td>0.98-9.88</td>
<td>.05</td>
</tr>
<tr>
<td>Model 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spontaneous</td>
<td>1.0 (reference)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Augmented</td>
<td>1.61</td>
<td>0.76-3.40</td>
<td>.22</td>
</tr>
<tr>
<td>Induced without pitocin</td>
<td>0.85</td>
<td>0.23-3.15</td>
<td>.81</td>
</tr>
<tr>
<td>Induced with only  pitocin</td>
<td>1.46</td>
<td>0.60-3.57</td>
<td>.41</td>
</tr>
<tr>
<td>Induced with only  prostaglandin</td>
<td>1.90</td>
<td>0.37-9.65</td>
<td>.44</td>
</tr>
<tr>
<td>Induced with pitocin + prostaglandin</td>
<td>4.54</td>
<td>1.66-12.42</td>
<td>.003</td>
</tr>
</tbody>
</table>
possible risk factors of low prevalence. Third, our analysis focuses on maternal outcomes after VBAC, and does not consider neonatal outcomes. This is of relevance, because there is a recent report that suggests that the rate of hypoxic ischemic encephalopathy is increased in the newborn infants of women who underwent an elective cesarean compared with VBAC (although the absolute risk is quite small).

In this large, generalizable, observational study of maternal VBAC safety, we found that the overall incidence of uterine rupture in those attempting VBAC is quite low. Based on our data, we believe that women with a prior cesarean should be offered VBAC, and women with a prior cesarean and prior vaginal delivery should be encouraged to VBAC.

References
Increased intrauterine frequency of *Ureaplasma urealyticum* in women with preterm labor and preterm premature rupture of the membranes and subsequent cesarean delivery

Armin Witt, MD,a,* Angelika Berger, MD,b Christian J. Gruber, MD,c Ljubomir Petricevic, MD,a Petra Apfalter, MD,c,d Christof Worda, MD,a Peter Husslein, MDa

Departments of Obstetrics and Gynecology,a Pediatrics, Division of Neonatology, b Gynecologic Endocrinology and Reproductive Medicine,c and Clinical Microbiology,d Vienna Medical School, Vienna, Austria

Received for publication December 14, 2004; revised March 9, 2005; accepted March 30, 2005

**KEY WORDS**

*Ureaplasma urealyticum*

Preterm delivery

Cesarean delivery

Colonization

Amniotic cavity

**Objective:** The purpose of this study was to evaluate the prevalence of microbial invasion of the amniotic cavity at the time of preterm cesarean delivery for therapy-resistant preterm labor or preterm premature rupture of membranes, which are events that commonly are induced by infection, and to compare this group of patients with a group of patients who underwent preterm cesarean delivery for indications other than preterm labor or preterm premature rupture of membranes.

**Study design:** We studied 207 consecutive women between 23 and 34 weeks of gestation who underwent cesarean delivery. These patients were divided into 3 groups according to the indication for cesarean delivery: patients with preterm labor (group 1), patients with preterm premature rupture of membranes (group 2), and patients with other indications (group 3). In the course of the surgical procedure, amniotic fluid, amniotic membrane, and placental tissue specimens were collected for the detection of pathogens.

**Results:** *Ureaplasma urealyticum* was detected in 43.9% (58/132) of the patients of groups 1 and 2, with no significant difference between these 2 subgroups. In group 3, which served as the comparison group, *Ureaplasma urealyticum* was isolated in only 2.7% (2/75) of the patients. *Ureaplasma urealyticum* as a single pathogen was more frequent than all obligate pathogens together (43.9% vs 39.3%).

**Conclusion:** Our results provide evidence for an association between intrauterine colonization with *Ureaplasma urealyticum* and both therapy-resistant preterm labor and preterm premature rupture of membranes.

© 2005 Mosby, Inc. All rights reserved.
Microbial invasion of the amniotic cavity is present in up to one half of patients with preterm labor (PL) and preterm premature rupture of the membranes (pPROM). Ureaplasma urealyticum, a small prokaryote that belongs to the class of molecules, is a microorganism that is isolated from the vagina of 40% to 80% of women. Numerous previous reports proposed that microbial invasion of the amniotic cavity with U. urealyticum is associated with adverse pregnancy outcome. Additionally, there is evidence that colonization of the respiratory tract of premature infants with U. urealyticum favors the development of chronic lung disease, which is a correlation that has not been confirmed by other reports. Infection of the amniotic cavity by U. urealyticum also has been implicated in the development of postcesarean endometritis. The collection of amniotic fluid by amniocentesis is the most common method for the diagnosis of an intraterine colonization in cases of PL and pPROM. There are several studies about microbial invasion of the amniotic cavity that was diagnosed by amniocentesis that reported prevalences for U. urealyticum between 10% and 28%. However, in the case of pPROM, amniocentesis is not always feasible because of a lack of amniotic fluid.

Previous publications have shown a correlation between the detection of U. urealyticum in placental tissue and premature onset of labor and pPROM, with placentas that were obtained after vaginal delivery. Other authors found that, among women who underwent cesarean delivery after PL, only 50% with microbrial colonization of the chorioamnion also had the microbial colonization of the amniotic fluid. The purpose of this study was to evaluate the frequency of microbial invasion of the amniotic cavity at the time of preterm cesarean delivery for therapy-resistant PL or pPROM. As a new approach, we collected material from amniotic fluid, placenta, and amniotic membrane during the course of the surgical procedure. The intrauterine detection rates of U. urealyticum and other pathogens were compared with patients who underwent preterm cesarean delivery for other indications.

Material and methods

To improve the detection rate of intrauterine colonization, an expanded infection screening program was introduced at the Departments of Gynaecology and Obstetrics and Neonatology of the Vienna University Hospital in May 2001. All consecutive preterm parturients between 23 and 34 weeks of gestation who signed the informed consent and underwent preterm cesarean delivery between May 2001 and October 2003 were included in the screening program. Women with multiple pregnancy and women who were delivered vaginally were excluded from the study. It is the policy of our Department to perform cesarean delivery in the case of a preterm delivery at <34 weeks of gestation, only multiparous women with cephalic presentation are excluded from this procedure.

The patients were divided into 3 groups according to the indication for cesarean delivery: patients with therapy-resistant PL (group 1), patients with pPROM (group 2), and patients with other indications for preterm cesarean delivery (such as HELLP [hemolysis, elevated liver enzymes, and low platelet count], preeclampsia, intrauterine growth retardation, and pathologic cardiotocography or Doppler ultrasonography (group 3). Cases with PL plus pPROM were assigned to the pPROM group. In the case of pPROM, vaginal smears for bacterial culture were performed at the time of admission, if feasible. In the case of PL or other indications we did not perform smears routinely. Demographic and clinical data were obtained from the individual patient medical records.

All patients with pPROM (group 2) routinely received intravenous ampicillin at a dose of 3 × 4 g or, in the case of allergy, cefuroxime at a dose of 3 × 1.5 g for a maximum of 7 days. Following fetal indications, cesarean deliveries were performed after a median of 2 days of antibiotic treatment. Patients with PL (group 1) received antibiotic treatment only if clinical or laboratory signs of infection were present. Patients who underwent preterm cesarean delivery for other indications (group 3) did not receive antibiotics. Labor was not induced in any patient. In patients with PL, tocolytics (usually Atosiban [Tractocile]) at the recommended daily dosage were given in the case of objective labor (cervical length, <20 mm; a positive fibronectin test or ≥4 contractions per 30 minutes).

In the course of the surgical procedure, amniotic fluid was collected by puncturing the amniotic membrane and subsequent aspiration with a sterile syringe. After the delivery of the fetus, the placenta was expressed by cordtraction and placed on a sterile surface. An approximate 1-cm³ portion of the fetal side and 2 cm² of the membrane were excised with a sterile scalpel blade. This prospective cohort study was approved by the institutional review board of the Vienna Medical School. Informed consent was obtained with the patients’ admissions to hospital.

Transport media

Amniotic fluid was transported by means of the PORT A CUL transport system (Becton Dickinson Microbiology Systems, Cockeysville, MD), and amniotic membrane and placental tissue specimens were placed immediately into PORT A GERM transport media.
Culture media and detection of pathogens

Amniotic fluid was centrifuged at 4000g for 10 minutes at room temperature, and the pellet was used for further analysis. Placenta and amniotic membrane tissues were thoroughly homogenized by means of sterile glass tissue grinders and vortexed before being cultured for Mollicutes and for fast-growing aerobic and anaerobic bacteria.

Amnion fluid, membranes, and placenta were cultured for *Mycoplasma hominis* and *U urealyticum* with the commercially available ready-to-use urea- and arginine-containing broth-based detection system for urogenital mycoplasmas, Mycroscreen (International Microbio, Signes, France), and A7 agar medium (Heipha, Biotest, Heidelberg, Germany). Mycroscreen is used together with A7 agar (modified Shepard medium) for diagnostic purposes because it is a fast method that can differentiate commensal growth (≤10³ color-changing units [CCU]/mL) from those likely to be involved in an infectious process (≥10⁴ CCU/mL).

Briefly, 3 drops of amnion fluid or in the case of placenta and membranes, homogenized specimens (ground in brain heart infusion) each were inoculated simultaneously onto A7 agar medium and A3 medium (the base with which the arginine- and urea-containing trays of the microscreen test system are inoculated).

One hundred fifty microliters and 100 μL of A3 suspension were added in the urea wells and arginine well of Mycroscreen, respectively. One drop of *M hominis* supplement (which contained growth factors) was added to the arginine wells before paraffin oil was added to all wells to ensure anaerobic growth conditions. Trays were incubated at 35°C for 24 hours before the results were read by the change of color of the indicator in the medium from yellow to bright pink (for urea) or orange red (for arginine) because of the production of ammonia, which makes the medium alkaline. If the urea well changes to bright pink within 24 hours, *U urealyticum* is very likely to be present in the specimen at a level of ≥10⁴ CCU/mL; if the arginine well appears orange red within 24 hours, *M hominis* is likely to be present at a concentration of ≥10⁴ CCU/mL.

If the Mycroscreen test produced unequivocal or negative results, the agar medium A7 was checked after 48 hours (anaerobic conditions, 35°C) microscopically (phase contrast, ×100) for the growth of mycoplasma and ureaplasm, which both were identified by their characteristic morphologic colonies. In the case of growth onto A7 agar only, colony counts were estimated to be ≤10⁷/ml as recommended by the manufacturer. For data analysis, all positive specimens were taken into account.

In addition, all specimens were cultured by conventional microbiologic methods with Columbia agar that contained 5% sheep blood (35°C, carbon dioxide, 2 days), McConkey (35°C, aerobic, 1 day), chocolate agar (35°C, carbon dioxide, 2 days) for aerobic Gram-positive and Gram-negative bacteria and Schaedler (35°C, anaerobic, 6 days) and blood agar that was supplemented with kanamycin and vancomycin (35°C, aerobic, 2 days) for anaerobic bacteria. For culturing *Gardnerella vaginalis*, Gardnerella Gélose Agar (BioMérieux) was used (35°C, microaerophil, 2 days). In addition, brain heart infusion broth (35°C, 2 days) was used for enrichment culturing.

Microbial isolates were divided into the following 3 groups: (1) *M hominis/U urealyticum*; (2) obligate pathogens: group B streptococci, *Escherichia coli*, *Enterococcus faecalis*, *Staphylococcus aureus*, *Pseudomonas aeruginosa*, *Acinetobacter* spp, *Enterobacter* spp, *Prevotella* spp, *Bacteroides* spp, *G vaginalis*, *Candida* spp; and (3) bacteria considered to be skin contaminants: coagulase-negative staphylococci, alpha hemolytic streptococci, *Bacillus* spp, *Corynebacteria* spp, and anhemolytic streptococci.

Intra-amniotic colonization was defined as the presence of a positive culture result for group 1 or 2 pathogens or both in either amniotic fluid, placenta, and/or amniotic membrane.

Statistics

Chi-squared and, where appropriate, Fisher’s exact tests were used to assess the statistical significance of cross tabulations.

Results

During the study period, there were 595 preterm deliveries between 23 and 34 weeks of gestation. Two-hundred ninety-six of those women who were delivered vaginally; 90 women underwent cesarean delivery without being included in the screening program. Within the study group, 209 women underwent preterm cesarean delivery for the following indications: 39 women (18.8%) for therapy-resistant PL, 94 women (45.0%) for pPROM, and 76 women (36.4%) for other indications (8.0% HELLP syndrome, 33.3% preeclampsia, 14.7% intrauterine growth retardation, 29.4% pathologic cardiotocography or Doppler ultrasonography, and 14.6% for maternal indications [such as nephropathy, epilepsy]). Two of the 209 patients were excluded from the final analysis because no culture results were available. The median time interval between pPROM
and cesarean delivery was 2 days (range, 0-50 days); in the case of PL, the median time interval was 3 days (range, 0-18 days). Patient characteristics are given in Table I. All patients were white, primarily middle class, and between 18 and 40 years of age.

Microorganisms that were found in amniotic fluid, amniotic membrane, and placenta in groups 1 to 3 are presented in Table II. *U* urealyticum was detected in 43.9% (58/132 specimens) of patients with PL (group 1) and pPROM (group 2), with no significant difference found between these 2 subgroups (chi-squared test, $P = .56$). In all *U* urealyticum positive cases, the colony counts were $\leq 10^4$ per ml. *U* urealyticum as a single pathogen was more frequent than all obligate pathogens together (43.9% vs 39.3%) without significant association to other pathogens. In group 3, 2.7% of the patients had detection of *U* urealyticum (2/75 patients) within the amniotic cavity.

One hundred thirty-eight of 207 women had complete data sets of all sites, the missing data were due to the unfeasibility of collection, especially in the case of pPROM. As expected, we found a highly significant correlation between microorganisms that were detected in the amniotic fluid and in the placental/amniotic membrane compartments, respectively. Of women who were negative for *U* urealyticum in the amniotic fluid, 84% of the results (n = 100 women) were negative and 16% of the results (n = 19 women) were positive in the placental/amniotic membrane compartment ($P < .001$, by chi-squared test). Overall, 38.75% of the patients with PL and pPROM of membranes had intravaginal *U* urealyticum at the time of admission. An attempt at inferring intrauterine from vaginal findings was not reliable because we did not perform vaginal smears routinely in the case of PL.

With regard to the duration of antibiotic administration, we found no difference in the prevalence of intravaginal *U* urealyticum between 0 and 4 days of antibiotic administration. Notably, most patients (79/97, 81.4%) who required antibiotic treatment received antibiotics for $\leq 4$ days. Beyond 4 days of antibiotic treatment, the number of cases was too small for further calculation. The results of a subgroup analysis of patients who did not receive antibiotics are shown in Table III (which represents culture results of amniotic fluid and culture results of placenta + amnion).

### Table I  Patient characteristics (n = 207)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Group 1 (PL)</th>
<th>Group 2 (pPROM)</th>
<th>Group 3 (comparison group)</th>
<th>$P$ value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients (n)</td>
<td>39</td>
<td>93</td>
<td>75</td>
<td>.77</td>
</tr>
<tr>
<td>Age (y)</td>
<td>29 ± 5</td>
<td>30 ± 6</td>
<td>31 ± 6</td>
<td>.07</td>
</tr>
<tr>
<td>Birth weight (g)</td>
<td>1314 ± 612</td>
<td>1179 ± 350</td>
<td>1106 ± 466</td>
<td>.001</td>
</tr>
<tr>
<td>Gestational age (wk)</td>
<td>28.4 ± 3.0</td>
<td>28.3 ± 2.4</td>
<td>29.6 ± 2.3</td>
<td>.23</td>
</tr>
<tr>
<td>Duration of antibiotic treatment (d)</td>
<td>2 ± 1.6</td>
<td>3 ± 1.9</td>
<td>ND</td>
<td>.23</td>
</tr>
</tbody>
</table>

Values are given as mean ± SD; ND, Not done.
* By 1-way analysis of variance.

### Table II  Distribution of microorganisms within the amniotic cavity in relation to indication (n = 207)

<table>
<thead>
<tr>
<th>Microorganism</th>
<th>Group 1 + 2* (PL + pPROM)</th>
<th>Group 1* (PL)</th>
<th>Group 2* (pPROM)</th>
<th>Group 3 (Comparison group)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients (n)</td>
<td>132</td>
<td>39</td>
<td>93</td>
<td>75</td>
</tr>
<tr>
<td><em>U</em> urealyticum</td>
<td>35 (26.5%)</td>
<td>11 (28.2%)</td>
<td>24 (25.8%)</td>
<td>2 (2.7%)</td>
</tr>
<tr>
<td>Pathogenic</td>
<td>29 (22.0%)</td>
<td>9 (23.1%)</td>
<td>20 (21.5%)</td>
<td>7 (9.3%)</td>
</tr>
<tr>
<td><em>U</em> urealyticum and pathogenic</td>
<td>23 (17.4%)</td>
<td>6 (15.4%)</td>
<td>17 (18.3%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>Contamination</td>
<td>9 (6.8%)</td>
<td>2 (5.1%)</td>
<td>7 (7.5%)</td>
<td>17 (22.7%)</td>
</tr>
<tr>
<td>None</td>
<td>36 (27.3%)</td>
<td>11 (28.2%)</td>
<td>25 (26.9%)</td>
<td>49 (65.3%)</td>
</tr>
</tbody>
</table>

* Significant difference to group 3 (chi-squared; $P < .001$); no significance between group 1 and group 2.

Comment

To our knowledge, the present study is the first to explore the prevalence of microorganisms that were recovered from the amniotic cavity in patients who underwent cesarean delivery <33 weeks 6 days of gestation. The novel approach to sample intrauterine material is easily feasible and may be helpful for further neonatal management. It is a limitation of the study that the indication for cesarean delivery was the result of the Department’s policy in most cases. In settings with a different policy, these results may not be obtained, and the clinical relevance of this novel approach of sampling...
would be reduced. Although clear evidence for the optimal mode of delivery in this early gestational weeks is missing, there is a general trend towards primary cesarean delivery in our country. We are aware of the fact that this policy is not very common in other countries, but this discussion is not the topic of the present study. However, we found a significantly higher rate of amniotic cavity *U urealyticum* in patients with therapy-resistant PL or pPROM compared with patients who underwent cesarean delivery for other indications. *U urealyticum* represented the far most prevalent single intrauterine pathogen in these patients.

Previous publications have assessed placental tissues for the presence of *U urealyticum* in patients with preterm vaginal delivery; even though these reports found a strong correlation among microbial invasion, preterm delivery, and an increase in perinatal morbidity and mortality rates, the interpretation of these results is somewhat problematic because of the mode of delivery. After vaginal delivery, it is not possible to differentiate between the colonization of the amniotic cavity and vaginal cross contamination as a result of the birth process. Furthermore, cesarean delivery offers the possibility to detect microorganisms from all compartments of the amniotic cavity. The prevalence of microbial invasion of the amniotic cavity with *U urealyticum* in our study, which affected 43.9% of patients in groups 1 and 2, is much higher compared with previous studies that reported data from amniocentesis of patients with pPROM. One possible explanation for that difference is the combined sampling of amniotic fluid and specimens of placenta and amniotic membranes in our study. The latter 2 specimen types consist of more cellular material compared with amniotic fluid. Thus, the possibility of the detection of *U urealyticum* was probably higher. Furthermore, in our setting, we optimized factors such as the transport media, the time between sampling, and the beginning of detection procedure.

We acknowledge that the fact that the exact location where portions of placental and amniotic membrane were excised was not defined is another limitation of the study. One can argue that the location could have been near the internal cervical os and that vaginal ascension therefore could be conceivable. At this point, it has to be acknowledged that the question “time of invasion” is not elucidated clearly. It is possible that the amniotic cavity became invaded as a result of PL or pPROM. The results of the current study do not indicate a significant difference in the intrauterine colonization of *U urealyticum* between patients with PL and pPROM. This finding weakens the theory of late ascension of microorganisms through ruptured amniotic membranes but favors the possibility of microbial ascension at the beginning of or during the early weeks of gestation. The assumption that colonization develops early in the course of pregnancy is also supported by studies that show a clear association between *U urealyticum* colonization and spontaneous abortion and by reports on adverse pregnancy outcome after amniotic fluid isolation of *U urealyticum* during genetic amniocentesis. Prophylactic erythromycin has not been shown to be of benefit, and this finding may well be due to the limited transplacental transfer of macrolide antibiotics.

With regard to the duration of antibiotic administration, we found no difference in the rate of intrauterine *U urealyticum* between 0 and 4 days of antibiotics, which

<table>
<thead>
<tr>
<th>Microorganism</th>
<th>Group 1: PL (n)</th>
<th>Group 2: pPROM (n)</th>
<th>Group 3: Comparison group (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amniotic fluid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients</td>
<td>25 (100%)</td>
<td>33 (100%)</td>
<td>70 (100%)</td>
</tr>
<tr>
<td><em>U urealyticum</em></td>
<td>5 (20%)*</td>
<td>3 (9.1%)</td>
<td>0</td>
</tr>
<tr>
<td>Pathogenic</td>
<td>1 (4%)</td>
<td>7 (21.2%)*</td>
<td>2 (2.9%)</td>
</tr>
<tr>
<td><em>U urealyticum</em> and pathogenic</td>
<td>2 (8%)*</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Contamination</td>
<td>2 (8%)</td>
<td>0</td>
<td>3 (4.3%)</td>
</tr>
<tr>
<td>None</td>
<td>7 (28%)</td>
<td>10 (30.3%)</td>
<td>49 (70%)</td>
</tr>
<tr>
<td>Not done</td>
<td>8 (32%)</td>
<td>13 (39.4%)</td>
<td>16 (22.9%)</td>
</tr>
<tr>
<td>Placenta/membrane</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients</td>
<td>25 (100%)</td>
<td>33 (100%)</td>
<td>70 (100%)</td>
</tr>
<tr>
<td><em>U urealyticum</em></td>
<td>8 (32%)*</td>
<td>10 (30.3%)</td>
<td>2 (2.9%)</td>
</tr>
<tr>
<td>Pathogenic</td>
<td>6 (24%)*</td>
<td>3 (9.1%)</td>
<td>4 (5.7%)</td>
</tr>
<tr>
<td><em>U urealyticum</em> and pathogenic</td>
<td>5 (20%)*</td>
<td>5 (15.2%)*</td>
<td>0</td>
</tr>
<tr>
<td>Contamination</td>
<td>1 (4%)</td>
<td>6 (18.2%)</td>
<td>14 (20%)</td>
</tr>
<tr>
<td>None</td>
<td>5 (20%)</td>
<td>9 (27.3%)</td>
<td>50 (71.4%)</td>
</tr>
</tbody>
</table>

* Significant to group 3 (P < .001); no significant difference between group 1 and group 2.
1 Significant to group 3 (P < .01); no significant difference between group 1 and group 2.
suggests that the short-term administration of ampicillin does not result in the selection of *U urealyticum*. We are aware of the possible interference of our data because of the routine administration of antibiotics in the case of pPROM, but it is unlikely that *U urealyticum* was selected if they were not present at the beginning of antibiotic prophylaxis; in the comparison group, *U urealyticum* was detectable in only 2.7%. Aside from that, in cases of PL, we found rather similar intrauterine detection rates, although those patients did not receive antibiotics routinely. To avoid the potential confounder “antibiotic administration,” we performed a subgroup analysis of patients who did not receive antibiotics. Furthermore, we split the culture results into amniotic fluid and placenta/membrane results. Although the number of patients was subsequently low, we found the same statistical significances between group 1 and 2 versus group 3 and no significant difference between groups 1 and 2 (even not between invasion rates of the amniotic fluid; *P* = .068).

Our observations are unrelated to the clinical significance of neonatal morbidity that is caused by *U urealyticum*, such as chronic lung disease, but it indirectly increases neonatal morbidity and mortality rates by decreasing the duration of pregnancy. The mechanism of *U urealyticum* colonization and its role in the cascade of inflammation has not been elucidated fully. In this context, several reports show an association of *U urealyticum* and elevated cytokines in the fetal, amniotic, and maternal compartments and the *U urealyticum*–induced production of proinflammatory cytokines by macrophages.

The approach of culturing amniotic fluid, placenta, and membranes is a simple and effective diagnostic measure, which allows microbial colonization to be diagnosed and adequate treatment for the neonate to be initiated early. Particularly in patients in whom amniocentesis cannot be performed, this method therefore may be associated with postpartum benefits for both mother and child. It is well-known that the diagnosis of early-onset infection, particularly the identification of the pathogens that are involved, remains a dilemma in preterm infants. The rate of culture-proven early onset sepsis is low in this patient population, although many neonates have clinical and laboratory evidence of early-onset sepsis. We therefore considered it an ideal compromise to identify microorganisms from the amniotic cavity of the mother to increase the detection rate of pathogens that are potentially responsible for neonatal infections. Additional research is being conducted currently to elucidate whether the assessment of infection parameters and culture results at the time of cesarean delivery can improve neonatal outcome.

In summary, this is the first study to demonstrate a significantly higher rate of amniotic cavity *U urealyticum* in patients who undergo cesarean delivery for PL or pPROM compared with patients who undergo preterm cesarean delivery for other indications. This finding supports numerous previous reports about the role of *U urealyticum* in the pathogenesis of preterm birth and can be interpreted as additional evidence. Regarding the true nature of the association that was found in this study, it is still not evident whether there is a causative role for *U urealyticum* colonization and preterm delivery or whether these 2 phenomena are merely the result of an obscure common factor. PL and pPROM constituted the major cause of preterm delivery in our collective. Any approach at preventing preterm delivery therefore has to focus on combating and controlling reproductive tract infection, given the association of PL and pPROM with *U urealyticum*.

References


Persistance of adverse obstetric and neonatal outcomes in monochorionic twins after exclusion of disorders unique to monochorionic placentation

Line Leduc, MD,* Larissa Takser, MD, PhD, Denyse Rinfret, RN

Maternal Fetal Medicine Division, Department of Obstetrics and Gynecology, Sainte-Justine hospital. Montréal, Québec, Canada

Received for publication June 19, 2004; revised March 16, 2005; accepted April 1, 2005

KEY WORDS
Perinatal outcomes in twins
Chorionicity
Preterm birth
Fetal growth restriction

Objective: This study was undertaken to assess obstetric and neonatal outcomes in dichorionic twins and monochorionic-diamniotic twins after exclusion of twin-to-twin transfusion syndrome and twin reversed arterial perfusion sequence.

Study design: Data from a tertiary center were collected in twin gestations between 1994 and 2002. Chorionicity was defined by standard echographic criteria and placental examination at delivery. Neonatal outcomes were compared between monochorionic and dichorionic gestations.

Results: This study included 503 women: 378 (75%) dichorionic and 125 (25%) monochorionic twin gestations. Monochorionic twin gestations had a higher risk of preterm deliveries between 30 and 34 weeks' gestation than pregnancies with dichorionic twins (P < .01). Monochorionic twins had a higher number of birth weight less than 10th percentile (P < .001) discordancy 25% or greater (P < .02), admission to neonatal intensive care unit (P < .03), and intraventricular hemorrhage grade 3 and 4 (P < .007) than dichorionic twins even after adjusting for gestational age.

Conclusion: Monochorionic diamniotic twins have a higher risk of perinatal complications than dichorionic twin gestations, even after exclusion of disorders unique to monochorionic placentation.

© 2005 Mosby, Inc. All rights reserved.
monochorionic twins. Although it has been reported that complications such as prematurity and impaired fetal growth have a role in the outcome of monochorionic and dichorionic twins, previous studies have included either small cohorts or severely discordant twins of various causes.4,15-18 Therefore, we undertook an observational hospital based cohort study to assess obstetric and neonatal outcomes in twin pregnancies in relation to chorionicity. Monoamniotic twins and complications unique to monochorionic placentation were excluded.

**Material and methods**

Data were prospectively collected in all twins monitored at our tertiary care center between April 1994 and January 2002. This database review was approved by our ethics board committee. Twin pregnancies complicated by TTTS, chromosomal and structural anomalies, monoamniotic twins, and intrauterine fetal death of 1 co-twin were not included in the analysis as well as any transfer to our center that occurred after 21 week’s gestation. The diagnosis of TTTS was made according to Quintero’s criteria19 and the presence of fetal echocardiographic findings suggesting this syndrome.20 Chorionicity was assessed by standard echographic criteria and by meticulous placental examination at delivery. A detailed level II ultrasound was performed on all patients between 16 to 20 weeks of gestation. From 20 weeks of gestation until delivery, all women had the same prenatal care with clinical visits at 1- to 2-week intervals, cervical examination at each visit after 24 weeks of gestation, and multivitamin supplementation. Serial ultrasound examinations were performed every 2- to 4-week interval depending of fetal growth, for fetal biometry, and amniotic fluid assessment. Doppler studies were undertaken as soon as fetal growth restriction was suspected. Fetal growth restriction was defined as an estimated fetal weight less than the 10th percentile according to the Canadian birth weight growth curve for twins.21 Discordance was defined as a 25% or greater difference in birth weights and was calculated by using the following formula: larger twin weight – smaller twin weight 

\[
\text{Discordance} = \frac{\text{Larger twin weight} - \text{Smaller twin weight}}{\text{Larger twin weight}} \times 100
\]

All women had a maternity work leave from 24 weeks of gestation until delivery. Delivery was indicated between 32 to 34 weeks of gestation when fetal growth restriction less than the third percentile was present in at least 1 co-twin, or at 39 weeks’ gestation for postdates or for any standard obstetric indications. Preeclampsia was defined according to the Canadian classification.22 All pregnant women were screened for gestational diabetes mellitus (GDM) between 24 and 28 weeks’ gestation. The screening for GDM was a 1-hour plasma glucose measurement after a 50-g glucose load given at any time of the day. If the 1-hour plasma glucose was 10.3 mmol/L or greater, GDM was confirmed. If the 1-hour plasma glucose was 7.8 to 10.2 mmol/L, a 75-g oral glucose tolerance test (OGTT) was conducted.23

Maternal demographic and obstetric data as well as neonatal data were recorded. Perinatal mortality was reported as the neonatal death in the first month of life. Perinatal morbidity was documented with the following variables: the proportion of deliveries that occurred 34 or less, 32 or less, and 30 or less weeks of gestation, the proportion of twins with a birth weight discordance 25% or greater, a birth weight less than the 10th percentile, and a very low birth weight (VLBW) less than 1500 g. Neonatal morbidity was defined by the proportion of neonates admitted to the neonatal intensive care unit (NICU), the rate of endotracheal intubation, and the rate of intraventricular hemorrhage (IVH) grade 3 or 4. Obstetric morbidity was described in using the following parameters: the rate of elective cesarean section and urgent cesarean section on the second twin. Finally, maternal morbidity was assessed with the prevalences of gestational diabetes requiring insulin and preeclampsia. Perinatal outcomes in monochorionic twins were compared with those of dichorionic twins. Finally, discordant twins were compared according to chorionicity.

Statistical analysis was performed in using unpaired t test for continuous variables such as: mean maternal age and gestational age at delivery. The \( \chi^2 \) test was performed for the following categorical variables: the rates of elective and urgent cesarean section, the rate of both vaginal deliveries, the incidence of preeclampsia and gestational diabetes, the sex of fetuses, the rate of preterm deliveries less than 34 weeks, 32 or less weeks, and 30 or less weeks.

Mixed procedure (SAS version 8.02, SAS Institute, Cary, NC)24 for repeated measures was used to estimate the differences in birth weight between monochorionic and dichorionic pregnancies taking into account within pair covariance. To compare the rate of discordant (more than 25% discordance) twin pairs, and birth weight less than 2500 g, \( \chi^2 \) test was performed. To compare the rate of discordant (more than 25% discordance) twin pairs, and birth weight less than 2500 g, \( \chi^2 \) test was performed.

### Table I

<table>
<thead>
<tr>
<th>Variables</th>
<th>DC</th>
<th>MC</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean maternal age (y)</td>
<td>30.6 ± 5.3</td>
<td>29.1 ± 6.1</td>
<td>.01</td>
</tr>
<tr>
<td>Gestational age at 1st US</td>
<td>18.7 ± 3.6</td>
<td>18.6 ± 3.7</td>
<td>NS</td>
</tr>
<tr>
<td>Elective CS</td>
<td>168/378 (44.4%)</td>
<td>57/125 (45.6%)</td>
<td>NS</td>
</tr>
<tr>
<td>Urgent CS on fetus 2</td>
<td>21/210 (10%)</td>
<td>11/68 (16.2%)</td>
<td>.01</td>
</tr>
<tr>
<td>Both vaginal deliveries</td>
<td>189/210 (90%)</td>
<td>57/68 (83.8%)</td>
<td>NS</td>
</tr>
<tr>
<td>% Preeclampsia</td>
<td>25/378 (6.6%)</td>
<td>12/125 (9.6%)</td>
<td>NS</td>
</tr>
<tr>
<td>% Gestational diabetes</td>
<td>53/370 (14.3%)</td>
<td>15/122 (12.3%)</td>
<td>NS</td>
</tr>
</tbody>
</table>

US, Ultrasound; CS, cesarean section.
than 10th percentile or less than 1500 g for monochorionic and dichorionic pregnancies, the General Estimating Equations (GEE) modelling (GENMOD) procedure was used.\textsuperscript{25} Crude odds ratio (OR) was calculated and then gestational age at birth was introduced in GEE model setting for adjusted OR.

### Results

During the study period, 781 twin pregnancies were delivered at our center. Of these, 278 cases were excluded from the study because of the diagnosis of TTTS (n = 77), intrauterine fetal death of 1 or both fetuses (n = 31), chromosomal or structural anomalies (n = 20), monoamniotic twins (n = 21), or for transfer to our center after 21 weeks’ gestation (n = 106). Twenty-three twin pairs were lost to follow-up. The study included 503 sets of twins with 378 dichorionic (DC) and 125 monochorionic (MC) twin pairs that represent 75% and 25% of the cohort, respectively.

Table I reports maternal data. The mean maternal age is statistically different in both groups. The mean gestational age at time of the first ultrasound is the same. The rate of elective cesarean section is about the same irrespective of chorionicity but urgent cesarean section on the second twin happens more frequently in MC twins.

Table II reports neonatal data. Overall, MC twins have an increased perinatal morbidity with an earlier mean gestational age at delivery. The proportion of deliveries equal or less than 34 and 32 weeks’ gestation are significantly higher in MC group. However, unexpectantly, the rate of preterm deliveries less than 30 weeks’ gestation did not vary significantly when chorionicity was taken into account. MC twins had systematically a lower birth weight with an average of 212 g or less for the female and 260 g or less for the male fetuses.

Table III shows the GEE analysis. The effect of monochorionicity on perinatal morbidity was still present for almost all studied variables and this, even after...
adjusting for gestational age at birth. However, the risk
to be VLBW or to be intubated does not vary with
chorionicity. The mortality rate was not greater for MC
twins.

Table IV details the perinatal outcome in discordant
twins according to chorionicity. Overall, discordant
twins do worse than concordant twins. However, dis-
cordant MC twins deliver at an earlier gestational age
with preterm birth 34 weeks or less and admission to the
NICU than discordant DC twins. When the GEE model
is applied (Table V) and results are adjusted for gesta-
tional age at birth, perinatal morbidity of discordant
MC twins is not greater than for discordant DC twins.
We do not observe proportionally more neonates with
low birth weight or VLBW, endotracheal intubation, or
IVH among the discordant MC twins. Finally, in our
series, discordance 25% or greater was a sign of growth
restriction less than the 10th percentile in 76% of cases
(19/25) but the positive predictive value still remains low
with only 48.7% (19/39) of cases that were detected with
this threshold.

**Comment**

Our study examined a cohort of twin pregnancies
followed at a single tertiary care center with a consistent
method of managing twins and providing appropriate
neonatal support. We compare the perinatal outcome in
DC and MC twins after exclusion of conditions unique
to MC twins such as TTTS and TRAP sequence. These
peculiar complications of MC twin gestation play an
important role in terms of morbidity and mortality in
twins. We observed a higher risk of perinatal morbidity
in MC diamniotic twins than DC twins, independent of
complications such as TTTS and TRAP sequence. Previous
studies performed on the same topic have either
included a small number of cases or considered all the MC
twin cases.7-11,26-28 These data support the
need for diagnosis of early chorionicity as well as close
sonographic surveillance independently by the diagno-
sis of diseases such as TTTS, TRAP sequence, and
monoamniocity.

The perinatal mortality was not different between the
MC and DC twins once TTTS and TRAP sequence have
been excluded. These observations may either be real or
insufficient number of patients may explain these results.
These data are different than previously reported,7-11,26-28
but these studies usually considered all MC twin cases.

In our study, the proportion of pregnancies with a
birth weight discordance 25% or greater was double that
of the MC group as well as the number of neonates less
than the 10th percentile. We observed a higher number of deliveries between 30 and 34 weeks in severely discordant MC twins. These results suggest an association between preterm delivery and severe discordance or fetal growth restriction. This association may result from increased obstetric interventions in cases with severe discordance, leading to prematurity. On occasion, discordance 25% or greater can be explained by the presence of a large for gestational age neonate rather than a growth-restricted fetus and therefore may lead to iatrogenic deliveries of appropriate growth fetuses. Moreover, as Blickstein et al previously reported, our cohort suggests that discordance was a sign of fetal growth restriction in approximately 50% of cases and therefore, whenever severe discordance is suspected, it remains crucial to exclude the presence of fetal growth restriction before undertaking any medical intervention.

Recently, Amaru et al published their results from a retrospective study, hospital-based cohort of twin gestations with 2 live births delivered at 24 weeks or later from 1992 to 2001. The authors tried to define whether discordance was an independent risk factor for adverse neonatal outcome and they concluded that discordance was not an independent risk factor for serious neonatal morbidity or mortality. These apparently controversial results are the consequence of different definition used in the 2 studies. First, we used twin norms to define appropriate for gestational age (AGA) and small for gestational age (SGA) rather than singleton norms, thus we may have included neonates who actually met the definitions of low birth weights because 76% and 65% of our MC and DC twins with a birth weight discordance 25% or greater were less than the 10th percentile; second, we used a threshold of 25% discordance rather than 20%, which more likely included a population of neonates at higher risk for SGA and adverse outcome. The differences in study design may explain the different results. Therefore, we suspect that the poor outcomes observed in our cohort more likely result from these neonates with fetal growth restriction and not from growth discordance itself. Discordant MC twins do not have a worse perinatal outcome than discordant DC twins in our setting. These findings are opposite to those of Victoria et al who reported greater morbidity in discordant MC twins. However, TTTS was not excluded from their cohort. These observations warrant further investigation to clarify whether iatrogenic prematurity or impaired fetal growth result in worse outcome.

Maternal morbidity was significant in women carrying twins for gestational diabetes and preeclampsia. Although these complications do not appear directly related to or associated with chorionicity, preeclampsia may result in preterm birth and occurs more frequently in twin pregnancy. Some authors have previously reported the same findings and prevalence in a cohort of twin pregnancies. Likewise, we documented a significant rate of gestational diabetes requiring insulin in our cohort. The prevalence of GDM is population specific and has been previously described to be higher than expected in the Canadian Founder people. The design of the study cannot, of course, address the issue of the incidence of GDM in twins because no control group was recruited in the singleton population.

Finally, the rate of cesarean section in the MC subgroup was significant. Most of them are the result of the choice of the attending physician when fetal growth restriction was documented. However, the incidence of urgent cesarean section in which vaginal delivery was planned is similar to what has been previously reported in the literature.

This study compared the perinatal outcome in DC and MC diamniotic twins and observed that MC twins have a higher risk of perinatal morbidity even after exclusion of complications unique to MC placentaion. These data support the need for early diagnosis of chorionicity as well as close sonographic surveillance of MC twin pregnancies, irrespective of the diagnosis of specific high-risk conditions such as TTTS, TRAP sequence, and monoamniocytosis.

References


Hypertensive disease in pregnancies complicated by systemic lupus erythematosus

Robert S. Egerman, MD,a,b Risa D. Ramsey, PhD,a Lu W. Kao, RN,a Jay J. Bringman, MD,a Andrew J. Bush, PhD,c Jim Y. Wan, PhDc

Departments of Obstetrics and Gynecology,a Internal Medicine,b and Preventive Medicine,c University of Tennessee Health Science Center, Memphis, TN

Received for publication December 13, 2004; revised March 8, 2005; accepted March 29, 2005

KEY WORDS
Pregnancy
Systemic lupus erythematosus
Preeclampsia
Hypertension

Objective: The purpose of this study was to assess the percentage of hypertensive disease in pregnancies complicated by systemic lupus erythematosus at a single institution.

Study design: We conducted a retrospective analysis of medical records between 1992 and 2003 of 68 pregnancies that were complicated by systemic lupus erythematosus from 48 parturients. Patients were categorized into 3 groups: no chronic hypertension (n = 49 women), chronic hypertension–no medication (n = 6 women) and chronic hypertension–treated (n = 13 women). Analyses of variance (with Tukey-Kramer adjusted follow-up evaluation) and chi-squared/Fisher’s exact tests were used for the analyses of continuous and categoric variables, respectively. Significance was defined by a probability value of ≤ .05.

Results: Chronic hypertension complicated 28% of systemic lupus erythematosus pregnancies. Mean systolic blood pressures at intake were significantly different between the normotensive and no chronic hypertension groups and between the chronic hypertension–no medication and chronic hypertension–treated groups; the differences in diastolic pressures reached significance only between the no chronic hypertension and the chronic hypertension–treated groups. Maternal age, gestational age at delivery, birth weight, lowest platelet count, and highest serum creatinine levels were similar between the hypertensive and the nonhypertensive groups. There were no differences in the percentage of aspirin or heparin treatments among the groups, but the percentage of the chronic hypertension–treated group who received steroids was significantly greater than the percentage of women who received steroids in the other 2 groups (P < .05). Preeclampsia developed in 23% of the no chronic hypertension pregnancies and in 32% of the hypertensive pregnancies (P = .54). When pregnancies that were treated with prednisone (n = 34 pregnancies) were compared with those pregnancies that were managed with other agents (n = 34 pregnancies), the percentages of preeclampsia were similar (26% and 24%, respectively; P = .78).

Conclusion: The percentage of parturients with systemic lupus erythematosus in whom preeclampsia develops is increased, regardless of the presence of underlying chronic hypertension. Prednisone therapy was not associated with a higher risk of preeclampsia in this series.

© 2005 Mosby, Inc. All rights reserved.

Presented at the 25th Annual Meeting of the Society for Maternal-Fetal Medicine, Reno, NV, February 7-12, 2005.

Reprints not available from the authors.

0002-9378/S - see front matter © 2005 Mosby, Inc. All rights reserved.
doi:10.1016/j.ajog.2005.03.073
Systemic lupus erythematosus (SLE) has a propensity to affect women of child-bearing age.\(^1\) Wide variations exist regarding the hazards of SLE on the parturient, and the course during pregnancy is difficult to portend.\(^2-4\)

Intuitively, patients with SLE with no renal involvement (as reflected by normal serum creatinine level), absence of an active urinary sediment (cellular casts, hematuria), and no underlying hypertension seem to have perinatal outcomes similar to that of the general population.\(^5\) We investigated whether normotensive patients with SLE had a lesser risk of the development of preeclampsia than those patients with SLE and underlying chronic hypertension.

**Material and methods**

After obtaining institutional review board approval from the University of Tennessee Health Science Center, we identified obstetric patients with an antecedent diagnosis of SLE by standard criteria through discharge diagnoses between the years 1992 and 2003. Sixty-eight pregnancies from 48 parturients were identified and studied retrospectively. Demographic variables included maternal age, race, a history of chronic hypertension, and treatment for chronic hypertension. Outcome variables included gestational age at delivery, birth weight, lowest platelet count, and highest serum creatinine level during pregnancy. Chronic hypertension was defined as systolic blood pressure of \( \geq 140\) mm Hg or diastolic blood pressure of \( \geq 90\) mm Hg as defined by the Joint National Committee on Detection, Evaluation, and Treatment of High Blood Pressure [JNC VI]) that occurred before the 20th week of gestation or, if blood pressure criteria were not met, by an a priori diagnosis of chronic hypertension under current treatment. Definitions of preeclampsia included either the development of proteinuria (\( > 300\) mg/24-hour collection) and systolic blood pressure of \( \geq 140\) mm Hg or diastolic blood pressure of \( \geq 90\) mm Hg in a patient with no underlying chronic hypertension that occurred after the 20th week of gestation, the development of HELLP (hemolysis, elevated liver function tests, low platelet count) syndrome as defined by Sibai,\(^6\) difficult to control hypertension that occurred in a patient whose condition was previously stable, or maternal signs of severe preeclampsia.

Patients were stratified into 1 of 3 categories: normotensive (No CHTN), underlying chronic hypertension that did not require treatment (CHTN-No Rx), and underlying chronic hypertension that required treatment (CHTN-Rx). Differences in means were assessed by analysis of variance (with Tukey-Kramer adjusted follow-up evaluation); differences in percentages were assessed by chi-squared/Fisher’s exact tests. Statistical significance was established by probability values of \( \leq .05\).

**Results**

Chronic hypertension complicated 28% of 68 pregnancies with SLE. Maternal age, gestational age at delivery, birth weight, lowest platelet count, and highest serum creatinine levels were similar between the hypertensive and the nonhypertensive groups (Table I). Interestingly, although the complications did not reach statistical significance, we noted a trend toward higher birth weights in the normotensive group when compared with the hypertensive groups, despite similar mean gestational ages at delivery.

Mean systolic blood pressures at intake were significantly different between the normotensive and CHTN-No...
Rx groups and between the CHTN-No Rx and CHTN-Rx groups, whereas the differences in diastolic pressures reached significance only between the No CHTN and the CHTN-Rx groups. Choice of antihypertensive agents that were used in this series were nearly equally distributed between calcium channel blockers (nifedipine), beta blockers (predominantly labetolol) and alpha-methyldopa. At delivery, significantly higher systolic and diastolic pressures were identified in the CHTN-Rx group compared with the other 2 groups. Of the 13 pregnancies in the CHTN-Rx group, more than a single antihypertensive agent was necessary in 2 pregnancies.

Maternal age, gestational age at delivery, birth weight, lowest platelet count, and highest serum creatinine levels were similar between the hypertensive and the nonhypertensive groups. There were no differences in the percentage of patients who received aspirin or heparin treatments among the groups (Table II); however, the percentage of the CHTN-Rx group who received steroid therapy was significantly greater than the percentage patients who received steroid therapy in the other 2 groups ($P < .05$). In comparing patients with No CHTN with all patients with hypertension (CHTN-No Rx and CHTN-Rx), preeclampsia developed in 23% of the No CHTN and in 32% of the hypertensive pregnancies ($P = .54$; Table III). No patients in the CHTN-No Rx group had preeclampsia, whereas 6 of 13 patients from the CHTN-Rx were diagnosed with preeclampsia. When comparing pregnancies that were treated with prednisone ($n = 34$ pregnancies) to those pregnancies that were treated with other agents ($n = 34$ pregnancies), the percentages of preeclampsia were similar (26% and 24%, respectively; $P = .78$; Table III).

Three pregnancy losses occurred among the 68 pregnancies (at 23, 24, and 31 weeks of gestation), for a perinatal mortality rate of 44 of 1000 births. Two losses were complicated by preeclampsia in the No CHTN group (losses at 23 and 31 weeks of gestation), and the remaining loss that was complicated by preeclampsia (24 weeks of gestation) was from the CHTN-No Rx group. The 2 early losses were from the same parturient whose pregnancy was complicated by antiphospholipid antibody syndrome. The prevalence of antiphospholipid antibody syndrome was unknown in this series because this information was not uniformly available.

### Table II Maternal treatment in patients with SLE

<table>
<thead>
<tr>
<th>Treatment</th>
<th>No CHTN (n = 49)</th>
<th>CHTN-No Rx (n = 6)</th>
<th>CHTN-Rx (n = 13)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirin</td>
<td>13 (26.5%)</td>
<td>1 (16.7%)</td>
<td>3 (23.1%)</td>
<td>NS</td>
</tr>
<tr>
<td>Heparin</td>
<td>13 (26.5%)</td>
<td>2 (33.3%)</td>
<td>3 (23.1%)</td>
<td>NS</td>
</tr>
<tr>
<td>Steroids</td>
<td>19 (38.8%)</td>
<td>3 (50.0%)</td>
<td>12 (92.3%)</td>
<td>&lt;.05</td>
</tr>
</tbody>
</table>

NS, Not significant.

### Table III Percentage of preeclampsia in pregnancies with SLE that were complicated by chronic hypertension (combined treatment and nontreatment groups) and the percentage of preeclampsia in pregnancies with SLE that were treated with steroids

<table>
<thead>
<tr>
<th>Complication</th>
<th>Preeclampsia (n)</th>
<th>Odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronic hypertension</td>
<td>Yes (n = 19)</td>
<td>6 (31.6%)</td>
</tr>
<tr>
<td></td>
<td>No (n = 49)</td>
<td>11 (22.5%)</td>
</tr>
<tr>
<td>Steroids</td>
<td>Yes (n = 34)</td>
<td>9 (26.0%)</td>
</tr>
<tr>
<td></td>
<td>No (n = 34)</td>
<td>8 (24.0%)</td>
</tr>
</tbody>
</table>

### Comment

SLE affects a multitude of organ systems with considerable variance in disease manifestation and progression between patients. The severity of renal involvement from SLE is a well-accepted indicator of debility. Hypertension that is associated with this disease either reflects or contributes to progressive renal dysfunction. Similarly, the contribution of underlying renal dysfunction on the severity of hypertension is apparent in advanced disease. Correspondingly, we sought to determine the role of SLE on hypertensive disease in pregnancy. In this retrospective analysis, patients with an antecedent diagnosis of SLE were stratified into 3 groups: normotensive, underlying chronic hypertension that did not require treatment, and underlying chronic hypertension that required treatment. We demonstrated an increased risk of preeclampsia whether the patient was normotensive (23%) or hypertensive (32%) in pregnancies that were complicated by SLE. Other investigators have reported a lower rate of preeclampsia, despite overt renal involvement from SLE (15%), which indicates that additional genetic or environmental factors are likely to be contributory.

The distinction of preeclampsia from a lupus flare is ambiguous many times. Maternal symptoms that are
consistent with previously known exacerbations (particularly joint involvement or fever) and laboratory indicators of autoimmune dysfunction (decreasing complement levels) are helpful. Also, superimposed preeclampsia can occur regardless of the overt clinical rheumatologic manifestations, which further encumbers an attempt at a distinct diagnosis. Criteria in this study regarding the development of preeclampsia were based on readily available data of maternal symptoms and signs and laboratory data. Difficult to control blood pressure, in a previously stable pregnant woman who receives antihypertensive therapy, was another clue of developing superimposed preeclampsia. Relying on discharge diagnoses and physician and nurses’ progress notes is a limitation of any retrospective analysis and recognizing the imperfection of this, we relied on the judgment of the treating physicians at the time in rendering the patients’ diagnoses.

Regarding outcomes in a subgroup of patients who required >1 antihypertensive agent, our number of patients in this category was too few to report meaningful results. Further, this retrospective review did not enable us to ascertain compliance among our patients who were receiving antihypertensive therapy.

In conclusion, SLE appears to contribute to an increased risk of preeclampsia, irrespective of the presence of an underlying maternal hypertensive disorder. Practitioners should remain vigilant during these pregnancies, given the unpredictability of this disease, and appropriate counseling should be offered to the patient during the periconceptional period.

References

Gestational age-specific predicted risk of neonatal respiratory distress syndrome using lamellar body count and surfactant-to-albumin ratio in amniotic fluid

Raymond Karcher, PhD,a Elizabeth Sykes, MD,a Daniel Batton, MD,b Zi Uddin, PhD,c Gary Ross, DO,d Elaine Hockman, PhD,e George H. Shade, Jr, MDf,*

Departments of Clinical Pathologya and Pediatrics,b William Beaumont Hospital, Royal Oak, MI; Department of Clinical Pathology,c St. John Detroit Riverview Hospital, Detroit, MI; Department of Quality Management,d St. John Oakland Hospital, Madison Heights, MI; Division of Computing and Information Technology,e Wayne State University; Department of Obstetrics and Gynecology, Sinai-Grace Hospital,f Wayne State University, School of Medicine, Detroit, MI

Received for publication March 8, 2005; accepted March 29, 2005

KEY WORDS
Lecithin/sphingomyelin ratio
Lamellar body count
Surfactant-to-albumin ratio (TDx-FLMII)
Amniotic fluid
Fetal lung maturity

Objective: This study was undertaken to study the statistical correlation between lecithin/sphingomyelin (L/S) ratio, percent phosphatidylglycerol (%PG), lamellar body count (LBC), and surfactant-to-albumin ratio (TDx-FLMII) in amniotic fluid (AF); and derive gestational age-specific (GA) predicted risk of neonatal respiratory distress syndrome (RDS) for LBC and TDx-FLMII.

Study design: AF specimens (238) were collected by transabdominal amniocentesis. L/S ratio, %PG, LBC, and TDx-FLMII were determined by established procedures. RDS diagnosis was ascertained by a neonatalogist, and statistical analyses were performed with the use of the SPSS software program (SPSS Inc, Chicago, Ill).

Results: Significant correlation was obtained among the 4 variables (L/S ratio, %PG, LBC, and TDx-FLMII). Independent linear regression analyses between L/S ratio versus LBC and TDx-FLMII provided acceptable correlation. Multiple regression analysis showed a significant (P < .001) contribution from TDx-FLMII and GA for predicting the L/S ratio. Receiver operating characteristic curve analysis provided the immature cutoffs (LBC = < 30.0 × 106/μL; TDx-FLMII = < 40.0 mg/g). Total accuracy (either positive or negative) for RDS was similar for LBC (75.5%) and TDx-FLMII (76.7%).

Conclusion: LBC and TDx-FLMII are equally accurate. GA-specific predicted risk of RDS by both tests significantly eliminated L/S ratio identified false positive cases of fetal lung maturity.

© 2005 Mosby, Inc. All rights reserved.

* Reprint requests: George H. Shade, Jr, MD, FACOG, FACPE, Department of Obstetrics and Gynecology, Sinai-Grace Hospital, Wayne State University, School of Medicine, Suite M541, 6071 W Outer Dr, Detroit, MI 48235.

E-mail: gshade@dmc.org

0002-9378/S - see front matter © 2005 Mosby, Inc. All rights reserved.
doi:10.1016/j.ajog.2005.03.080
It is of concern to the obstetrician to ensure fetal lung maturity (FLM) in the newborn infant to limit the risk of respiratory distress syndrome (RDS). Gluck et al first described the prenatal prediction of FLM on the basis of the analysis of the lecithin/sphingomyelin (L/S) ratio in amniotic fluid (AF). The percent phosphatidylglycerol (%PG) was later added to the L/S ratio, and these 2 tests are regarded as the “gold standard” of FLM in the neonate.

The L/S ratio and %PG are time-consuming, expensive, and not readily available around the clock. These limitations are of concern and inconvenience to obstetricians. Several attempts were made to replace the L/S ratio and %PG tests with more readily available tests. The purpose of this study was to evaluate the 2 currently readily available tests, ie, lamellar body count (LBC) and surfactant-to-albumin ratio (the TDx-FLMII), as potential replacements for the L/S ratio and %PG.

Gluck et al first established FLM by a yes/no (Boolean) result; however, Tanasijevic et al observed that the rising rate of cesarean sections is a result of physician and hospital behavior rather than inherent patient characteristics. Recently McElrath et al published the GA-specific predicted risk of RDS using the currently used second-generation assay named TDx-FLMII.

Materials and methods

AF specimens (238) were procured from participating hospitals (February 2003-June 2004). All the infants were delivered within 48 hours of amniocentesis, except in 7 cases where repeat amniocentesis was performed 4 to 7 days later. The hospital’s human investigation committee approved the project, and it was carried out without the knowledge of the obstetricians and neonatologists involved. Although the overall rate of cesarean sections for uncomplicated deliveries at our associated hospitals is 32%, in these higher risk cases, 65% of the infants were delivered by cesarean section. It has been recently reported that the rising rate of cesarean sections is a result of physician and hospital behavior rather than inherent patient characteristics.

The clinical diagnosis of RDS in the newborn infant was made by the presence of respiratory distress (tachypnea, retractions, and/or nasal flaring) shortly after delivery and a persistent need for respiratory support (oxygen or positive pressure) for more than 24 hours along with a typical chest radiograph (reticular granular appearance of pulmonary parenchyma). Neither newborn pneumonia nor transient tachypnea was diagnosed as RDS. Maternal and neonatal characteristics are described in Table 1.

All AF specimens were centrifuged at 500g for 5 minutes before analysis. Physical appearance of the AF (color, presence of meconium or red blood cell count, or turbidity) was noted for each specimen, but this had no bearing on the statistical analysis because every specimen was assayed. The L/S ratio and %PG in AF were determined by 2-dimensional thin-layer chromatography and reflectance densitometry. This procedure is essentially the original Kulovich and Gluck method.

### Table 1 Maternal and neonatal characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Maternal</strong></td>
<td></td>
</tr>
<tr>
<td>Age (y)</td>
<td></td>
</tr>
<tr>
<td>18-29</td>
<td>72</td>
</tr>
<tr>
<td>30-39</td>
<td>155</td>
</tr>
<tr>
<td>40-45</td>
<td>11</td>
</tr>
<tr>
<td>GA* at amniocentesis (wk)</td>
<td></td>
</tr>
<tr>
<td>33-34</td>
<td>7</td>
</tr>
<tr>
<td>35-36</td>
<td>113</td>
</tr>
<tr>
<td>37-38</td>
<td>111</td>
</tr>
<tr>
<td>39-40</td>
<td>4</td>
</tr>
<tr>
<td><strong>Clinical characteristics</strong></td>
<td></td>
</tr>
<tr>
<td>Premature labor</td>
<td>4</td>
</tr>
<tr>
<td>Preeclampsia/ eclampsia</td>
<td>13</td>
</tr>
<tr>
<td>Premature rupture of membranes</td>
<td>2</td>
</tr>
<tr>
<td>Twin gestations</td>
<td>7</td>
</tr>
<tr>
<td>Breech presentation</td>
<td>11</td>
</tr>
<tr>
<td>Previous cesarean section</td>
<td>72</td>
</tr>
<tr>
<td>Elective cesarean section</td>
<td>151</td>
</tr>
<tr>
<td>Diabetes mellitus (type 1 or 2)</td>
<td>83</td>
</tr>
<tr>
<td>Gestational diabetes</td>
<td>9</td>
</tr>
<tr>
<td>Fetal distress</td>
<td>12</td>
</tr>
<tr>
<td>Hemoglobinopathy</td>
<td>0</td>
</tr>
<tr>
<td>Rh sensitivity</td>
<td>3</td>
</tr>
<tr>
<td>Substance abuse</td>
<td>0</td>
</tr>
<tr>
<td>Fetal malformations</td>
<td>4</td>
</tr>
<tr>
<td>Meconium-stained AF</td>
<td>3</td>
</tr>
<tr>
<td>Placenta previa</td>
<td>33</td>
</tr>
<tr>
<td>Cholestasis of pregnancy</td>
<td>8</td>
</tr>
<tr>
<td>IUGR</td>
<td>16</td>
</tr>
<tr>
<td>Hypertension</td>
<td>4</td>
</tr>
<tr>
<td><strong>B. Neonate</strong></td>
<td></td>
</tr>
<tr>
<td>No RDS†</td>
<td>207</td>
</tr>
<tr>
<td>RDS</td>
<td>13</td>
</tr>
<tr>
<td>No data</td>
<td>18</td>
</tr>
</tbody>
</table>

**IUGR**: Intrauterine growth rate.

† GA not determined accurately in 3 patients.

The total number of patients listed exceeds 238, because some patients had multiple clinical characteristics.

Seven patients had repeat amniocentesis.
predicted in this procedure, whenever the L/S ratio is greater than 2.0 and PG is present. However, because an L/S value of 2.0 still carries a small risk of RDS, we used a maturity cutoff of 2.5. A %PG of greater than 0.5 was used to indicate presence so this result could be evaluated statistically as a continuous variable. In 12 specimens, the %PG was not available because of procedural limitations.

LBC (number of lamellar bodies per microliter of amniotic fluid) was determined using the platelet channel on the Sysmex XE-2100 (Toa Medical Electronics, Los Alamitos, CA). TDx-FLM II results expressed as milligrams surfactant per gram albumin were obtained by using the TDx (Abbott Diagnostics, Abbott Park, IL).

Statistical analyses (receiver operating characteristic curve [ROC] analysis, analysis of variance, $\chi^2$, t test, and Levene’s test for equality of variances, linear, multiple, and multiple logistic regression) were preformed with the SPSS, version 12 (SPSS Inc, Chicago, IL).

## Results

LBC showed no statistical difference between spun (5 minutes at 500 g) and nonspun AF (paired $t = 0.006$, $n = 24$). For consistency, all AF specimens were centrifuged before the assay of the 4 parameters (L/S ratio, %PG, LBC, and TDx-FLM II). Results of all 4 variables were significantly higher in diabetic (80) than in nondiabetic (129) patients, and Levene’s test for equality of variances showed a greater variability in %PG ($P < .001$), and TDx-FLM II ($P < .001$) in the diabetic group as well. There was no significant difference in L/S variances. Significant correlation ($P < .001$) was observed among the 4 variables. Acceptable correlation from independent linear regression analysis was determined between L/S ratio versus TDx-FLM II, and LBC [a: TDx-FLM II = 13.26 + 8.9 L/S ($r = 0.79$, $P < .001$); b: LBC = 2.77 + 12.0 L/S ($r = 0.58$, $P < .001$)].

Multiple regression analyses presented significant contributions from TDx-FLM II and LBC in predicting the L/S ratio:

(a) $L/S = 0.032 \text{PG} + 0.066 \text{TDx-FLM II} - 0.131 \text{GA} + 5.456$

(b) $L/S = 0.088 \text{PG} + 0.016 \text{LBC} + 0.015 \text{GA} + 2.240$

For (a), $R = 0.806$, $P < .001$. Coefficients for both PG ($P = .021$) and TDx-FLM II ($P < .001$) are both significant. For (b), $R = 0.640$, $P < .001$. Coefficients for PG ($P < .001$) and LBC ($P < .001$) are both significant.

With the use of the “gold standard” (L/S ratio > 2.5 and %PG > 0.5 as cutoffs for FLM), ROC curve analysis provided values for immature cutoffs for LBC and TDx-FLM II: (a) LBC less than 30 $\times 10^3/\mu$L (area under the curve [AUC] = 0.966, sensitivity [SEN] = 90.9%, specificity [SPEC] = 89.7%); (b) TDx-FLM II less than 40.00 mg/g (AUC = 0.978, SEN = 93.1%, SPEC = 90.2%). These immaturity cutoffs of less than 30 $\times 10^3/\mu$L for LBC and less than 40.00 mg/g for TDx-FLM II provided a high index of agreement between the 2 categorical variables ($kappa = 0.68 \pm 0.054$, $t = 10.35$, $P < .001$). Total agreement was 86.8% (21.7% at risk and 65.1% not at risk of RDS). The 13.2% nonagreement between TDx-FLM II and LBC was evenly divided between the 2 patterns of disagreement. This indicated that both tests (LBC and TDx-FLM II) are equally reliable in the prediction of RDS.

With the use of the “gold standard” established in our laboratory for FLM cutoffs for L/S ratio (> 2.5) and %PG (> 0.5), and the ROC curve analysis-derived cutoffs from the present data for LBC (> 30 $\times 10^3/\mu$L) and TDx-FLM II (> 40.00 mg/g), the total accuracy values (either positive or negative) for RDS for the 4 tests in amniotic fluid were as follows: (a) L/S ratio = 86.8%, (b) %PG = 65.8%, (c) TDx-FLM II = 76.7%, and (d) LBC = 75.5%. For TDx-FLM II, SEN was 93.3%, SPEC 94.2%, positive predictive value 80.0%, and negative predictive value 98.3%. For LBC, the results for these 4 indices were as follows: 90.0%, 91.8%, 73.0%, and 97.4%, respectively. By using the same cutoffs (described previously) for the 4 parameters, and the clinical diagnosis of the newborn infant as determined by the neonatologist, we calculated the percentage of correct and incorrect predictions for these 4 laboratory tests for FLM (Table II).

GA-specific predicted risk of RDS (Table III) values expressed as percentages for TDx-FLM II were calculated from multiple logistic regression analysis: Probability

<table>
<thead>
<tr>
<th>Predictor With RDS</th>
<th>Without RDS</th>
<th>% Accurately predicted</th>
<th>% Inaccurately predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>TDx-FLM II</td>
<td>92.3% (12/13)</td>
<td>77.6% (159/205)</td>
<td>7.7% (1/13)</td>
</tr>
<tr>
<td>LBC</td>
<td>84.6% (11/13)</td>
<td>75.2% (155/206)</td>
<td>15.4% (2/13)</td>
</tr>
<tr>
<td>%PG</td>
<td>92.3% (12/13)</td>
<td>67.0% (128/191)</td>
<td>7.7% (1/13)</td>
</tr>
<tr>
<td>L/S</td>
<td>61.5% (8/13)</td>
<td>89.4% (168/188)</td>
<td>38.5% (5/13)</td>
</tr>
</tbody>
</table>

* Using FLM cutoffs of TDx-FLM II (40), LBC (30), %PG (0.5), and L/S (2.5).
(event) = 1/1 + e^{-Z}, where \( Z = 0.240 \times TDx-FLM_{II} – 0.422 \times GA \). Similarly, the GA-specific risk of RDS (Table III) values for LBC were calculated from multiple logistic regression analysis: Probability (event) = 1/1 + e^{-Z}, where \( Z = 0.125 \times LBC \times 0.066 \times GA – 4.386 \).

Both of these probabilities correlated well (\( R = 0.82, P < .001 \)), and the mean difference was only 0.0034 (\( t = 0.325, P = .746 \)). The TDx-FLM II mean probabilities for RDS (13) and no RDS (202) were significantly different (0.486 vs 0.897), with \( F(1,213) = 36.37, P < .001 \). Similarly, the LBC mean probabilities for RDS (13) and no RDS (202) were also significantly different (0.549 vs 0.885), with \( F(1,213) = 32.82, P < .001 \).

Comments

In obstetric practice there is always a finite possibility of RDS. Except in a few cases (eg, severe preeclampsia or clinical chorioamnionitis, where the mother and fetal well-being can be jeopardized), obstetricians prefer to have evidence of neonatal FLM before iatrogenic termination of pregnancy. Sometimes a simplistic answer, ie, “mature” or “immature” based on an L/S ratio can be misleading as documented by Pinette et al.\(^\text{18}\)

We found that in 5 pregnancies (cases A, D, G, J, and M) the L/S ratio was indicative of FLM (L/S \( < 2.5 \)), when in actuality all 5 infants had RDS. The GA-specific predicted risks of RDS (Table IV) for the same 5 cases on the basis of TDx-FLM II and GA using the equation in the results section were as follows: A = 5%, D = 33%, G = 11%, J = 82%, and M = 1%. Similar GA-specific predicted risks of RDS (Table IV) based on LBC were as follows: A = 31%, D = 32%, G = 15%, J = 57%, and M = 6%. It is clear from our data as well as the recommendations of others\(^\text{15,17,18}\) that if the physicians at our institutions had access to GA-specific percent predicted risk of RDS data, at least 4 of 5 cases of RDS could have been predicted on the basis of LBC results, and 3 of 5 on the basis of TDx-FLM II results.

Our data suggested identical statistical and clinical performance for both LBC and TDx-FLM II. We recommend that the GA-specific predicted risk of RDS tables should replace the misleading cutoff values of mature, indeterminate, and immature for these 2 FLM indices (LBC and TDx-FLM II). We also recommend

### Table III

<table>
<thead>
<tr>
<th>GA (wk)</th>
<th>TDx-FLM II (mg/g)</th>
<th>LBC (Sysmex XE-2100) (#/µL × 10^{-3})</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;16</td>
<td>34</td>
<td>95 80 56 29 11 4 1 &lt;1 &lt;1 &lt;1 &lt;1 1</td>
</tr>
<tr>
<td>16-20</td>
<td>35</td>
<td>97 81 56 38 16 5 2 &lt;1 &lt;1 &lt;1 &lt;1 &lt;1</td>
</tr>
<tr>
<td>21-25</td>
<td>36</td>
<td>98 91 75 49 22 8 3 &lt;1 &lt;1 &lt;1 &lt;1 &lt;1</td>
</tr>
<tr>
<td>26-30</td>
<td>37</td>
<td>96 94 82 58 30 12 4 1 &lt;1 &lt;1 &lt;1 &lt;1 &lt;1</td>
</tr>
<tr>
<td>31-35</td>
<td>38</td>
<td>99 96 87 66 39 17 6 2 &lt;1 &lt;1 &lt;1 &lt;1 &lt;1</td>
</tr>
<tr>
<td>36-40</td>
<td>39</td>
<td>99 97 91 76 49 23 9 3 &lt;1 &lt;1 &lt;1 &lt;1 &lt;1</td>
</tr>
<tr>
<td>41-45</td>
<td>40</td>
<td>100 98 94 83 60 31 12 4 1 &lt;1 &lt;1 &lt;1 &lt;1 &lt;1</td>
</tr>
<tr>
<td>46-50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51-55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56-60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>61-65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>66-70</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table IV

<table>
<thead>
<tr>
<th>Case no.</th>
<th>GA</th>
<th>DM</th>
<th>L/S ratio</th>
<th>PG</th>
<th>TDx-FLM II*</th>
<th>LBC*</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>35</td>
<td>No</td>
<td>2.8</td>
<td>Negative</td>
<td>38 (5)</td>
<td>23 (31)</td>
</tr>
<tr>
<td>B</td>
<td>35</td>
<td>No</td>
<td>2.1</td>
<td>Negative</td>
<td>23 (66)</td>
<td>16 (51)</td>
</tr>
<tr>
<td>C</td>
<td>35</td>
<td>No</td>
<td>1.1</td>
<td>Negative</td>
<td>7 (99)</td>
<td>3 (84)</td>
</tr>
<tr>
<td>D</td>
<td>34</td>
<td>Yes</td>
<td>2.6</td>
<td>Negative</td>
<td>27 (33)</td>
<td>23 (32)</td>
</tr>
<tr>
<td>E</td>
<td>36</td>
<td>Yes</td>
<td>1.7</td>
<td>Negative</td>
<td>27 (53)</td>
<td>20 (48)</td>
</tr>
<tr>
<td>F</td>
<td>36</td>
<td>No</td>
<td>2.0</td>
<td>Negative</td>
<td>25 (60)</td>
<td>19 (41)</td>
</tr>
<tr>
<td>G</td>
<td>37</td>
<td>Yes</td>
<td>2.7</td>
<td>Negative</td>
<td>38 (11)</td>
<td>29 (15)</td>
</tr>
<tr>
<td>H</td>
<td>36</td>
<td>No</td>
<td>2.2</td>
<td>Negative</td>
<td>28 (50)</td>
<td>34 (9)</td>
</tr>
<tr>
<td>I</td>
<td>36</td>
<td>No</td>
<td>1.3</td>
<td>Negative</td>
<td>12 (98)</td>
<td>5 (80)</td>
</tr>
<tr>
<td>J</td>
<td>37</td>
<td>No</td>
<td>2.7</td>
<td>Negative</td>
<td>23 (82)</td>
<td>13 (57)</td>
</tr>
<tr>
<td>K</td>
<td>35</td>
<td>No</td>
<td>1.0</td>
<td>Negative</td>
<td>23 (66)</td>
<td>9 (72)</td>
</tr>
<tr>
<td>L</td>
<td>37</td>
<td>No</td>
<td>2.4</td>
<td>Negative</td>
<td>30 (45)</td>
<td>9 (69)</td>
</tr>
<tr>
<td>M</td>
<td>37</td>
<td>No</td>
<td>3.7</td>
<td>1.4%</td>
<td>51 (1)</td>
<td>37 (6)</td>
</tr>
</tbody>
</table>

DM, Diabetes mellitus.

* Percent predicted (in parenthesis) risk of RDS on the basis of TDx-FLM II or LBC and GA using equation described in the Results section.
that in view of the poor concordance among various instruments for the assay of LBC, a cooperative effort among the laboratories using exactly the same hematology analyzer should provide a better GA-specific percent predicted risk table of RDS.

References

The mean weekly increment of amniotic fluid TDx-FLM II ratio is constant during the latter part of pregnancy

Ibrahim Bildirici, MD,a Christopher N. Moga, MD,a Ann M. Gronowski, PhD,b Yoel Sadovsky, MDa,c

Departments of Obstetrics and Gynecology,a Pathology and Immunology,b and Cell Biology and Physiology,c Washington University School of Medicine, St Louis, MO

Received for publication November 16, 2004; revised March 11, 2005; accepted March 30, 2005

KEY WORDS
Fetal lung maturity
TDx-FLM II
Amniocentesis

Objective: The purpose of this study was to determine the mean weekly increment in amniotic fluid TDx-FLM II ratio during the latter part of pregnancy.

Study design: All women who underwent >1 amniocentesis for the determination of fetal lung maturity between 1998 and 2004 were identified retrospectively. Clinical information and TDx-FLM II ratios were collected from the participant’s chart and analyzed.

Results: The gestational age of all participants at the first TDx-FLM II test was 31.2 to 37.5 weeks (mean, 34.7 ± 1.4 weeks of gestation). The median interval between the 2 tests was 7 days (range, 5-36 days). We found that the mean weekly increment of TDx-FLM II was 14.4 ± 9.9 mg/g (surfactant to albumin) and remained constant across the gestational ages.

Conclusion: The mean weekly increment of TDx-FLM II is 14.4 ± 9.9 mg/g and is constant during the latter part of pregnancy. This information, combined with the gestational age, should be useful in treating women with an initial immature test.

Hyaline membrane disease (HMD) is a major cause of morbidity and death in the newborn infant. In 1998 it was ranked among the 4 leading causes of infant deaths in the United States. The use of biochemical indices of fetal lung maturity has been instrumental in the assessment of newborn risk of HMD. TDx-FLM is an automated fluorescence polarization assay that determines the ratio of surfactant (expressed in milligrams) to albumin (expressed in grams) in amniotic fluid samples. TDx-FLM requires less amniotic fluid, exhibits reduced operator variance compared with traditional tests, and costs less. The reliability of TDx-FLM results has been validated in diverse obstetric populations, including women with diabetes mellitus. Importantly, when combined with gestational age, TDx-FLM serves to predict the probability of HMD and has become the test of choice in many institutions. Surprisingly, little is known about the increment in TDx-FLM ratio over time. Such information may enhance the ability of the clinician to extrapolate the risk of HMD after an initial estimate of lung immaturity, primarily when the result is combined with gestational age. This risk estimate may help to determine the timing of repeat amniocentesis for documentation of fetal lung maturity before a planned preterm delivery. We sought to determine the weekly increment in TDx-FLM ratio during the third trimester, when testing is
clinically meaningful. Our null hypothesis was that the mean weekly increment in TDx-FLM is constant during the latter part of pregnancy. We tested our hypothesis using the TDx-FLM II assay, a second-generation polarization assay that measures surfactant-to-albumin ratio as described for TDx-FLM. Notably, the results of TDx-FLM and TDx-FLM II are numerically equivalent. We evaluated the change in TDx-FLM II in women who had undergone >1 TDx-FLM II test between 31 and 38 weeks of gestation.

Methods

The Human Studies Committee at Washington University School of Medicine approved the study. All pregnant women who underwent >1 fetal lung maturity assessment with TDx-FLM II assay at Washington University in St. Louis, Mo, at Barnes Jewish Hospital between 1998 and 2004 were identified. All participants carried a singleton pregnancy. The gestational age was determined by a combination of the last menstrual period and ultrasound examination before 20 weeks of gestation or by ultrasound examination, only if the last menstrual period was not known or if there was >7 days discrepancy between the ultrasound examination and the last menstrual period. The gestational age was determined by ultrasound examination up to 28 weeks of gestation only when it was concordant with gestational age assignment by last menstrual period by <7 days. Only women who had a repeat TDx-FLM II assessment at least 5 days after the initial one was included. Women were excluded for non-reliable gestational age or if they did not meet the gestational age criteria described earlier. Women were also excluded for ruptured membranes, chorioamnionitis, a fetus with chromosomal or structural anomalies, or abnormalities of the amniotic fluid, including blood or meconium contamination.

TDx-FLM II analysis was performed according to the manufacturer’s instructions (Abbott Laboratories, North Chicago, IL). The upper limit of detection was 160 mg/g, and the lower limit of detection was 10 mg/g. In our laboratory, this assay has the following total imprecision: at a mean of 25 mg/g, covariance, 4.2%; at a mean of 49 mg/g, covariance, 3.9%; at a mean of 97 mg/g, covariance, 4.3%; and using patient-derived amniotic fluid pool, we obtained a mean of 20 mg/g, covariance, 11%.

Indications for fetal lung maturity assessment included previous uterine surgery (classic cesarean delivery, >2 low transverse cesarean deliveries or multiple myomectomies) that required a cesarean delivery (n = 29 pregnancies), maternal medical comorbidities or a history of late pregnancy fetal death in utero (n = 20 pregnancies), preterm labor with advanced cervical dilation (n = 12 pregnancies), placenta previa (n = 10 pregnancies), intrauterine growth restriction (n = 5 pregnancies), unexplained vaginal bleeding (n = 2 pregnancies), a large for gestational age fetus (n = 2 pregnancies), worsening mild preeclampsia (n = 2 pregnancies), malpresentation (n = 2 pregnancies), and a previous abdominal cerclage (n = 1 pregnancy). For each participant, we calculated the increment in TDx-FLM II ratio per day and multiplied this value by 7 to derive the increment in TDx-FLM II during an interval of 1 week. All statistical analyses were performed with SPSS software (version 11.0; SPSS Inc, Chicago, IL). We performed a linear regression analysis to test whether gestational age at the time of the first amniocentesis or the value of initial TDx-FLM II

### Table I Participant characteristic

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother</td>
<td></td>
</tr>
<tr>
<td>Age (y)*</td>
<td>28.8 ± 6.7 (17-46)</td>
</tr>
<tr>
<td>Nulliparous (n)</td>
<td>13 (15%)</td>
</tr>
<tr>
<td>Race (n)</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>46 (54%)</td>
</tr>
<tr>
<td>Black</td>
<td>39 (46%)</td>
</tr>
<tr>
<td>Diabetes mellitus (n)</td>
<td>19 (22%)</td>
</tr>
<tr>
<td>Pregestational (n)</td>
<td>11 (13%)</td>
</tr>
<tr>
<td>Gestational (n)</td>
<td>8 (9%)</td>
</tr>
<tr>
<td>Newborn infant</td>
<td></td>
</tr>
<tr>
<td>Gender: Female (n)</td>
<td>29 (34%)</td>
</tr>
<tr>
<td>Gestational age at delivery (mo)</td>
<td>36.9 ± 1.3 (33.6-40.4)</td>
</tr>
</tbody>
</table>

* Data are given as mean ± SD (range).

Figure 1 Gestational age distribution for all participants, assessed at the time of the initial amniocentesis.
ratio predicted the weekly increment in TDx-FLM II ratio. The Student $t$ test and analysis of variance were used for comparison of weekly changes in TDx-FLM II values among participant groups. A probability value of $<.05$ was considered significant.

**Results**

Our study included 85 participants who underwent 2 consecutive amniocentesis procedures for the determination of fetal lung maturity (Table I). Although 15 participants underwent a third amniocentesis, these additional data were not included in the analysis. The mean gestational age at the time of the first TDx-FLM II test was $34.7 \pm .4$ weeks (range, 31.2-37.5 weeks; Figure 1). The mean time interval between the 2 amniocentesis procedures was $10 \pm 5.1$ days (range, 5-36 days; median, 7 days). The TDx-FLM II ratios for all participants are presented in Figure 2. The mean value of weekly TDx-FLM II increments between the first and second amniocentesis for all participants was $14.4 \pm 9.9$ mg/g (95% CI, 12.3-16.5; median, 12.7 mg/g).

Using a linear regression (Figure 3), we found that the mean weekly increment in TDx-FLM II did not change significantly with advancing gestational age between 31 and 38 weeks ($P = .335$). Women at $\leq 34$ and $> 34$ weeks of gestation had a mean weekly increment of $13.0 \pm 10.0$ mg/g and $15.0 \pm 9.9$ mg/g, respectively ($P = .42$). Furthermore, as shown in Figure 4, the mean weekly increment in TDx-FLM II did not significantly change as a function of the initial TDx-FLM II ratio that ranged between 10 and 55 mg/g ($P = .284$). In addition, maternal race, diabetic status, or fetal gender did not significantly impact the mean weekly increment in TDx-FLM II (not shown).

Of all 85 participants, 46 women received steroids for the enhancement of fetal lung maturity, which included dexamethasone in 60% of the women and betamethasone in 40% of the women. The timing of steroid administration is shown in Table II. The mean weekly increment in TDx-FLM II among all 46 women who received steroids was similar to those who did not receive steroids (Table II). The timing of steroid administration relative to consecutive amniocentesis procedures and corresponding weekly increments in TDX-FLM II

![Figure 2](image_url) The TDx-FLM II ratios according to participants’ gestational ages.
is summarized in Table II. As expected, the mean weekly increment in TDx-FLM II in the group of women who received steroids immediately after the initial amniocentesis and before the second amniocentesis was increased. However, the number of participants in each of the subcategories presented in Table II was too small for meaningful conclusions. Twelve of the 85 newborn infants in our study were admitted to the special care nursery or the newborn intensive care unit. The range of their last TDx-FLM II was 12.9 to 41.4 mg/g.

**Comment**

The results support our null hypothesis that the mean weekly increment of 14.4 ± 9.9 mg/g in TDx-FLM II ratio is relatively constant between 31 and 38 weeks of gestation. In addition, the observed mean weekly increment in TDx-FLM II did not significantly change across a wide range of initial TDx-FLM II values. Similar results regarding the increment in lecithin/sphingomyelin ratio were reported recently in an abstract form by Deering et al.18 Notably, our data are based on 2 sequential amniocentesis procedures in a large group of women and not on several procedures in individuals. Therefore, the finding of a relatively constant mean weekly TDx-FLM II increment during the latter part of pregnancy does not rule out the possibility that in an individual the weekly TDx-FLM II increment may accelerate transiently. Nevertheless, our data effectively rule out the presence of a specific time point between 31

![Figure 3](image3.png) **Figure 3** The weekly TDx-FLM II increment as a function of gestational age. A scatter plot for the weekly increment in TDx-FLM II, according to the corresponding gestational ages at the time of initial amniocentesis. A linear regression line that represents the best fit is shown (P = .335; β coefficient, 0.106; R² = 0.011). The dotted line represents 95th percentile for all cases.

![Figure 4](image4.png) **Figure 4** The weekly TDx-FLM II increment for each participant as a function of initial TDx-FLM II value. A linear regression line that represents the best fit is shown (P = .284; β coefficient, 0.118; R² = 0.014).

<table>
<thead>
<tr>
<th>Steroids*</th>
<th>Gestational age (wk)</th>
<th>Weekly TDx-FLM II increment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not administered (n = 34; 40%)</td>
<td>35.6 ± 1.0</td>
<td>15.4 ± 11.4</td>
</tr>
<tr>
<td>Administered (n = 46; 54%)†</td>
<td>34.0 ± 1.4†</td>
<td>14.1 ± 9.2</td>
</tr>
<tr>
<td>Steroids before the initial amniocentesis</td>
<td>23</td>
<td>11.1 ± 4.4</td>
</tr>
<tr>
<td>Steroids after the initial amniocentesis</td>
<td>18</td>
<td>18.7 ± 12.4</td>
</tr>
<tr>
<td>Steroids before and after the initial amniocentesis</td>
<td>3</td>
<td>9.1 ± 6.8</td>
</tr>
</tbody>
</table>

* Information on steroid administration was not available for 5 women.
† The exact timing of steroid administration was uncertain for two women.
P < .001, when compared with the gestational age of women who did not receive steroids. All women who did not receive steroids were at a gestational age of >34 weeks. The numbers for the subgroups of participants who received steroids was too small for meaningful conclusions.
and 38 weeks of gestation when such acceleration might occur. Whereas 15 participants in our study underwent a third amniocentesis, the data from the additional procedure were not included in our analysis. Inclusion of these additional data would not have changed our estimate of the mean weekly increment (14.3 ± 9.6 mg/g). The exclusion of 8 cases with a weekly increment above 95th percentile confidence interval (Figure 1) would have yielded a more conservative estimate of the weekly increment (11.8 ± 5.6 mg/g). Last, we noted that, whereas most subjects underwent 2 consecutive tests in 1 to 2 weeks, in 4 women this interval was >20 days. Although the weekly increment for these women was not markedly different from those with an interval of 1 week, an assumption of linearity for such a long period might have introduced a small error into the weekly estimates. We determined the power of our study (α = .05) to find a relevant difference in mean weekly TDx-FLM-II ratio increment between participants at a gestational age above or below 34 weeks. Based on our data, our study had a power of 84% to detect a difference of 7 mg/g in the weekly TDx-FLM II increment between the 2 groups.

We also found that maternal age, parity, race, or fetal gender did not alter the weekly increment in TDx-FLM II. We noted that women who were carrying male fetuses were over-represented in our series (Table I). Although this could be a mere coincidence, it might also reflect the parents’ expectation of a more favorable outcome for preterm female infants and therefore less enthusiasm for amniocentesis for female fetuses. Nevertheless, in our series the mean weekly increment in TDx-FLM II was not statistically different for women who were carrying male or female fetuses. Interestingly, we did not observe a difference in TDx-FLM II increment between diabetic and nondiabetic women. Our data are consistent with the possibility that the weekly increment in fetal lung maturity testing in diabetic women takes place at a pace similar to that of nondiabetic women, yet the process may begin at a later gestational age.19

Our retrospective study was not designed to interrogate the influence of steroids on fetal lung maturation. Although we found that the weekly increment in TDx-FLM II was not significantly different in women who received steroids at any time point and those women who did not, the mean weekly increment in TDx-FLM II in women who received steroids immediately after the initial amniocentesis was increased. However, the small number of participants precluded a meaningful conclusion. Women who received steroids were at a significantly younger gestational age than those women who did not receive steroids. In addition, among all women who received steroids, there was an average of 3-week interval between the steroid administration and the initial amniocentesis. It is likely that steroid administration impacted fetal lung maturity in many of the study participants, but not necessarily during the weekly interval that was assessed in this study.

TDx-FLM is among the most commonly used clinical tests for the prediction of fetal lung maturity.7-13,20,21 Indeed, MacKenzie et al22 described the influence of 10-unit increment in TDx-FLM and the change in the risk of HMD. Conclusions from our study, combined with gestational age, may assist the practitioner in extrapolating the risk of HMD and in optimizing the timing of repeat fetal lung maturity testing.

Acknowledgment

We thank Lori Rideout for administrative assistance.

References


Maternal plasma concentrations of IGF-1, IGFBP-1, and C-peptide in early pregnancy and subsequent risk of gestational diabetes mellitus

Chunfang Qiu, MD, MS,a,∗ Surab Vadachkoria, MD, PhD,a Lois Meryman, BA,a Ihunnaya O. Frederick, MPH,a Michelle A. Williams, ScD,a,b

Objective: Insulin-like growth factor-1 (IGF-1) and IGF binding protein-1 (IGFBP-1) may be important determinants of glucose homeostasis. We examined the association between circulating concentrations of IGF-1, IGFBP-1 in early pregnancy and development of gestational diabetes mellitus (GDM).

Study design: Maternal plasma (collected at 13 weeks) IGF-1, IGFBP-1, and C-peptide were measured using immunoassay. Relative risks (RR) and 95% CIs were calculated.

Results: The percentage of the cohort that developed GDM was 5.8% (n = 804). Free IGF-1 and IGFBP-1 were inversely associated with GDM risk, while C-peptide was positively associated with GDM risk (P for trend test <.05). Women with free IGF-1 ≥1.08 ng/mL experienced a 69% reduced risk of GDM (CI 0.12-0.75) compared with women having concentrations <0.80 ng/mL. There was a 57% reduced risk of GDM among women with IGFBP-1 ≥68.64 ng/mL (RR = 0.43, CI 0.18-1.05). Women with C-peptide ≥3.00 ng/mL experienced a 2.28-fold increased risk of GDM (CI 1.00-5.19) compared with women who had concentrations <1.45 ng/mL.

Conclusion: These associations may help to further elucidate the pathologic process of GDM.

Emerging evidence indicates that insulin-like growth factor-1 (IGF-1), and insulin-like growth factor binding proteins (IGFBPs), are involved in a myriad of physiologic and pathologic processes related to reproductive endocrinology,1 bone metabolism,2 and the growth of various cancers.3 The structural homology between IGF-1 and insulin and the hypoglycemic activity regulated by IGFBPs also suggest that IGF-1 and its binding proteins have an intrinsic role in glucose metabolism and homeostasis.4 To date, 6 binding proteins regulating IGF-1 bioactivity have been characterized.5 A large proportion of IGF-1 in circulation is bound to IGFBP-3.5 The free fraction of IGF-1, which is the biologically active form and comprises less than 1% of the total

---

KEY WORDS
IGF-1
IGFBP-1
C-peptide
Gestational diabetes mellitus

Supported in part by an award from the National Institutes of Health (R01-HD/HL 32562).

∗ Reprint requests: Dr Chunfang Qiu, Center for Perinatal Studies, Swedish Medical Center, 747 Broadway (Suite 4 North), Seattle, WA, 98122.

E-mail: Chun-Fang.Qiu@Swedish.org

© 2005 Mosby, Inc. All rights reserved.


0002-9378/S - see front matter © 2005 Mosby, Inc. All rights reserved.
IGF-1 in circulation, is thought to be controlled by rapid alterations in IGFBP-1 concentrations. IGF-1 can increase rates of glucose disposal in skeletal muscle independent of insulin. Results from experiments with transgenic animals have shown that inactivation of the IGF-1 gene results in marked decreases in circulating concentrations of IGF-1, which is also associated with hyperinsulinemia and insulin insensitivity.

The production of IGF-1 is primarily regulated by growth hormones (GH). There is in vitro evidence that insulin enhances IGF-1 production by either direct regulation of the hepatic GH receptor or a permissive effect on postreceptor events. Insulin also has an important role in regulating IGF-1 bioactivity through inversely regulating hepatic IGFBP-1 production. Increased bioavailability of IGF-1 may lead to greater inhibition of GH release from the pituitary gland, and reduction in production of IGF-1 in the liver.

Results from studies of IGF-1 and related binding proteins in patients with gestational diabetes mellitus (GDM), a condition that is biochemically and epidemiologically similar to type 2 diabetes, have also been inconsistent, and most studies were case-control design. In a prospective design, noted that increased total IGF-1 concentrations (measured 4.5 years before diagnosis, on average) was associated with a 50% reduced risk of glucose intolerance and type 2 diabetes. The authors reported that total IGF-1 was inversely related to 2-hour glucose concentrations, and this relationship was only present in subjects with IGFBP-1 concentrations below the median (IGFBP-1 = 25 µg/L and P = .002). The authors did not measure free IGF-1 concentrations.

Results from studies of IGF-1 and related binding proteins in patients with gestational diabetes mellitus (GDM), a condition that is biochemically and epidemiologically similar to type 2 diabetes, have also been inconsistent, and most studies were case-control design. In a prospective design, noted that increased total IGF-1 concentrations (measured 4.5 years before diagnosis, on average) was associated with a 50% reduced risk of glucose intolerance and type 2 diabetes. The authors reported that total IGF-1 was inversely related to 2-hour glucose concentrations, and this relationship was only present in subjects with IGFBP-1 concentrations below the median (IGFBP-1 = 25 µg/L and P = .002). The authors did not measure free IGF-1 concentrations.

Because concentrations of free IGF-1 (ie, the readily dissociable fraction of IGF-1) are dependent on a complex interplay between the synthesis and release of total IGF-1, IGFBP-1, we also assessed the relation between maternal plasma free IGF-1 concentrations and GDM risk.

**Material and methods**

**Overall study design and population**

The population for the present analysis was drawn from participants of the ongoing Omega study, which is designed to examine maternal dietary and lifestyle risk factors of preeclampsia and GDM. This study has been described in detail elsewhere. Women who initiated prenatal care before 16 weeks' gestation were eligible for the study. Participants were invited to provide blood samples and to participate in an interview that took place during the 13th week of gestation, on average. Maternal and infant records were reviewed. The procedures used in this study were in agreement with the protocols approved by the Institutional Review Boards of Swedish Medical Center and Tacoma General Hospital. All participants provided written informed consent.

The study population for this report is from the first 1000 participants who were enrolled in the Omega study (year 1996-2000). During this period, 1219 eligible women were approached and 1000 (82%) agreed to participate. A total of 968 participants provided blood samples. Women found to have chronic hypertension (n = 45) and gestational diabetes mellitus (n = 4) were excluded. Women who experienced a spontaneous abortion or who had an induced abortion were excluded (n = 22). Also excluded were those women for whom the outcome of pregnancy was unknown because of any of the following reasons: moved, delivered elsewhere, and/or missing medical records (n = 46). Women with insufficient serum sample for laboratory analyses (n = 47) were excluded from these analyses. A cohort of 804 remained for analysis.

**Data collection and diagnostic procedures**

The diagnosis of GDM was made using the recently revised guidelines set forth by the American Diabetes Association (ADA). Women were diagnosed with GDM if 2 or more of the 4 diagnostic 100 g 3-hour oral glucose tolerance tests (OGTT) glucose concentrations exceeded the following criteria: 18 fasting > 95 mg/dL; 1-hour post-challenge > 180 mg/dL; 2-hour post-challenge > 155 mg/dL; 3-hour post-challenge > 140 mg/dL. From this cohort, we identified 47 confirmed GDM cases and 757 women who did not develop GDM.

From structured questionnaire and medical records, we obtained covariate information including maternal age, height, prepregnancy weight, reproductive and medical histories, and medical histories of first-degree family members. Prepregnancy body mass index (BMI) was calculated as weight in kg divided by height in meters squared. Maternal nonfasting blood samples, collected at 13 weeks’ gestation (interquartile range: 8–16 weeks), were frozen at −80°C until analysis. Plasma total IGF-1, free IGF-1, IGFBP-1, and C-peptide (a marker of endogenous pancreatic insulin secretion) concentrations were measured using enzyme immunoassay (Diagnostic Systems Laboratory, Inc, Webster, TX) with the intra- and interassay coefficients of variation all < 10%. All assays were performed without knowledge of pregnancy outcome.
Statistical analytic procedures

Women were grouped according to tertiles determined by the distribution of plasma IGFBP-1 concentrations among the entire cohort. We examined frequency distributions of maternal sociodemographic characteristics and medical and reproductive histories according to plasma IGFBP-1 tertiles. Spearman’s correlation coefficients were used to measure the closeness of a linear relationship between maternal plasma IGF-1, IGFBP-1, and C-peptide concentrations and selected characteristics. We estimated the relative association between varying tertiles of plasma IGF-1, IGF-BP1, and risk of GDM using Stata (version 7.0) software (Stata, College Station, TX). We fitted generalized linear models to derive risk ratios (RR) and 95% CIs.19 In multivariable analyses, we evaluated linear trends in risk by treating the tertiles as a continuous variable. To assess confounding, we entered variables into a generalized linear model 1 at a time, and then compared the adjusted and unadjusted risk ratios. Final generalized linear models included covariates that altered unadjusted risk ratios by at least 10%, as well as those covariates of a priori interest (eg, maternal age, nulliparity, prepregnancy BMI, and family history of type 2 diabetes mellitus). Statistical significance was not used as a criterion for selecting confounders. Maternal race, parity (as a continuous variable), and smoking status, as well as physically inactivity during pregnancy, were not confounders in this analysis.

Results

The characteristics of women in the study cohort according to tertiles of plasma IGFBP-1 concentrations are summarized in Table I.

Maternal plasma free IGF-1 and IGFBP-1 concentrations were inversely correlated with prepregnancy BMI ($r = -0.14$ and $r = -0.34$, respectively, $P < .01$). Both determinants were also inversely correlated with maternal plasma leptin ($r = -0.09$, $P = .01$ and $r = -0.17$, $P < .01$, respectively), as well as C-reactive protein (CRP) ($r = -0.23$, $P < .01$ and $r = -0.13$, $P < .01$, respectively). These concentrations were not correlated with maternal age, gestational age at blood collection, or infant birth weight. As marker of pancreatic insulin secretion, plasma C-peptide concentrations were significantly correlated with GDM risk factors, such as prepregnancy BMI ($r = 0.23$), triglyceride ($r = 0.29$), leptin ($r = 0.25$), and CRP concentrations ($r = 0.15$). Total IGF-1 concentrations were inversely correlated with CRP concentrations ($r = -0.15$, $P < .01$), but with no other factors (Table II).

### Table I

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Entire cohort n = 804</th>
<th>Tertiles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1 (Low) n = 268</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>Plasma IGFBP-1 (ng/mL)</td>
<td>!35.96</td>
<td>!35.96-68.63</td>
</tr>
<tr>
<td>Maternal Age (y)*</td>
<td>32.0 ± 0.3</td>
<td>32.4 ± 0.3</td>
</tr>
<tr>
<td>Maternal race/ethnicity</td>
<td>32.4 ± 0.3</td>
<td>31.7 ± 0.3</td>
</tr>
<tr>
<td>Non-Hispanic white</td>
<td>83.6</td>
<td>82.5</td>
</tr>
<tr>
<td>African American</td>
<td>2.5</td>
<td>3.0</td>
</tr>
<tr>
<td>Other</td>
<td>12.7</td>
<td>13.4</td>
</tr>
<tr>
<td>Unmarried</td>
<td>11.2</td>
<td>13.4</td>
</tr>
<tr>
<td>≤12 years’ education</td>
<td>4.0</td>
<td>4.9</td>
</tr>
<tr>
<td>Nulliparous</td>
<td>14.8</td>
<td>18.7</td>
</tr>
<tr>
<td>Family history of diabetes</td>
<td>6.3</td>
<td>9.3</td>
</tr>
<tr>
<td>Smoker</td>
<td>17.5</td>
<td>17.9</td>
</tr>
<tr>
<td>No exercise</td>
<td>19.9</td>
<td>13.1</td>
</tr>
<tr>
<td>Prepregnancy BMI (kg/m²)</td>
<td>20.0</td>
<td>54.2</td>
</tr>
<tr>
<td>20.0-24.9</td>
<td>15.6</td>
<td>21.3</td>
</tr>
<tr>
<td>25.0-29.9</td>
<td>8.7</td>
<td>18.3</td>
</tr>
<tr>
<td>≥30.0</td>
<td>12.9 ± 0.2*</td>
<td>12.6 ± 0.2</td>
</tr>
<tr>
<td>Gestational age at blood collection (wk)</td>
<td>12.9 ± 0.2*</td>
<td>12.6 ± 0.2</td>
</tr>
</tbody>
</table>

* Mean ± SE = mean ± standard error.
GDM risk was 3.25-fold higher among women with C-peptide concentrations ≥3.00 ng/mL, as compared with those women whose concentrations were <1.45 ng/mL (95% CI 1.47-7.18) (Table III). This association was attenuated after adjustment for potential confounding by maternal age, parity, family history of diabetes, and prepregnancy BMI (RR = 2.28, 95% CI 1.00-5.19).

Compared with women in the lowest tertile (total IGF-1 <154.18 ng/mL), women having intermediate total IGF-1 concentrations (154.18-215.04 ng/mL) experienced a 1.71-fold (95% CI 0.82-3.58) increased risk of developing GDM after accounting for the potential confounders. The corresponding RR for women with the highest tertile of total IGF-1 (>215.05 ng/mL) was 1.47 (95% CI 0.70-3.08). However, this association did not reach statistical significance.

As seen in Table III, risk of GDM decreased across successively higher tertiles of free IGF-1 concentrations (RRs: 1.00, 0.60, and 0.24; P value for trend = .01). Relative to women in the lowest tertile of maternal plasma free IGF-1 concentrations (<0.80 ng/mL), women with concentrations in the most upper tertile (≥10.08 ng/mL) experienced a 76% reduced risk of GDM (RR = 0.24; 95% CI 0.10-0.58). After adjustment for confounders, the RR was attenuated slightly (adjusted RR = 0.31; 95% CI 0.12-0.75). A similar pattern was seen for IGFBP-1. The risk of GDM decreased across successively higher tertiles of IGFBP-1 concentrations (RRs: 1.00, 0.72, and 0.43; P value for trend = .06). There was a 57% reduced risk of GDM among women with IGFBP-1 concentrations ≥68.64 ng/mL (RR = 0.43, 95% CI 0.18-1.05) compared to women with concentrations in the lowest tertile (<35.96 ng/mL) after adjusting for confounders.

### Comment

GDM, a heterogeneous complication of pregnancy, is characterized by carbohydrate intolerance of variable severity with onset or first recognition during pregnancy. Insulin resistance and β-cell dysfunction are thought to be major determinants of its development. Its pathophysiologic and epidemiologic characteristics in many ways resemble that of type 2 diabetes. An expanding body of evidence documents similarities between GDM and the metabolic syndrome. Our finding of an association between elevated C-peptide in early pregnancy and subsequent GDM risk suggests that maternal hyperinsulinemia and/or insulin resistance is detectable about 15 weeks before GDM is typically diagnosed. Our findings are consistent with results from other studies. Kautzky-Willer et al reported that C-peptide was 30% higher (P < .05) in subjects with GDM than in controls or in nonpregnant lean women. These investigators also noted that β-cell sensitivity to glucose for insulin release was decreased by 40% to 50% (P < .01) in subjects with GDM. Winkler et al reported that C-peptide was 6.0 ± 2.7 and 3.1 ± 1.7 mg/L in 30 GDM cases and 35 pregnant controls, respectively (P < .01).

A finding pertaining to total IGF-1 is in agreement with observations from some case control studies where plasma concentrations of this analyte were determined at delivery. Latham et al reported that there was no difference of total IGF-1 concentrations between women with GDM and controls. Our findings and those of Latham, however, are inconsistent with findings reported by Sandhu or Hughes et al, who noted that total IGF-1 concentrations (measured in samples collected in the third trimester) were higher in women with GDM than in controls.

Although we are not aware of other published reports regarding maternal free IGF-1 concentrations and GDM risk, our observation of an inverse association between plasma free IGF-1 and GDM risk is in agreement with other studies involving metabolic syndrome. For instance, Healde reported that elevated free IGF-I and IGFBP-1 concentrations were associated with decreased risk of metabolic syndrome (OR = 0.46, 95% CI 0.22-0.96; OR = 0.58, 95% CI 0.44-0.76, respectively).

Our finding of an inverse association between IGFBP-1 and subsequent GDM risk is not in agreement with a report from O’Leary et al. In a case-control study where maternal concentrations of IGFBP-1 were measured after the clinical diagnosis of GDM, the authors reported that IGFBP-1 levels was higher in GDM cases than controls. However, our results are consistent with a large body of literature documenting a
similar association with insulin resistance, impaired glucose tolerance, and risk of metabolic syndrome in men and nonpregnant women. For instance, Heald et al reported that IGFBP-1 concentrations were significantly lower in subjects with impaired glucose tolerance when compared with controls. The authors further noted a 40% reduction in risk of impaired glucose tolerance for every 2.7 ng/mL increase in IGFBP-1 concentration (OR = 0.6, 95%CI 0.49-0.71).

There are several possible explanations for why results from existing studies are not in agreement, the most likely being differences in study design and analytical techniques. Differences in the sensitivity and specificity of assays may have resulted in considerable variability in IGF-1 and IGFBP-1 concentrations. Differences in population characteristics such as age, overweight and obesity status, as well as different treatments and variability in insulin sensitivity across populations may account for conflicting reports. In addition, distortion from uncontrolled confounding secondary to gestational age at blood collection and other maternal factors may have been present in many of the previous studies.

Our study has several important strengths. Determination of IGF-1 and IGFBP-1 concentrations using plasma collected in early pregnancy served to clarify the temporal relationship between the biomarkers and subsequent risk of GDM. Additionally, the high follow-up rate achieved in our study (>95%) also minimized possible bias. However, our study has several limitations that merit consideration. A single measurement of IGF-1 and IGFBP-1 is not likely to provide a time-integrated measure of maternal status during the index pregnancy. Longitudinal studies with serial measurements are needed to further elucidate the mechanisms and pathophysiologic consequences of reduced free IGF-1 and IGFBP-1 concentrations during pregnancy. Studies are also needed to help determine the source of intra- and interperson variation in IGF-1 (and their binding proteins) synthesis, release, and circulation in maternal and fetal compartments. Moreover, although we excluded women with a medical diagnosis of pregestational diabetes, we cannot, with absolute certainty, exclude the possibility that some subjects in our study had undiagnosed diabetes or other abnormalities of glucose tolerance before or in early pregnancy. Universal glucose tolerance testing in early pregnancy is not part of the standard obstetric care. Several observations, however, serve to attenuate concerns about undiagnosed diabetes in this study population.

<table>
<thead>
<tr>
<th>Table III</th>
<th>Relative risk (RR) and 95%CI of GDM according to tertile of percentage of IGF-1 and its binding proteins, Seattle and Tacoma, Washington, 1996-2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurement</td>
<td>Tertile of measurement</td>
</tr>
<tr>
<td>C-peptide (ng/mL)</td>
<td>Interval</td>
</tr>
<tr>
<td>Cases, n</td>
<td>8</td>
</tr>
<tr>
<td>Cohort, N</td>
<td>265</td>
</tr>
<tr>
<td>RR (95%CI)</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td>RR* (95%CI)</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td>Total IGF-1 (ng/mL)</td>
<td>Interval</td>
</tr>
<tr>
<td>Cases, n</td>
<td>12</td>
</tr>
<tr>
<td>Cohort, N</td>
<td>267</td>
</tr>
<tr>
<td>RR (95%CI)</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td>RR* (95%CI)</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td>Free IGF-1 (ng/mL)</td>
<td>Interval</td>
</tr>
<tr>
<td>Cases, n</td>
<td>25</td>
</tr>
<tr>
<td>Cohort, N</td>
<td>265</td>
</tr>
<tr>
<td>RR (95%CI)</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td>RR* (95%CI)</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td>IGFBP-1 (ng/mL)</td>
<td>Interval</td>
</tr>
<tr>
<td>Cases, n</td>
<td>26</td>
</tr>
<tr>
<td>Cohort, N</td>
<td>268</td>
</tr>
<tr>
<td>RR (95%CI)</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td>RR* (95%CI)</td>
<td>1.00 (referent)</td>
</tr>
</tbody>
</table>

* Adjusted for maternal age, nulliparity, family history of diabetes, and prepregnancy BMI.
consistent with observations in other settings. Hence, the likelihood that our findings are due entirely to the prevalence of undiagnosed pregestational diabetes in this cohort is very low. Third, some RRs are statistically unstable due to our small sample size. Finally, the current study population was primarily focused on the primiparous women, and the cohort was mainly well-educated white women. Therefore, caution must be taken when generalizing our results.

The physiologic processes underlying the role of IGF-1 in glucose regulation are unclear. However, it is known that infusion of IGF-1 will suppress glucose counter-regulatory hormones, such as glucagon and GH, and may speed up oxidation of lipids. IGFBP-1 concentrations may be secondary to maternal secretion or hepatic insulin sensitivity, both of which are important components in glucose regulation. Further studies are needed to evaluate these hypotheses and to comprehensively evaluate the precise roles that IGFs and IGFBPs play in insulin resistance and the development of GDM. Future studies may also elucidate the extent to which reduced IGFBP-1 concentrations may be secondary to maternal hyperinsulinemia and insulin resistance.

In summary, we have shown that alteration in maternal plasma free-IGF-1 and IGFBP-1 concentrations are evident in pregnancies destined to develop GDM as early as 13 weeks’ gestation. The findings from this prospective study are generally consistent with the reports of previous investigations. Our findings are also consistent with a much larger body of evidence from experimental, clinical, and epidemiologic investigations, which suggest that the IGF system is an important mediator of glucose homeostasis in humans. Taken together with the available literature, our results suggest that dysregulation of the synthesis and release of IGF-1 and IGFBP-1 may be of etiologic importance in GDM.

Acknowledgments

The authors are indebted to staff at the Center for Perinatal Studies, Swedish Medical Center.

References


Protein Z in patients with pregnancy complications

Florence Bretelle, MD, PhD,a,* Dominique Arnoux, PhD,b Raha Shojai, MD,c Claude D’Ercole, MD,c José Sampol, PhD,d,e Françoise Dignat, PhD,d,e Laurence Camoin-Jau, PhDd,e

Service de Gynécologie-Obstétrique, Pr. B. Blanc, Centre Hospitalo-Universitaire de la Conception,a Laboratoire d’Hématologie, Pr G. Sebahounb; and Service de Gynécologie-Obstétrique, Pr. L Boubli, Centre Hospitalo-Universitaire Nordc; Laboratoire d’Hématologie, Pr. J. Sampol, Centre Hospitalo-Universitaire de la Conceptiond; UMR Inserm 608, Laboratoire d’Hématologie et d’Immunologie, UFR de Pharmacie,e Marseille, France

Received for publication July 24, 2004; accepted April 1, 2005

KEY WORDS
Protein Z
Preeclampsia
Intrauterine growth restriction
Intrauterine fetal demise

Objective: This study was undertaken to evaluate the association between protein Z concentration and pregnancy complications.

Study design: A prospective case-control study was conducted over a 2-year period to evaluate the prevalence of protein Z deficiency in pregnancy complications. Protein Z levels were measured at the time of diagnosis of complications such as preeclampsia, intrauterine growth restriction, and intrauterine fetal demise. Protein Z deficiency was defined as a plasma level below 1.2 mg/L. In addition to patients presenting with pregnancy complications, healthy age-matched nonpregnant and pregnant women were invited to participate.

Results: A total of 145 women were included in the study: 50 nonpregnant women, 34 healthy pregnant women, 29 women with preeclampsia, 25 women presented with intrauterine growth restriction, and 7 women with intrauterine fetal demise. The median protein Z level was similar in healthy pregnant and nonpregnant women (1.63 [0.47-3.1] mg/L and 1.69 [0.7-3] mg/L, respectively). Three women with normal pregnancies had a low protein Z level (8.8%), compared with 8 patients presenting with intrauterine growth restriction (33.3%) and 8 patients with intrauterine fetal demise (50%). Compared with normal pregnancy, the frequency of decreased protein Z was significantly higher in cases of intrauterine growth restriction and in intrauterine fetal demise (relative risk [RR] 1.96, 95% CI 1.16-3.32; P = .041 and RR 3.36, 95% CI 1.65-6.8; P = .0031, respectively), but not in preeclampsia (RR 1.6, 95% CI 0.9-2.8; P = .23). Placenta histologic examination revealed vascular lesions in 50% of patients with protein Z deficiency and in 33% of patients with normal levels of protein Z (RR 0.84; 95% CI 0.6-1.2).

Conclusion: Protein Z deficiency is associated with late fetal demise and intrauterine growth restriction. The pathophysiologic role of protein Z deficiency, either congenital or caused by the presence of specific antibodies remains unclear and should be further investigated.

© 2005 Mosby, Inc. All rights reserved.
Protein Z is a vitamin K-dependent plasma glycoprotein synthesized by the liver. It plays a role in the physiologic inhibition of coagulation by acting as a cofactor in the inactivation of factor Xa. Conflicting findings on plasma protein Z levels in patients with ischemic stroke have been reported and the thrombotic risk associated with this deficiency is still debated. Protein Z deficiency has also been described in fetal loss between the 8th and 15th week of gestation. Moreover, protein Z deficiency has recently been associated with poor pregnancy outcome. The pathophysiology of early fetal loss shows similarities with that of late pregnancy complications, which include preeclampsia, intrauterine restriction (IUGR), and fetal demise (IUFD). An abnormality of trophoblast invasion associated with thrombotic risk factors, such as thrombophilia or infection may be involved. A recent hypothesis emphasizes the role of inflammation and atherosclerosis in preeclampsia. Indeed, atherosclerotic lesions and thrombosis have been found on the arterial wall of uteroplacental vessels. Because protein Z is implicated in atherosclerosis, we conducted a study to investigate the prevalence of its deficiency in patients with pregnancy complications.

**Material and methods**

After informed consent, patients presenting with preeclampsia, IUGR, or IUFD and healthy pregnant and nonpregnant women were included in a study approved by the local ethics committee of our institution (Conception Hospital, Marseille, France). The study was conducted over a 2-year period (2002-2003). Preeclampsia was defined as a diastolic arterial blood pressure greater than 90 mm Hg, and a systolic blood pressure greater than 140 mm Hg, associated with proteinuria (more than 300 mg/24 hours). The blood pressures were taken at admission, the inclusion day after at least half an hour bed rest.

Isolated IUGR was defined as ultrasonographic measurement less than the fifth percentile for gestational age. IUFD was defined by ultrasound examination as a visible fetus without cardiac activity after 15 weeks of gestation. The exclusion criteria for women whose pregnancies were complicated by preeclampsia, IUGR, and IUFD were the presence of congenital malformations or chromosomal abnormalities in the fetus, recent cytomegalovirus infection, trauma, drug or alcohol abuse during pregnancy, and clinical signs of maternal preeclampsia as previously defined. In addition to patients presenting with pregnancy complications, age-matched women seen for routine gynecologic examination (non-pregnant women) and healthy pregnant primigravid women with no history of medical illness attending the routine antenatal clinics, were invited to participate as controls. The blood samples were obtained at the enrollment for the healthy pregnant and nonpregnant women.

Blood samples were collected into 0.129 mol/L sodium citrate (3.8%) and centrifuged, and plasma was stored at –80°C. All pathologic cases were included in the study at the time of diagnosis.

No patient had vitamin K deficiency, or vitamin K antagonist treatment. Protein Z was measured with the use of a commercial enzyme-linked immunosorbent assay (ELISA) kit (kindly provided by Diagnostica Stago, Asnières, France). The interassay of 15 assays was 8% and the within assay variation was 3%. The specificity and sensitivity was 92% and 34%, respectively. Each measurement of protein Z was performed in duplicate. All patients were screened for thrombophilia (antithrombin, protein C, Leiden mutation of factor V, and 20210 mutation of prothrombin gene) and for the presence of antiphospholipid antibodies (anticardiolipin, anti-β2GP1, lupus anti-coagulant). Protein S levels were not taken into consideration because all the patients were tested during pregnancy.

The level of 1.2 mg/L was selected as a cutoff to define protein Z deficiency, because it corresponded to the 10th percentile among the normal population (including pregnant and nonpregnant healthy women).

Placentas of patients with complications were fixed within the 24th hour in 10% neutral buffered formalin solution and embedded in paraffin blocks. Sections were obtained from placental villi and stained with hematoxylin and eosin. Histologic examination of placentas including general aspect, presence of infarction or intervillous thrombosis was performed.

Thrombi were diagnosed when aggregates of fibrin were seen adherent to the luminal aspect of fetal stem vessels, usually on endothelial cushions, occluding less than 20% of the lumen.

**Statistical analysis**

Data were expressed as the median and range, and statistical analysis was performed with GraphPad statistical software. \( P < .05 \) was considered to represent significance. The correlation between parameters and markers of preeclampsia, IUGR, or IUFD was evaluated with a Spearman test. The statistical analysis was performed by use of the 1-way analysis of variance method (ANOVA) or Mann-Whitney for quantitative variables as indicated and \( \chi^2 \) was used for categorical variables.

**Results**

Over this 2-year period, there were 61 cases of pregnancy complications that met inclusion criteria, including 29
preeclampsia and 25 isolated IUGR. In the group of IUFD (n = 16), 7 patients presented with isolated IUFD without IUGR or preeclampsia; the other patients had either preeclampsia (n = 2) or IUGR (n = 7). In addition, 34 normal pregnancies and 50 healthy, nonpregnant, age-matched women were enrolled. Protein Z levels and patient characteristics are described in Table I. The overall median level of protein Z was 1.63 with a range of 0.47 to 3.10 mg/L. Levels of protein Z were not different among the different pregnancy periods between 19 to 29, 30 to 36, and 37 to 42 weeks of gestation (the median among the different pregnancy periods were 1.73 ± 0.6, 1.75 ± 0.7, and 1.62 ± 0.6, respectively; \( P = .85 \)). The percentage of protein Z-deficient women was elevated in pregnancies complications (n = 21) compared with normal pregnancies (n = 3) (34.5% vs 8.8%; RR 1.55, 95% CI 1.2-2; \( P = .0065 \) (Table II). Patients with a protein Z level less than 1.2 mg/L represented 20.6% of the preeclampsia group (n = 6) (RR 1.6, 95% CI 0.9-2.8; \( P = .23 \)), 33.3% of the group with isolated IUGR (n = 8) (RR 1.96, 95% CI 1.16-3.32; \( P = .041 \)) and 50% of the group with IUFD (n = 8) (RR 3.36, 95% CI 1.65-6.8; \( P = .0031 \)). After exclusion of the patients presenting with IUGR or with preeclampsia, among the 7 remaining: 4 patients had a protein Z level less than 1.2 mg/L (57%) (RR 2.1, 95% CI 0.9-5; \( P = .01 \)). Protein Z deficiency was significantly more frequent in isolated IUGR and in IUFD (Figure). In our population, gestational age did not influence protein Z levels (\( r = 0.056, P = .64 \)).

The results of biologic screening for thrombophilia are reported in Table II. Among patients with protein Z levels below 1.2 mg/L, 2 patients were carriers of a heterozygous Leiden mutation of factor V and 1 also had anticardiolipin antibodies. Another patient with protein Z deficiency had lupus anticoagulant. These 3 patients had severe pregnancy complications: all presented with UFID, which was associated with a severe preeclampsia (blood pressure greater than 160/100 mm Hg associated with proteinuria greater than 4 g/24 hours) in 1 patient and with IUGR in the 2 others. In the group of patients with IUFD, 2 patients had a history of unexplained thrombosis. None of the patients with low level of protein Z presented with vitamin K deficiency and all had normal levels of other vitamin K-dependent factors.

The histologic examinations of the placentas revealed thrombi or ischemic lesions in 50% of patients with

### Table I Characteristics of healthy pregnant women and patients with preeclampsia, isolated IUGR, and IUFD

<table>
<thead>
<tr>
<th></th>
<th>PE (n = 29)</th>
<th>Isolated IUGR (n = 25)</th>
<th>IUFD (n = 16)</th>
<th>Normal pregnancies (n = 34)</th>
<th>Nonpregnant women (n = 50)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fetal weight at delivery (g)</td>
<td>1490 [400-4325]</td>
<td>1225 [500-2590]</td>
<td>1300 [400-3100]</td>
<td>3300 [2880-3900]</td>
<td>—</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>Systolic blood pressure (mm Hg)</td>
<td>160 [130-200]</td>
<td>120 [100-135]</td>
<td>140 [110-200]</td>
<td>110 [90-130]</td>
<td>—</td>
<td>&lt;.0001†</td>
</tr>
<tr>
<td>Diastolic blood pressure (mm Hg)</td>
<td>90 [80-110]</td>
<td>70 [60-90]</td>
<td>80 [60-110]</td>
<td>80 [50-90]</td>
<td>—</td>
<td>&lt;.0001†</td>
</tr>
<tr>
<td>Proteinuria (g/24h)</td>
<td>2.7 [0.4-6.6]</td>
<td>0.3 [0-3]</td>
<td>2.5 [0-6.6]</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Protein Z (mg/L)</td>
<td>1.75 [0.48-3.1]</td>
<td>1.79 [0.75-2.87]</td>
<td>1.57 [0.78-3.1]</td>
<td>1.63 [0.47-3.01]</td>
<td>1.69 [0.7-3]</td>
<td>ns</td>
</tr>
<tr>
<td>25th</td>
<td>1.32</td>
<td>1.24</td>
<td>0.88</td>
<td>1.29</td>
<td>1.41</td>
<td>ns</td>
</tr>
<tr>
<td>75th</td>
<td>2.12</td>
<td>2.21</td>
<td>2.65</td>
<td>2.26</td>
<td>2.3</td>
<td></td>
</tr>
</tbody>
</table>

Blood samples were collected at the inclusion. Gestational ages are given at sampling. Values are represented as median and [range], and upper and lower quartiles for protein Z. Difference between groups was analyzed by the Kruskall-Wallis and post hoc test.

* \( P < .05 \); all groups are different from normal pregnancy.
† \( P < .05 \); PE and IUFD groups are different from normal pregnancies and IUGR.
‡ \( P < .05 \); all groups are different from PE.
§ \( P < .05 \); PE group is different from normal pregnancies.
∥ \( P < .05 \); IUFD is different from PE and IUFD.

### Table II Characteristics of studied population for thrombophilias and antibodies screening

<table>
<thead>
<tr>
<th></th>
<th>Complicated pregnancies (n = 61) %</th>
<th>Normal pregnancies (n = 34) %</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>3.6</td>
<td>0</td>
<td>ns</td>
</tr>
<tr>
<td>PC</td>
<td>5.4</td>
<td>0</td>
<td>ns</td>
</tr>
<tr>
<td>FVL</td>
<td>3.4</td>
<td>2.9</td>
<td>ns</td>
</tr>
<tr>
<td>20210a</td>
<td>0</td>
<td>0</td>
<td>ns</td>
</tr>
<tr>
<td>aPL</td>
<td>9</td>
<td>0</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Protein Z deficiency</td>
<td>34.5</td>
<td>8.8</td>
<td>&lt;.01</td>
</tr>
</tbody>
</table>

AT, Antithrombin deficiency; PC, protein C deficiency; FVL, Leiden mutation of factor V; 20210a, mutation of 20210 allele of prothrombin gene; aPL, antiphospholipid antibodies and lupus anticoagulant; Protein Z deficiency, protein Z levels <1.2 mg/L.
Protein Z deficiency is a recently recognized factor in thrombotic events. It has been implicated in early fetal loss and in complicated pregnancies. Protein Z deficiency has also been reported to be common in the antiphospholipid syndrome, which is known to be associated with recurrent fetal losses. In the current study, we found a high prevalence of protein Z deficiency in patients with IUFD or in IUGR. In the IUFD group, half of the patients had a protein Z level below 1.2 mg/L. This rate is greater than the 22% that has been reported in patients with 1 early fetal loss between 10 and 19 weeks but is comparable with that reported in pathologic conditions such as ischemic stroke. We studied late IUFD (mean gestational age: 28 weeks), but the high prevalence of protein Z deficiency compared with the previous studies cannot be explained by gestational age, because our results seem to show that protein Z levels are stable throughout pregnancy. IUFD subgroup analysis, after exclusion of patients presenting with preeclampsia and IUGR, was not significant as well. This may be due to confounding factors or to the small effective of this subgroup. Instead, protein Z deficiency may be implicated in IUFD through the constitution of thrombotic lesions in the placental vascular bed. Nevertheless, placental examinations showed thrombi to be present in 50% of low protein Z levels cases and in 33% of the remaining cases. This result was not statistically significant and might also be due to the effective of the study.

Atherosclerosis could be a feature shared by complicated pregnancies of vascular origin and ischemic stroke, in which protein Z deficiency has also been implicated. Moreover, lesions similar to those found in atherosclerosis have been observed in examination of placentas from women with preeclampsia. Deposits of protein Z have been shown in vascular lesions of patients with atherosclerosis. It could be postulated that protein Z could similarly be detected in the placental vascular bed, suggesting that protein Z deficiency could be secondary to pregnancy complications rather than causal. Against this hypothesis, Gris et al found an association between pregnancy complications and low levels of protein Z measured at least 6 months after the last obstetric events. This makes the possibility of protein Z consumption in the placental vascular bed unlikely. Even if thrombophilia predisposes to preeclampsia or IUGR, other factors such as modifications of inflammatory response or previous infection may also play a role. Protein Z deficiency may be one element of a complex series of events leading to the clinical manifestations. In our study, the 3 patients in whom protein Z deficiency was associated to other thrombophilic abnormalities with particularly unfavorable outcomes. Protein Z deficiency may thus increase the risk of vascular events in association with other predisposing factors as described in the antiphospholipid syndrome.

It has been suggested that in severe preeclampsia and unexplained fetal death, antiprotein Z antibodies could cause protein Z inhibition. We did not test our population for these antibodies, but this also might be an explanation for the decreased protein Z in complicated pregnancies. Protein Z deficiency thus seems to play a role in the physiopathology of pregnancy complications such as IUGR and IUFD. These findings should be validated in a larger study.

References


A free radical scavenger, edaravone, inhibits lipid peroxidation and the production of nitric oxide in hypoxic-ischemic brain damage of neonatal rats

Jesmin I. Noor, MD,a Tomoaki Ikeda, MD,a Yuto Ueda, MD,b Tsuyomu Ikenoue, MDa,*

Departments of Obstetrics and Gynecologya and Psychiatry,b Faculty of Medicine, University of Miyazaki, Japan

Received for publication November 15, 2004; revised February 28, 2005; accepted March 30, 2005

Objective: The purpose of this study was to elucidate a role for edaravone, a free radical scavenger 3-methyl-1-phenyl-2-pyrazolin-5-one, in neonatal hypoxic-ischemic brain damage. We determined the level of thiobarbituric acid reactive substances as an index of lipid peroxidation and nitric oxide metabolites as nitric oxide production.

Study design: Seven-day-old Wistar rats were subjected to left common carotid artery ligation followed by 2 hours of 8% oxygen exposure. Then, the rats were administered edaravone (9 mg/kg) or saline solution intraperitoneally. Cerebrospinal fluid was withdrawn just before the rats were killed at 2, 5, 24, and 48 hours after hypoxia, and brains were removed. The thiobarbituric acid reactive substances and nitric oxide metabolites levels were measured in the brain tissue and cerebrospinal fluid, respectively.

Results: On the ligated side, edaravone significantly decreased thiobarbituric acid reactive substances levels at 5 and 24 hours after hypoxia, compared with saline group (P < .01). Edaravone significantly decreased the nitric oxide metabolites level in the cerebrospinal fluid only at 5 hours, compared with saline group (P < .01).

Conclusion: Edaravone potently and transiently inhibited lipid peroxidation and the production of nitric oxide in the neonatal rat brain after hypoxic-ischemic insult.

© 2005 Mosby, Inc. All rights reserved.

KEY WORDS
Edaravone
Lipid peroxidation
Nitric oxide metabolite
Hypoxic-ischemic brain damage

The free radical scavenger 3-methyl-1-phenyl-2-pyrazolin-5-one, edaravone, was endorsed by the Ministry of Health, Labor and Welfare of Japan in April 2001 for the treatment of acute cerebral infarction in adult patients. Previous studies have shown the neuroprotective and antioxidative mechanisms of edaravone in adult ischemic animal models.1-5 For example, edaravone dose-dependently prevented brain edema that was induced by hemispheric embolization by suppressing the production of lipoxygenase metabolites.1 Edaravone has been also reported to prevent the peroxidative vascular endothelial cell damage caused by hydroperoxyeicosatetraenoic acid in vitro study.2 Furthermore, edaravone was effective in the treatment of adult rat brain edema because of focal ischemia by inhibiting both non-enzymatic free radical reactions and lipoxygenase activity in vitro.3 Postischemic treatment with edaravone significantly

* Reprint requests: Tsuyomu Ikenoue, MD, Department of Obstetrics and Gynecology, Faculty of Medicine, University of Miyazaki, 5200 Kihara, Kiyotake, Miyazaki 889-1692, Japan.
E-mail: satomako@fc.miyazaki-med.ac.jp

doi:10.1016/j.ajog.2005.03.069
inhibited arachidonate-induced brain edema in the adult Mongolian gerbil. Edaravone may function as a neuroprotective agent that counteracts oxidative neurotoxicity that arises from activated microglia, because it occurs in either inflammatory or neurodegenerative disorders of the central nervous system. Taken together, these results suggest that edaravone has a neuroprotective and antioxidant effect in cerebral hypoxic-ischemic brain damage in adults.

We previously examined the effects of edaravone on hypoxic-ischemic brain damage in neonatal rats. Edaravone was administered intraperitoneally at a dosage of 3, 6, or 9 mg/kg after hypoxia-ischemia, then the brains were examined histologically in the infarct areas by 2, 3, 5-triphenyltetrazolium chloride staining. We reported that edaravone is neuroprotective in neonatal hypoxic-ischemic rats in a dose-dependent manner. Neonatal asphyxia involves various systemic organs, unlike adult cerebral infarction, such as the heart, kidney, lung, and brain; thus, oxidative stress plays an important role in the pathogenesis of these organs. At birth, the newborn infants encounter an increase in oxygen level compared with its intrauterine environment. Especially in the preterm newborn infant, this increase in oxygen level is exacerbated by the low efficiency of natural antioxidant systems in newborn infants. Therefore, in newborn infants, especially the preterm infants are highly susceptible to the risk of oxidative stress, which eventually leads to free radical oxidative damage. Taken together, these results suggest that neonatal brains are easily prone to oxidative stress–related brain damage.

These studies and our previous results warranted us to elucidate a role for edaravone in the mechanisms that are involved in neuroprotection in neonates. Lipid peroxidation is initiated as a result of free radical formation and maintains the generation in a chain reaction that ultimately damages the cell membrane and causes cell death. Inhibition of lipid peroxidation may be effective to prevent free radical–mediated brain damage. Our aim was to examine the antioxidant property of edaravone by investigating its effect on lipid peroxidation and nitric oxide metabolite (NOx) production in neonatal rat brain after hypoxia-ischemia.

**Study design**

**Animals**

Pregnant Wistar rats were purchased from Japan Charles River (Shizuoka, Japan). All the animals were maintained at the Experimental Animal Center, at controlled ambient temperature of 23°C ± 1°C, with 50% ± 10% relative humidity. The rats were allowed to deliver spontaneously, and the pups were reared with their dams until the time of experiment. The experimental design was reviewed and approved by the Committee for the Ethics on Animal Experiments in Miyazaki Medical College, University of Miyazaki. Seven-day-old Wistar rats were divided into 3 groups: control (n = 7 rats), edaravone (n = 24 rats), and saline (n = 24 rats). Edaravone and saline groups underwent surgery; the control group rats were not subjected to surgery and were administered neither edaravone nor saline.

**Surgery**

Seven-day-old Wistar rats (n = 48) were subjected to a modified Levine’s procedure for the production of hypoxic-ischemic injury, as previously reported. In brief, each rat was anesthetized with ether, and the left carotid artery was sectioned between double ligatures with 4-0 surgical silk. Then the rats were allowed to recover for 1 to 2 hours and exposed to 2 hours of hypoxia in a plastic container that was perfused with a mixture of humidified 8% oxygen that was balanced with nitrogen. The temperature inside the container was kept at 33°C, which is the usual temperature while the pup is huddling with the mother.

**Drug administration**

After the rats were exposed to 2 hours of hypoxia, they were immediately administered intraperitoneally with edaravone (n = 24 rats) at dose of 9 mg/kg or saline solution (n = 24 rats). In the present study, the dosage of edaravone at 9 mg/kg was selected according to our previous study in which 9 mg/kg showed a potent neuroprotection of the rat brain after hypoxia-ischemia insults. Edaravone was supplied by Mitsubishi Pharma Co (Tokyo, Japan).

**Aspiration of cerebrospinal fluid (CSF)**

The CSF was aspirated at 2, 5, 24, and 48 hours (n = 6 rats at each point) after hypoxia according to our previous method. In brief, CSF was obtained from rat pups by transcutaneous cisternal puncture with a 27-gauge needle. When the tap contained blood, samples were discarded. All samples were centrifuged for 10 minutes at 10,000 rpm, and the clear supernatant was taken. CSF was stored at −20°C until the time of assay. The level of NOx as an index of nitric oxide production in the brain was measured in the CSF by high-performance liquid chromatography. We did not use brain tissue homogenates to measure NOx production, because these homogenates are prone to aerobic contamination during the process of homogenization.

**Extraction of brain tissue**

The edaravone and saline groups (n = 24 rats per group) were killed at 2, 5, 24, and 48 hours after hypoxia,
immediately after aspiration of CSF. The control rats were killed on day 7. The brain, excluding the cerebellum and the olfactory lobes, were immediately removed and sectioned into 2 hemispheres (right and left), and each hemisphere was weighed. Phosphate-buffered saline solution was added 10 times the weight of each hemisphere and then homogenized by a homogenizer (Polytron PT 3000; Brinkmann Instruments, Inc, Westbury, NY). Brain tissue was homogenized within 8 minutes after death. The homogenates were stored at −84°C until assay.

**Thiobarbituric acid reactive substance (TBARS) assay**

TBARS was measured according to the method of Ohkawa et al. In brief, 0.1 mL of brain homogenate, 0.1 mL of 8.1% sodium dodecylsulfate, 2 mL of 20% acetate buffer (pH 3.5), 1.5 mL of 1.0% sodium thiobarbiturate, and 0.6 mL of distilled water were added and mixed. The mixture was incubated at 100°C for 60 minutes. TBARS were then extracted with a mixture of n-butanol and pyridine (15:1, vol/vol). The fluorescence of the extracted solution was analyzed by spectrofluorometry, under the following condition: excitation wavelength, 515 nm; emission wavelength, 553 nm.

**Statistical analysis**

Statistical analysis was carried out with 2-way (group × time course) analysis of variance, followed by Newman-Keuls’s test for multiple comparisons. A probability value of < .05 was considered significant.

**Results**

The TBARS level of the control was 35.5 ± 2 μmol/mg brain tissue. Figure 1, A, shows the comparison of TBARS level in the ligated side of cerebral hemisphere between the edaravone and saline solution groups. The TBARS level significantly decreased after edaravone administration compared with the saline solution group at 5 and 24 hours after hypoxia (21.2 ± 4.4 μmol/mg vs 225.7 ± 61.8 μmol/mg [P < .01]; 59.2 ± 9.8 μmol/mg vs 273.4 ± 42.8 μmol/mg [P < .01] brain tissue, respectively). No significant difference was observed between the 2 groups at 2 or 48 hours after hypoxia (39.5 ± 3.9 μmol/mg vs 55.3 ± 16.1 μmol/mg; 82.5 ± 9.8 μmol/mg vs 103.8 ± 16.8 μmol/mg brain tissue; respectively). Figure 1, B, shows the comparison of TBARS level in the non-ligated side of the cerebral hemisphere. No significant difference was observed between the edaravone and saline group at any time point. In the saline group, TBARS level in the ligated side was significantly decreased at 48 hours compared with 5 and 24 hours after hypoxia (103.8 μmol/mg ± 16.8 vs 227.7 ± 61.8 μmol/mg [P < .01] and 103.8 ± 16.8 μmol/mg vs 238.4 ± 42.75 μmol/mg [P < .01] brain tissue; respectively).

**Figure 1** Effect of edaravone on lipid peroxidation in neonatal Wistar rats after hypoxic-ischemic insult. The TBARS level was measured in homogenized brain tissue of both right and left hemisphere. A, The sequential changes of TBARS level in the left hemisphere of saline–treated (closed circles), edaravone (open circles), and control group (closed squares). The # denotes P < .01, saline vs edaravone; the * denotes P < .01, control vs saline. B, The sequential changes of TBARS level in the right hemisphere of saline–treated (open diamond), edaravone-treated (closed diamond), edaravone-treated (closed squares), and control group (closed squares). The data represent mean ± SEM. The saline group (n = 6) and edaravone group (n = 6) at each time point; control (n = 7).

The NOx level in the CSF after the administration of edaravone or saline solution after hypoxia. In the edaravone group, the NOx level was significantly decreased compared with saline solution group at only 5 hours after hypoxia (203.3 ± 16.4 pmol/10 μL vs 273.4 ± 8.1 pmol/10 μL; P < .05). No significant difference was observed at the 2-, 24-, and 48-hour point between the edaravone and saline solution groups.
Comment

Our study showed for the first time the inhibitory effects of edaravone on lipid peroxidation and NOx production in the neonatal rat brain after hypoxic-ischemic insult. In acute brain injury, oxygen free radicals are thought to cause oxidative damage to a variety of molecular and cellular targets, which include proteins, DNA, lipids, mitochondria, and membrane structures.11 Oxygen free radicals of potential importance in cerebral ischemia include hydroxyl and superoxide radicals. Among the free radicals, hydroxyl radical is a potent radical and can initiate lipid peroxidation in the cell membrane.12,13 Cell damage can be prevented by detoxification of free radicals, which eventually will prevent the progress of lipid peroxidation. Edaravone has been reported to scavenge the reactive oxygen free radicals such as hydroxyl, superoxide, and peroxyl radicals.14,15

We analyzed the effect of edaravone on the progress of lipid peroxidation in hypoxic-ischemic neonatal brain tissue by measuring the levels of TBARS. The TBARS assay is 1 of the commonly used methods for the measurement of free radical reactions indirectly, is easy to use to study the effects of different treatments on lipid peroxidation, and can be applied to crude biologic extracts. The specificity of TBARS to measure the level of lipid peroxidation has been criticized by certain investigators.16-18 Goda and Marnett17 showed that the quantitative analysis of malonaldehyde level in biologic fluids can be measured by high-performance liquid chromatography with electrochemical detection. Kosugi and Kikugawa18 reported that the TBARS test may not reflect the components of malonaldehyde alone but may also reflect the components of malonaldehyde, alk-2-enals, alka-2,4-diennals, and hydroperoxide functions. In the present experiment, the purpose was not to measure the absolute values of lipid peroxidation but to compare the values of edaravone-treated and saline solution–treated groups. Therefore, we used a relatively crude but widely used assay19-23 that has been accepted as an empiric window for the examination of the complex process of lipid peroxidation in various biologic extracts.22 In our study, the TBARS levels increase in the left hemisphere of the saline solution group. The reason behind this could be re-oxygenation after hypoxic/ischemic insult that resulted in an increase of free radical production, which suggests an increased lipid peroxidation level in the neuronal cell membrane.12 This increased level of lipid peroxidation persisted until 24 hours after the hypoxic insult, which is consistent with the study of Hsiang et al24 in which they showed that lipid peroxidation was significantly high until 24 hours after diffuse brain injury in rats. However, the TBARS level significantly decreased, which can be related to a long-term compensatory mechanism that involves the modulation and synthesis of the activity of enzymes that are related to reactive oxygen species.20 We have shown that an administration of edaravone significantly suppressed the TBARS level until 24 hours after the hypoxic insult. Taken together, our study demonstrates that edaravone can prevent the progress of lipid peroxidation.

There are 3 major mechanisms that are involved in neurotoxicity produced by nitric oxide radicals. First, nitric oxide radical react with super oxide radicals to form highly toxic peroxynitrite anions, which easily decompose to produce hydroxyl radicals and nitrates25,26 and eventually attack the cell membrane. Second, nitric oxide that is produced in the donor neuron can attack the DNA of the target neuron directly. In the target neuron, a great amount of adenosine triphosphate is consumed through poly adenosine diphosphate–ribose synthetase activation for repairing the damaged DNA, which leads to energy depletion and cell death.27 Finally, nitric oxide attacks the mitochondrial electron–transfer system and injures the iron/sulfur protein–containing enzymes that leads to an inhibition of glycolysis and mitochondrial respiration.28 Nitric oxide is a free radical synthesized by nitric oxide synthase (NOS) in the endothelial cells and neurons in response to increased intracellular calcium levels. Nitric oxide has dual action on hypoxic-ischemic brain damage and is not always neurotoxic, in fact nitric oxide also can be neuroprotective. Nitric oxide produced by NOS in the neurons is mainly neurotoxic; however, it is resistant to excitatory stimulation, which sometimes can be neuroprotective. On the
other hand, nitric oxide that is produced by NOS in the neurons gives mainly neuroprotection by vasodilatation and increases microcirculation in the ischemic areas. Nitric oxide that is produced by inducible nitric oxide synthase is mainly neurotoxic to the surrounding tissues.\textsuperscript{29} Citrulline, nitric oxide, and water are produced from arginine, whereas reduced nicotinamide adenine dinucleotide phosphate and oxygen by NOS.\textsuperscript{30-32} Reoxygenation during reperfusion provides oxygen to sustain neuronal viability and also provides oxygen as a substrate for numerous enzymatic oxidation reactions that produce reactive oxidants\textsuperscript{12}; as a result, nitric oxide is synthesized during the reperfusion phase. Nitric oxide radicals are oxidized within seconds to the more stable oxidation products nitrite and nitrate in vivo. Thus, assays of nitrites and nitrate are used widely as indicators of nitric oxide formation. Because direct and selective measurement of nitric oxide is still very difficult, the measuring of NO\textsubscript{x} demonstrates the nitric oxide production more precisely. The generation of nitric oxide is associated predominantly with cytotoxic effects. We therefore investigated the effects of edaravone on the production of NO\textsubscript{x}, which can be any degraded products of nitric oxide metabolite (such as NO\textsubscript{2} and/or NO\textsubscript{3}) after hypoxic-ischemic insult. We showed that the NO\textsubscript{x} level increased in the CSF at 5 hours, which eventually decreased within 24 hours of the insult. This suggests that free radical formation may be a transient phenomenon, whereas the progress of lipid peroxidation continues for a longer duration after hypoxic/ischemic brain damage. Edaravone suppressed the NO\textsubscript{x} levels in the CSF, which indicates the potent scavenging action on NO\textsubscript{x} and thus the prevention of the free radical–mediated brain damage.

Edaravone, for the following reasons, may be the drug of choice in the treatment of neonatal hypoxic/ischemic encephalopathy in newborn infants. Edaravone has shown to have minimal adverse effects since its inception.\textsuperscript{7} It can act in the state of highly oxidative stress,\textsuperscript{1,3-7,12} the condition the newborn infants are prone to in a deficient antioxidant system because of physiologic hyperoxygenation. Last, it can prevent oxidative stress–related pathogenesis in neonatal infarction.\textsuperscript{6} As drug of choice in cerebral infarction in adult patients, edaravone may have a greater beneficial effect on hypoxic-ischemic neonatal brain damage than on adult brain damage.

In conclusion, our study showed 2 important properties of edaravone, which are of great clinical importance in the prevention of neonatal hypoxic/ischemic brain damage. First, edaravone can prevent the progress of lipid peroxidation and, second, can decrease the production of NO\textsubscript{x} in neonatal hypoxia-ischemia. These data indicate that edaravone may be a highly effective and potent drug for the prevention of hypoxic-ischemic encephalopathy in newborn infants.

\textbf{Acknowledgments}

We thank Ms Hiroko Taniguchi for her research assistance.

\textbf{References}


25. Beckman JS, Beckman TW, Chen J, Marshall PA, Freeman BA. Apparent hydroxyl radical production by peroxynitrite: implications for endothelial injury from nitric oxide and superoxide. Proc Nat Acad Sci USA 1990;87:1620-4.


Embryogenesis of fused umbilical arteries in human embryos

Shigehito Yamada, MD,a,b,* Junzo Hamanishi, MD,b Shozo Tanada, MD,d Mitsuhiro Tachibana, MD,e Rokuro Mimura, MD,e Shingo Fujii, MD,b Kohei Shiota, MDa,c

Congenital Anomaly Research Center,a Departments of Gynecology and Obstetrics,b Anatomy & Developmental Biology,c Kyoto University Graduate School of Medicine, Kyoto, Japan; Department of Obstetrics and Gynecology,d Division of Pathology,e Hyogo Prefectural Amagasaki Hospital, Amagasaki, Hyogo, Japan

Received for publication December 17, 2004; revised March 3, 2005; accepted March 30, 2005

KEY WORDS
Umbilical artery
Intrauterine growth retardation
Human embryo

Objective: The objective of this study was to elucidate the embryologic basis of fused umbilical arteries in the human.

Study design: Twenty-nine human embryo specimens at Carnegie stages 11 through 15 (4-5 weeks after fertilization) were examined histologically, with special reference to the development of umbilical arteries.

Results: All embryos at Carnegie stage 11 and 12 had fused umbilical arteries, and 66% of Carnegie stage 13 embryos and 29% of Carnegie stage 14 embryos still had the condition. None of the embryos at Carnegie stage 15 or older had fused umbilical arteries, but there were always 2 arteries present in their umbilical cords.

Conclusion: Our data suggest that (1) a single umbilical artery splits into 2 as the developmental stage of the embryo advances, (2) that fused umbilical arteries represent a remnant of the embryonic phenotype, and (3) that fused umbilical arteries are embryologically distinct from true single umbilical artery.

Obstetric sonographic examination sometimes reveals fused umbilical arteries (FUA) in the fetus, which usually occurs near the placental cord insertion.1 FUA is a condition in which 2 umbilical arteries fuse with each other through their course in the umbilical cord (Figure 1).2 Clinically, FUA is encountered more frequently than single umbilical artery (SUA), which is one of the common cord anomalies in humans, and is often associated with congenital malformations such as musculoskeletal, cardiovascular, and urogenital anomalies (25%-50%).3-6 Because the association with malformations is rare (0.2%-1%) in newborn babies with FUA,1,7 FUA may be pathologically distinct from true SUA. Therefore, this condition should be diagnosed correctly during the prenatal period or at the time of delivery to avoid misdiagnosis as true SUA. Such misdiagnosis...
could be avoided by careful examination of the number of vessels throughout the entire length of the cord. Fox and Benirschke and Kaufmann proposed that SUA should be diagnosed by sectioning the umbilical cord at multiple sites and checking the number of vessels at both the placental and fetal ends at birth, even when the delivery was uneventful. The frequency of FUA is uncertain, but recently, Fujikura examined >700 placentae and estimated the frequency of FUA to be 3.1%. Although only approximately 30 cases of FUA have been reported to date in the literature, Fujikura claimed that this condition is not rare and that many cases are overlooked by obstetricians. The embryologic basis of FUA has not been well illustrated to date. In the present report, we describe the histologic findings of umbilical arteries in staged human embryos at 4 to 5 weeks after fertilization and discuss the embryogenesis of FUA.

Material and methods

Since 1961, >40,000 human conceptuses have been placed in the Kyoto Collection of Human Embryos, in a collaborative effort of several hundred obstetricians. In most of the cases, pregnancies were terminated for social reasons, mostly in healthy women during the first trimester of pregnancy (Maternity Protection Law of Japan). The terminations of pregnancy were performed.

Figure 1 Transverse sections of the umbilical cord of a newborn baby with intrauterine growth retardation. The mother was a 27-year-old nulliparous woman with pregnancy-induced hypertension. The cesarean delivery was performed at 28 weeks of gestation because of the cessation of fetal growth. A-D are arranged in a fetal body to placental sequence. A, Two arteries (arrowhead) and 1 vein are observed in the cord. B, The walls of the arteries are partially fused (arrowhead). C, The arterial walls are continuous (arrowhead). D, Near the placental end of the cord, 1 artery and 1 vein are recognized. The bar equals 1 mm.
Figure 2  Transverse sections of a human embryo at CS12 (Kyoto collection #50622). Sections A-E are arranged in a rostrocaudal sequence. A, Bilateral dorsal aortae are fused in the midline of the embryo. At this stage, the fused portion of the bilateral dorsal aorta is narrow. B, Bilateral dorsal aortae (arrowhead) are branched and distribute to primordial abdominal viscera. In the abdominal wall, umbilical arteries fuse with each other (arrow) and flow into the cordal portion of the umbilical artery. C, Bilateral dorsal aortae (arrowhead) descend in front of the neural tube. In the abdominal wall, 2 umbilical arteries can be seen (arrow) beside the allantois (Al). D, Bilateral dorsal aortae (arrowhead) descend in front of the neural tube. The allantois can be observed between the bilateral umbilical arteries (arrow). Two umbilical veins are fused with each other. E, In the caudal portion of the embryo, the descending dorsal aorta turns toward the abdominal wall and is continuous with the bilateral umbilical arteries (arrowhead). In the umbilical cord, 1 artery and 1 vein can be recognized. F, The illustration shows the course of the umbilical arteries in this embryo. The bilateral dorsal aortae descend in the embryonic body and turn forward to the future pelvic region of the embryo. They are continuous with truncal portion of the umbilical arteries (UAt). The umbilical arteries ascend in the anterior abdominal wall (the future medial umbilical ligaments) and run into the umbilicus. The umbilical arteries are fused with each other where they run into the umbilical cord (UAc). Al, Allantois; CS, connecting stalk; G, primitive gut; NT, neural tube; S, somites; TPC, thoracoperitoneal cavity; UAc, cordal portion of umbilical artery; UAt, truncal portion of umbilical artery; V, umbilical vein. The bar equals 100 μm.
mainly by curettage procedures. Further details of the embryo collection have been described in previous reports.8-11

In the present study, serial histologic sections of 29 human embryos at Carnegie stages (CS) 11 to 15 were examined.12 The distribution of the embryos by CS was 2 cases at CS 11 (27 days after fertilization), 3 cases at CS 12 (30 days), 9 cases at CS 13 (32 days), 7 cases at CS 14 (34-35 days), and 8 cases at CS 15 (36 days).9,13 Their sections were examined serially, with special reference to the umbilical cord and umbilical vessels. Embryos generally were sectioned transversely, but some cases were sectioned in other planes. The planes of section are described in the Figures.

This study, which used the human embryo specimens, was approved by the Institutional Review Board of Kyoto University Graduate School of Medicine.

Results

Umbilical arteries arise from the bilateral internal iliac arteries. During early embryonic stages, 2 aortae exist bilaterally beside the spine (the right and left dorsal aortae). Our histologic observation demonstrated that the 2 dorsal aortae communicate with each other in 4-week human embryos (Figure 2, A) to fuse and become a single aorta. The bilateral dorsal aortae had some branches in the pelvic region and were continuous with umbilical arteries (Figure 2, B-E). The umbilical arteries that run along the anterior abdominal wall (the future medial umbilical ligaments) were found to ascend beside the allantois (Figure 2, C and D), exit the embryonic body, and run into the umbilical cord (Figure 2, E). In younger embryos at CS 11 to 14, the fusion point of umbilical arteries was observed in the umbilical cord (Figure 2, F). At CS 11 and 12, all the embryos that were examined had FUA, as seen in Figure 2, B. The point of fusion of the umbilical arteries was located in the umbilical cord, just outside the abdominal wall. In 6 of 9 CS 13 embryos (66%), umbilical arteries were found to fuse near the abdominal wall (Figure 2, B). In the other 3 CS 13 embryos, 2 arteries were recognized throughout the length of the umbilical cord and fusion of umbilical arteries was not observed. At CS 14, FUA was found in 2 cases among 7 embryos that were observed (29%). A CS 14 embryo is shown in Figure 3.
in which 2 umbilical arteries coalesce into 1 in the cord to form FUA. The aorta curves in the future pelvic region (Figure 3, A) and runs cranioventrally along the neural tube (Figure 3, B). The aorta bifurcates and splits into bilateral umbilical arteries in the abdominal wall (Figure 3, C and D). Shortly after they left the abdominal cavity, the 2 umbilical arteries fused with each other to form FUA (Figure 3, E, F, and G). At CS 15, all the 8 cases that were examined had 2 umbilical arteries (Figure 4) as far as we could examine the cords, although the distal (placental) region of their cord had been removed to variable extents during the operations.

**Comment**

FUA was first described by Chantler et al in a mother with diabetes mellitus and pregnancy-induced hypertension. After their first report, this anomaly has been described mainly as a symptom that is encountered in
ultrasound sonographic prenatal diagnosis. Although only approximately 30 cases of FUA have been reported to date in the literature, Fujikura estimated the frequency of FUA to be 3.1% on the basis of his careful examination of > 700 placentae and claimed that this condition may be overlooked in many cases.

Although most reported cases of FUA are associated with a normal perinatal outcome, the association of FUA with Hallermann-Streiff syndrome and patent urachus has been suggested. Our case that is shown in Figure 1 was associated with severe intrauterine growth retardation, which has not been suggested previously. It is not clear whether such associations are merely due to chance, and further studies are needed to clarify the pathologic significance of these associations.

The development of umbilical arteries was discussed by Monie, on the basis of his examination of umbilical arteries in human embryos of 2.9 to 8.5 mm in crown-rump length. He showed that the right and left umbilical arteries existed in the embryonic body (the truncal portion of the umbilical artery) at all stages of human development, but 2 arteries were not always present in the umbilical cord (the cordal portion of the umbilical artery). He observed that umbilical arteries formed a plexus in the umbilical cord at the earliest stage (2.9-mm embryo) and that later they fused with each other to form a single artery by the 3.4-mm stage. The single artery became “split” into 2 vessels from the truncal toward the placental ends. This splitting process was considered to result in the formation of 2 umbilical arteries.

In our observation of the embryo cases in the Kyoto Collection, all the embryos at CS 11 and 12 had FUA. Their fusion point was recognized in the truncal portion of umbilical arteries. In some cases at CS 13 (66%) and CS 14 (29%), the fusion point of umbilical arteries was observed in the truncal or umbilical portion of umbilical arteries, which seemed more distal than that observed in CS 12 embryos. In the remaining cases at CS 13 or 14, the fusion of arteries was not recognized in the cord. In embryos older than CS 15, the umbilical cord had been cut near the embryonic body, and it was not possible to follow the entire course of umbilical arteries in the cord, but there were always 2 arteries in the cord, as far as we could observe. Therefore, it appeared that the fusion point of the umbilical arteries moves distally from the embryonic end as the developmental stage advances.

FUA is considered to be formed by the failure of the umbilical artery to split into 2 throughout the entire length of the umbilical cord. A major cause of FUA may be Hyrtl’s anastomosis of 2 umbilical arteries, which is observed commonly near the placental end of the cord and is considered to represent a normal physiologic fusion of the arteries. Although Monie categorized a 6.0-mm embryo with FUA as “abnormal,” similarly to those with SUA, we found that normal embryos at CS 13 (4.0-6.0 mm crown-rump length) and CS 14 (5.0-7.0 mm crown-rump length) sometimes had FUA, which most likely represents a normal developmental process. Therefore, it appears that the SUA described by Monie included not only true SUA but also nonpathologic FUA. Our present study is the first study that has examined as many as 29 human embryo specimens at the developmental stage when an SUA splits into 2 and provides a new insight into the development of umbilical arteries.

Acknowledgment

We thank Dr Hiroyuki Koshiyama, Kitano Hospital, Osaka, for stimulating discussion and helpful comments; the collaborating technicians in the Division of Pathology, Hyogo Prefectural Amagasaki Hospital, for their contribution; and Dr Murray Smith, University of New South Wales, Sydney, Australia, for his critical reading of the manuscript and helpful comments.

References

Insulin and fatty acids regulate the expression of the fat droplet-associated protein adipophilin in primary human trophoblasts

Uriel Elchalal, MD, W. Timothy Schaiff, PhD, Steven D. Smith, BA, Eli Rimon, MD, Ibrahim Bildirici, MD, D. Michael Nelson, MD, PhD, Yoel Sadovsky, MD

Departments of Obstetrics and Gynecology and Cell Biology and Physiology, Washington University School of Medicine, St. Louis, MO

Received for publication October 25, 2004; revised March 15, 2005; accepted April 1, 2005

Objective: This study was undertaken to test the hypothesis that insulin and fatty acids regulate adipophilin expression in cultured human trophoblasts.

Study design: Cytotrophoblasts isolated from term human placentas were cultured in the absence or presence of insulin (10 nmol/L), and a mix of oleic and linoleic acid in serum-free medium. The expression of adipophilin as well as the fatty acid transport proteins (FATP) 2, 3, 4 and 6 was examined. Fat accumulation was quantified by BODIPY staining and fat uptake determined using [3H]-oleic acid.

Results: A combination of insulin and fatty acids enhanced the expression of adipophilin (2.3-fold, \( P < .05 \)). In contrast, the expression of FATPs was unchanged. Furthermore, insulin and fatty acids increased the accumulation of fat droplets in trophoblasts by 4- to 5-fold (\( P < .05 \)), but had no effect on oleic acid uptake.

Conclusion: Insulin and fatty acids enhance the expression of adipophilin and the formation of fatty acid droplets in term human trophoblasts.

Supply of fatty acids to the fetus is particularly high during the third trimester of pregnancy, when the fetus more than doubles its size.\(^1\) Because de novo lipid synthesis in the placenta is insufficient, transplacental transport of maternal fatty acids is paramount to the developing fetus. Not surprisingly, maternal serum lipid levels are elevated during pregnancy.\(^1,2\) Although the mix of fatty acids delivered to the fetus is largely determined by maternal serum lipoproteins and fatty acid composition, the placenta is capable of regulated preferential transfer of long chain polyunsaturated fatty acids. In addition, the fetus is unable to synthesize n-3 and n-6 unsaturated fatty acids, and therefore depends on maternal transport for these essential fatty acids.\(^1,2\)

Cell types that rely on the supply of fatty acids, such as adipocytes, cardiocytes and hepatocytes, express specific fatty acid transport proteins. Several fatty acid...
transporters have been identified in trophoblasts, including fatty acid transport proteins (FATPs), placental membrane fatty acid binding protein (p-FABPpnm), and fatty acid translocase (FAT/CD36).1,3 Adipophilin and its murine ortholog, adipocyte differentiation-related protein (ADRP), are associated with cellular lipid droplets and implicated in cellular fatty acid uptake and storage of neutral lipids in adipocytes.4-6 We previously demonstrated that adipophilin is expressed in human villous trophoblasts in vivo and in cultured primary human trophoblasts in vitro.7 Moreover, we showed that the expression of adipophilin is regulated by the insulin-sensitizer peroxisome proliferator-activated receptor-γ (PPARγ),7 a protein that plays a pivotal role in development of the murine fetoplacental unit and differentiation of human trophoblasts.8,9 Consistent with these findings, Tarrade et al10 demonstrated enhanced accumulation of fatty acids during cytotrophoblast differentiation into syncytiotrophoblasts. Ablation of PPARγ is associated with a reduced number of fatty acid droplets in the murine placenta.8 Taken together, these findings suggest that accumulation of fatty acids in trophoblasts is necessary for fetal growth and is regulated during placental development and trophoblast differentiation.

The placenta in women with poorly controlled gestational diabetes mellitus (DM) and type-2 DM is exposed to hyperinsulinemia11 and increased supply of nutrients. In the current work we surmised that insulin, acting in a hyperlipidemic environment, regulates fatty acid accumulation in human trophoblasts. We hypothesized that insulin combined with fatty acids regulate the expression of adipophilin and fat accumulation in trophoblasts. To test this hypothesis we analyzed the influence of insulin, in the presence or absence of fatty acids, on adipophilin expression as well as fatty acid uptake and accumulation in cultured term human trophoblasts.

Materials and methods

Trophoblast isolation and culture

Our study was approved by the Institutional Review Board of Washington University School of Medicine. Placentas were obtained immediately after term singleton deliveries after uncomplicated pregnancies. Cytotrophoblasts were isolated by the trypsin-DNase, percoll (Sigma, St. Louis, MO) gradient centrifugation method described by Kliman et al,12 with modifications.13 Non-adherent cells and syncytiotrophoblast fragments were removed after 4 hours by washing 3 times with phosphate-buffered saline (PBS) solution, and cytотrophoblasts were cultured in Earle’s medium 199 supplemented with 10% fetal bovine serum (Hyclone, Logan, UT) and antibiotics in a standard tissue culture atmosphere of 5% CO2 and 20% O2. After 24 hours in culture the cells were washed and the medium replaced with fresh serum-containing or serum-free medium, supplemented where indicated by insulin (10 nmol/L, Sigma) dissolved in 0.1% insulin-free and fatty acid-free albumin (Sigma), a mixture of 400 μmol/L oleic acid and 800 μmol/L linoleic acid in albumin (Sigma), or vehicle control. These concentrations were selected on the basis of published information and the company’s recommendations, and were designed to assure that the saturating levels of fatty acids were well dissolved in the culture medium. All cells were harvested 48 hours after plating. The phosphatidylinositol 3-kinase (PI3K) inhibitor LY 294002 (10 μmol/L, Cell Signaling Technology, Beverly, MA) was added to some of the cell cultures.

RNA and protein expression

At the end of the culture period the plates were rinsed with PBS. RNA was purified with the use of TriReagent (Molecular Research Center, Inc, Cincinnati, OH) according to the manufacturer’s instructions. DNase I (Ambion, Austin, TX) was added to the purified RNA after plating. The phosphatidylinositol 3-kinase (PI3K) inhibitor LY 294002 (10 μmol/L, Cell Signaling Technology, Beverly, MA) was added to some of the cell cultures.

The fold change values were calculated by using geometric means of individual assays.7 Samples consisted of 20 to 30 μg per lane, which were analyzed with the use of monoclonal anti-adipophilin antibody (1:100 RDI, Flanders, NJ) or polyclonal goat anti-β-actin (1:1000, Santa Cruz Biotech, Santa Cruz, Calif), followed by the corresponding horseradish peroxidase-linked secondary antibodies (1:2000, Santa Cruz). Chemiluminescence was analyzed with the use of Epichemi-3 darkroom (UVP BioImaging...
System, Upland, CA). Protein density was normalized to β-actin expression.

**Lipid accumulation assay**

After 2 washes in PBS, the cells were fixed in 2% paraformaldehyde for 20 minutes at room temperature. Each plate was washed 2 additional times and the cells permeabilized with 1 mL of Triton-X-100 0.1% in PBS for 5 minutes, followed by 2 PBS washes. The wells were covered with BODIPY 493/503 (Molecular Probes, Eugene, OR), which was dissolved in ethanol for a stock solution of 0.1%, and further diluted 1:100 in PBS before assay. After 15 minutes of exposure to BODIPY, the cells were washed 4 times in PBS. Cell nuclei were counterstained with TO-PRO 3-iodide 642/661 (Molecular Probes), rinsed in PBS, mounted with GelMount (Biomeda, Foster City, CA), and viewed under a Nikon E800 microscope equipped with epifluorescence optics and a Nikon C1 confocal imaging package (Nikon, Tokyo, Japan). Images (n = 18) of 6 random 0.25 mm² microscopic fields were captured, and the number of cells stained for BODIPY 493/503 along with the total number of nuclei for each field was recorded. Lipid droplets were quantified from digital images by using Analysis digital imaging software (Soft Imaging Corp, Lakewood, CA). Nuclei were manually counted from the same images. Each experiment was performed in duplicate.

**Fatty acid uptake assay**

Transport of fatty acids into cultured trophoblasts was assayed in triplicates with radiolabeled oleic acid with the use of a modification of a previously described assay. Uptake medium consisted of a 1:1 ratio of fatty acid-free bovine serum albumin (BSA) (Sigma) and unlabeled oleic acid (Nu Chek Prep, Elysian, MN) at 80 μmol/L each, spiked with [3H]-oleic acid (Perkin-Elmer, Boston, MA) at 1 × 10^7 dpm/mL. The transport medium was prepared by adding the unlabeled and tritiated oleic acid dropwise with stirring to a 60°C solution of BSA in Hank’s balanced salt solution (HBSS, Sigma). The solution was stirred until cleared, then allowed to cool to room temperature. Trophoblasts were cultured in 6-well plates for 24 hours, and then changed to serum-free medium supplemented with insulin or fatty acids, as noted. At the end of the culture period the cells were incubated 1 hour at 37°C in serum-free Dulbecco modified Eagle (DME) medium (Invi- trogen/Gibco, Carlsbad, CA). After incubation, the culture plates were allowed to equilibrate to room temperature for 10 minutes. Transport medium (800 μL) was added to each well after HBSS rinse, and incubated at room temperature for 3 minutes. The uptake was stopped by the addition of 3.5 mL per well ice-cold stop solution that consisted of 0.1% fatty acid-free BSA and 500 μmol/L phloretin (Sigma) in PBS. The wells were rinsed 2 additional times with ice-cold stop solution, the cells were lysed with 500 μL of 0.2% sodium dodecyl-sulfate (SDS) and the lysates added to 10 mL ScintiVerse II scintillation fluid (Fisher Scientific, Fair Lawn, NJ). The wells were rinsed with additional 500 μL 0.2% SDS and pooled into their respective scintillation vials. Samples were counted on a Packard 2500TR Liquid Scintillation Analyzer (Packard, Downers Grove, IL). Nonspecific association of labeled fatty acid with trophoblasts was determined with the use of duplicate cultures assayed in the presence of 500 μmol/L phloretin. In these cultures, phloretin was added to the incubation solution, washes, and transport medium. Specific transport of fatty acids was calculated by subtracting counts in the presence of phloretin from counts in the absence of phloretin.

### Table 1

<table>
<thead>
<tr>
<th>Name</th>
<th>Accession no.</th>
<th>Real-time quantitative PCR</th>
<th>Direction and site sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adipophillin</td>
<td>NM_001122</td>
<td>F(326~346)</td>
<td>GCGAGAAGACGTTGAAGAC</td>
</tr>
<tr>
<td></td>
<td>gi: 34577058</td>
<td>R(390~372)</td>
<td>TCTGAATGATGGCCAGAC</td>
</tr>
<tr>
<td>FATP-1</td>
<td>NM_198580</td>
<td>F(2952~2973)</td>
<td>CCGGAATTGACTGGAACCCTT</td>
</tr>
<tr>
<td></td>
<td>gi: 38524615</td>
<td>R(3065~3048)</td>
<td>CACGAGTCAGAGGTTC</td>
</tr>
<tr>
<td>FATP-2</td>
<td>NM_003645</td>
<td>F(1829~1852)</td>
<td>CTTTGACACTGATGTGACTAT</td>
</tr>
<tr>
<td></td>
<td>gi: 23111065</td>
<td>R(1908~1884)</td>
<td>CATGTAGTCATAGTGCTGCTGAT</td>
</tr>
<tr>
<td>FATP-3</td>
<td>NM_024330</td>
<td>F(2098~2122)</td>
<td>CAAGAGACCTCCTAAACGACAGAAT</td>
</tr>
<tr>
<td></td>
<td>gi: 13236578</td>
<td>R(2180~2158)</td>
<td>GAAGGCAAAGGTGCGACAG</td>
</tr>
<tr>
<td>FATP-4</td>
<td>NM_005094</td>
<td>F(460~478)</td>
<td>GCCGAAGTCAGAGGCAAA</td>
</tr>
<tr>
<td></td>
<td>gi: 40807356</td>
<td>R(540~522)</td>
<td>CCGTGGTGGCCTTCTCA</td>
</tr>
<tr>
<td>FATP-6</td>
<td>NM_014031</td>
<td>F(542~559)</td>
<td>ACCGCGCGGATGCCAA</td>
</tr>
<tr>
<td></td>
<td>gi: 13325054</td>
<td>R(606~590)</td>
<td>CCGGCTAGAGGTTGAATCTC</td>
</tr>
<tr>
<td>18S</td>
<td>M10098</td>
<td>F(1049~1069)</td>
<td>CGAACCCTGACTTCCGTTCT</td>
</tr>
</tbody>
</table>
Statistical analysis

Data are expressed as mean ± SEM. All comparisons were made by either paired t test or analysis of variance (ANOVA) with post-hoc Student–Newman–Keuls test, where appropriate (Primer of Biostatistics McGraw-Hill NY, NY). Significance was determined at $P < .05$.

Results

We initially examined the influence of insulin on the expression of adipophilin messenger RNA (mRNA) in human trophoblasts, derived from normal term pregnancies and cultured over a 48-hour period. By using real-time quantitative PCR, we found that the effect of insulin on the expression of adipophilin mRNA was maximal after 4 hours in culture and at a concentration of 10 nmol/L (Figure 1, A and B). Notably, these experiments were performed in culture medium that contained 5 mmol/L of glucose and serum. To determine the influence of serum components on adipophilin mRNA expression, we repeated the experiment in the absence of 10% fetal bovine serum. The medium also contained 25 mmol/L of glucose, a mixture of 400 μmol/L oleic acid and 800 μmol/L linoleic acid, or both. As shown in Figure 1, C, in the absence of serum the expression of adipophilin mRNA was increased in all paradigms. In contrast, the effect of glucose or fatty acids was insignificant.

To examine the influence of insulin and fatty acids on adipophilin expression, we exposed the cells to insulin in serum-free medium, in the absence or presence of the fatty acid mixture. As shown in Figure 2, A, the combination of insulin and fatty acids resulted in enhanced adipophilin mRNA expression. The effect of insulin or fatty acids alone was insignificant. Interestingly, basal as well as induced adipophilin mRNA levels were reduced by using the PI3K inhibitor LY 294002. A Western analysis of adipophilin expression in trophoblasts provided further support to our findings, demon-

**Figure 1** The influence of insulin on the expression of adipophilin mRNA in cultured term human trophoblasts. After culture in serum-containing medium the cells were washed and cultured for an additional 24 hours in the absence or presence of insulin. **A**, Time-dependent effect of insulin (10 nmol/L). The x-axis indicates the period of insulin exposure before harvest. **B**, Concentration-dependent effect of insulin, added 4 hours before harvest. **C**, The effect of culture medium conditions on adipophilin mRNA. The effect of added glucose (25 mmol/L) or fatty acids (400 μmol/L oleic acid and 800 μmol/L linoleic acid) was insignificant. The reduced expression of adipophilin by addition of 10% fetal bovine serum was significant at all experimental paradigms ($P < .05$, paired t test). Each result was determined in duplicate, and represents the mean ± SEM obtained from 3 different placentas.
Straining enhanced expression of adipophilin protein in response to a combination of insulin and fatty acids (Figure 2, B). To assess whether the effect of insulin and fatty acids was specific to adipophilin we measured the expression of mRNA of several FATPs that are expressed in trophoblasts. Notably, FATP5 is not expressed in human trophoblasts, and the expression of FATP1 was too low for meaningful conclusions (data not shown). As shown in Figure 3, the influence of insulin and fatty acids on mRNA for FATP transcripts was statistically insignificant. Together, these data demonstrate that insulin combined with fatty acids selectively upregulate the expression of adipophilin, and that basal as well as induced expression of adipophilin depend on PI3K activity.

Because adipophilin is associated with formation of fatty acid droplets, we examined the influence of insulin and fatty acids on fatty acid droplets in trophoblasts. For this purpose we stained trophoblasts with BODIPY fluorophore 493/503, which specifically stains cellular lipid droplets. As shown in Figure 4, we found that exposure to insulin and fatty acids for 6 hours increased the number and total area of fatty acid droplets in cultured trophoblasts. Similar results were observed
Figure 4  The influence of insulin and fatty acids on the level of fatty acid droplets in cultured trophoblasts. A, Representative photomicrographs of primary trophoblasts (n = 3). Cultures were exposed for 6 hours to insulin, fatty acids or both, until harvested at 48 hours. Cells were stained by BODIPY fluorophore 493/503 for fat droplets (green) and TO-PRO (642/661) for nuclei (blue). Bar = 10 μm. B, Quantitative analysis of the number or area of fatty acids droplets in trophoblasts, performed as described in Materials and methods. The influence of combined insulin and fatty acid supplementation was significant (P < .05, ANOVA). Similar results were obtained when the cells were exposed to insulin and fatty acids for 24 hours (not shown).
after 24 hours of exposure to insulin and fatty acids (not shown). To determine whether stimulation of fatty acid uptake by insulin underlies fatty acid accumulation in trophoblasts we analyzed the influence of insulin, in the absence or presence of fatty acids, on trophoblast uptake of tritiated oleic acid. As shown in Figure 5, we found that the uptake of oleic acid (15%) by LY 294002 in the absence of insulin was significant ($P < .05$, paired $t$ test).

**Comment**

We found that a combination of insulin and the long chain fatty acids oleic and linoleic acid increase adipophilin expression in cultured primary term human trophoblasts. This effect was specific, and demonstrated using analysis of both mRNA and protein. The time course and concentration dependence of insulin were similar to those we previously established. We noted that either insulin or fatty acids alone exhibited a weak and statistically insignificant influence on adipophilin expression. Because PI3K regulates pivotal pathways in insulin signaling, we tested for the effect of LY 294002, a selective inhibitor of PI3K, on adipophilin expression. Interestingly, we found that LY 294002, inhibited both basal and induced expression of adipophilin, suggesting a role for PI3K in regulation of adipophilin expression in trophoblasts. However, the fold induction of adipophilin expression by insulin and fatty acids was unchanged, even in the presence of LY 294002. Consistent with this finding, LY 294002 had a weak inhibitory effect on oleic acid uptake in the presence of fatty acids, but not insulin. In addition, we found that the enhancement of adipophilin expression by combined insulin and fatty acids was associated with fat accumulation in trophoblasts but not with increased uptake of fatty acid. These data suggest that insulin and fatty acids enhance fat storage in trophoblasts, and that PI3K may play a role in adipophilin expression and fat accumulation in these cells.

Fat readily diffuses through cell membranes, including placental trophoblasts. In addition, several types of fatty acid transporters are expressed in the membrane and cytoplasm of diverse cell types, where they facilitate the transfer and intracellular channeling of fatty acids across membranes. Insulin regulates the expression and subcellular location of different fatty acid transporters in adipocytes, myocytes and cardiocytes, leading to enhanced fatty acid uptake by these cells. Although adipophilin plays a role in cellular fat uptake, its regulation is incompletely understood. Oxidized low-density lipoproteins (LDL) increase adipophilin expression in macrophage foam cells and monocytes. Enhanced expression of adipophilin in macrophages by acetylated LDL promotes triglyceride and cholesterol storage and reduces cholesterol efflux. We recently demonstrated that PPAR$\gamma$ increases the level of adipophilin expression in trophoblasts. Consistent with this notion, insulin stimulates the activity of PPAR$\gamma$ in a ligand independent fashion. In addition, the expression of adipophilin, but not the lipid-associated proteins perilipin and S3-12 proteins, is increased in obese but not lean Zucker rats. Together, these findings suggest that insulin and fatty acids influence adipophilin expression and fat accumulation in trophoblasts through regulation of the transcriptional activity of PPAR$\gamma$.

We used a long chain fatty acid mix of oleic and linoleic acids. These fatty acids represent only a fraction of fatty acid components in maternal blood. Other fatty acids, as well as cholesterol and triglycerides, may also play an important role in total fat transport and uptake by trophoblasts. Similarly, we did not analyze the influence of insulin and/or fatty acids on the expression and function of other fat transporters. Furthermore, we did not pursue the subcellular location of the transporters. These variables may be relevant to the activity of several transporters, such as CD36/fatty acid translocase. The regulation of fatty acid transporters in

---

**Figure 5** The effect of insulin and fatty acids on the uptake of [$^{3}$H]-oleic acid in cultured trophoblasts ($n = 3$), performed in duplicates in the absence or presence of LY 294002 (10 $\mu$mol/L) as described in Materials and methods. Only the reduction in oleic acid uptake (15%) by LY 294002 in the absence of insulin was significant ($P < .05$, paired $t$ test).
trophoblasts may have direct implications for pregnancy-related hyperinsulinemia and its effect on placental substrate availability.

Acknowledgments

We thank Elena Sadovsky and Lori Rideout for technical assistance and manuscript preparation.

References

12. Kliman HJ, Nestler JE, Sermasi E, Sanger JM, Strauss JM. Purification, characterization and in vitro differentiation of cyto-
Objective: In this study, human preterm amnion cells were investigated in 3-dimensional (3D) cell-matrix culture systems in an attempt to design therapeutic strategies for preterm premature rupture of the membranes.

Study design: Three-dimensional collagen I and fibrin cell–containing biomatrices were created to mimic the architecture of native amnion. Amnion mesenchymal cells were embedded in 3D matrices, and epithelial cells were placed on top of these matrices. Cell viability and morphology were visualized by DiI-ac-LDL, F-actin, and nuclear staining. Proteolytic activity of matrix metalloproteinases (MMPs) was investigated using gelatine zymography.

Results: Preterm amnion epithelial and mesenchymal cells cultured in collagen I and fibrin matrices assume cell morphologies similar to those observed in vivo. Mesenchymal cells were capable of remodelling collagen I, as seen by extensive volume contraction, by 40% at day 1 and 80% at day 5. Matrix contraction was independent of the presence of epithelial cells, and could not be inhibited by GM6001 and/or aprotinin. No contraction was observed in fibrin matrices over 8 days. The migratory response of mesenchymal cells cultured in 3D fibrin matrices supplemented with fibronectin was associated with specific activated MMP-9.

Conclusion: Three-dimensional fibrin matrices might be useful in amnion cell tissue engineering, including cell-matrix transplantation.

Preterm premature rupture of the membranes (PPROM) occurs in approximately 1% of all pregnancies, and is associated with 30% to 40% of preterm deliveries. It is the leading identifiable cause of preterm delivery. Both PPROM and preterm birth carry a high risk of maternal morbidity and neonatal morbidity and mortality. Several treatments for premature rupture of the membranes...
(PROM) in human fetal membrane models have been proposed. The application of maternal blood clot patches has been recognized as a potential option for therapy of PPROM. However, the intra-amniotic deposition of a mixture of platelets in fibrin cryoprecipitate, called 'amnio-patch,' as well as a transvaginally applied intracervical fibrin sealant, have been only partially successful in treatment of PPROM. A recent single case report for treatment of PPROM used a collagen plug in combination with additional fixation with fibrin glue adhesives. Although the sealing was initially effective, PPROM reoccurred after two weeks. Despite their apparent potential, none of these treatment methods have been introduced into routine clinical use. Failure of those trials may have resulted from the lack of integration of the plugs at the site of regeneration, and the lack of a subsequent healing response. Based on the literature, the healing response after implantation of the plug strongly depends on recruitment of cells from adjacent native tissue. In addition, the healing process may be facilitated and accelerated by tissue-specific cells provided within a biocompatible plug on which these cells can proliferate.

The aim of this study was to investigate the behaviour of human amnion cells in 3-dimensional (3D) collagen I and fibrin matrices, with the ultimate objective to develop a clinically relevant cell-matrix system for treatment of PPROM.

**Material and methods**

**Isolation and culture of human preterm amnion epithelial and mesenchymal cells**

Preterm amnion epithelial and mesenchymal cells were isolated and cultured as described previously. Amnion tissues were collected immediately after elective cesarean section from 6 preterm (28–36 week) placentas with attached membranes from women without PROM, signs of infection, or chromosomal abnormalities. The main indications for the elective cesarean sections were preeclampsia, intrauterine growth retardation, and placenta previa. Amnion and chorion leaf tissues were separated by blunt dissection, and the amnion was cut approximately 2 cm from the placental disc to avoid the ‘zone of altered morphology.’ The samples were washed in phosphate-buffered saline (PBS, Sigma, Buchs, Switzerland) to remove cellular debris and blood, minced, and treated 4 times with 0.25% trypsin (Gibco BRL, Basel, Switzerland) for 15 minutes. The first digestion supernatant, consisting primarily of red blood cells, was discarded. Epithelial cells obtained during subsequent trypsin treatments were collected by centrifugation at 1000 rpm for 5 minutes, and resuspended in a basal medium comprising a 1:1 mixture of Ham’s F-12 and Dulbecco’s modified Eagle’s medium (DMEM, Gibco) containing 10% heat-inactivated fetal bovine serum (FBS, Gibco), 100 U/mL penicillin (Gibco), and 100 μg/mL streptomycin (Gibco). Mesenchymal cells were isolated from the remaining amnion tissue after complete (>98%) removal of epithelial cells (evaluated microscopically). The tissue was minced and incubated with 2 mg/mL collagenase A (Roche Diagnostics, Rotkreuz, Switzerland) at 37°C for 120 minutes. Dispersed mesenchymal cells were collected by centrifugation at 1000 rpm for 5 minutes. Epithelial and mesenchymal cell viability as assessed by Trypan blue dye exclusion was found to be over 97%. Both cell types were cultured further in basal medium supplemented with 50 ng/mL epidermal growth factor (EGF, Sigma), 2.5 μg/mL insulin (Sigma), 5 μg/mL transferrin (Sigma), and 0.1 ng/mL triiodothyronine (T3, Sigma). This medium is further referred to as 'supplement medium.' The cells were maintained in culture until confluence in a humidified atmosphere at 37°C and 5% CO2.

**Coculture of preterm amnion epithelial and mesenchymal cells in 3D collagen I matrices**

The 3D collagen I matrices used for coculture of embedded preterm amnion epithelial and mesenchymal cells were used to create a cell-matrix system that mimics the structure of native amnion (Figure 1). Preterm amnion mesenchymal cells (1 × 10^6) were mixed with 100 μL neutralized collagen I solution (BD Biosciences, Bedford, MA). The final concentration of collagen I solution was 2 mg/mL. Samples were placed in 96-well MicroWell® plates (Nunc, Wiesbaden, Germany). Polymerization was allowed to proceed for 15 minutes at 37°C. Subsequently, 1 × 10^6 preterm amnion epithelial cells were added on top of the collagen I matrices. Cell-matrix systems were cultured in 'supplemented basal medium' at 37°C, 5% CO2, in a humidified atmosphere.

**Assessment of global collagen I matrix contraction**

Eight experimental groups of collagen I matrices were examined as outlined in Table I. These 8 experimental groups were performed using epithelial and mesenchymal cells of 3 different preterm amnion membranes. Mesenchymal cells were either embedded into 3D collagen I matrices (Group I, III, V, VII), or matrices containing mesenchymal cells were covered by epithelial cells (Group II, IV, VI, VIII). In some indicated cases GM6001 (an inhibitor of MMP-1, MMP-2, MMP-3, MMP-8, and MMP-9, Chemicon, Juro Supply GmbH, Lucerne, Switzerland) and/or aprotinin (an inhibitor of serine proteases, Sigma) were mixed into the collagen I solution before polymerization at a final concentration of 10 μmol/L and 50 μg/mL, respectively. Native cell-free collagen I matrices were used as controls. The size of 3D collagen I matrices was analyzed by phase contrast and fluorescence microscopy using a Leica DMIL (Leica, Glattbrugg, Switzerland) equipped with a Leica DC 300F.
digital camera (Leica). Collagen I matrix contraction was monitored daily and processed using LeicaQ Win Image Analysis software (Leica Imaging System, Ltd, Cambridge, England, UK) and Photoshop® 7.0 (Adobe Systems Inc, San Jose, CA). The size of 3D collagen I matrices were calculated by subtracting the matrix size in square micrometers from the size measured at t = 0 (0 hours). Values were expressed as a percentage of the size at t = 0 (100%).

Preparation of fibrin matrices

Fibrin matrices were prepared according to standard protocols by mixing the following components to the final concentrations: 2 mg/mL fibrinogen (Fluka AG, Buchs, Switzerland) in 10 mmol/L Tris-buffered saline, pH 7.4, 0.5 U/mL factor XIII (kindly provided by Baxter AG, Vienna, Austria), 2 NIH U/mL human thrombin (Sigma), and 2.5 mmol/L CaCl2. Collagen I, fibronectin (Bireoeba, Basel, Switzerland), or laminin-1 (Sigma) were included at 20 mg/mL gel before initiation of fibrin polymerization by addition of thrombin (Sigma).

Coculture of preterm amnion epithelial and mesenchymal cells in fibrin-based hydrogel matrices

Three-dimensional fibrin matrices were performed using epithelial and mesenchymal cells from 3 preterm amnion membranes. Preterm mesenchymal cells (2 × 10^4) were suspended in 200 μL fibrinogen solution containing different extracellular matrix molecules at 20 μg/mL before initiation of fibrin polymerization. Samples were placed in 8-well glass culture dishes (Nunc Lab-Tek TMII Chamber slide system, VWR International AG, Dietikon, Switzerland). Polymerization induced by thrombin was allowed to continue for 5 minutes at 37°C. Subsequently, the surface of 3D fibrin matrices were coated with either collagen I, fibronectin, or laminin at 20 μg/mL, and allowed to polymerize for another hour at 37°C. Thereafter, 2 × 10^4 preterm amnion epithelial cells were seeded on top of the 3D matrices (Figure 1). Cells-matrix systems were cultured for 8 days as described above.

Cytochemical analysis

To detect the uptake of 1,1'-dioctadecyl-3,3,3',3'-tetramethylindocarbocyanine-labeled acetylated LDL (DiL-ac-LDL, BTI, Inc, Stoughton, Mass), preterm amnion epithelial and mesenchymal cells were incubated with DiI-ac-LDL (2.4 μg/mL) at 37°C for 1 hour. Cells were then fixed with 2% paraformaldehyde for 10 minutes, washed 3× with PBS, permeabilized with 0.2% Triton X-100, and incubated with rhodamine-labelled phalloidin (Molecular Probes, Eugene, Ore) in order to visualize F-actin filaments. Finally, 4′,6-diamidino-2-phenylindole, dihydrochloride (DAPI, Molecular Probes) was used to stain the nuclei of all cells. The fluorescently labelled cells...
were imaged with a confocal laser scanning microscopy (Leica SP2, Leica). In some indicated cases, the data were further processed using Imaris software (Bitplane AG, Zurich, Switzerland).

**Gelatin zymography**

Gelatin zymography was used to investigate the proteolytic activities of MMPs expressed by preterm amnion mesenchymal cells embedded into 3D fibrin matrices. Collagen I, fibronectin, or laminin-1 was included into fibrin matrices before coagulation. Native fibrin matrices without additional extracellular matrix molecules were used as controls. The culture supernatants were collected at day 1, 3, 5, and 7. The gelatinolytic activities of MMPs in cell-matrix culture supernatants were assessed as previously described.\(^{16}\) Briefly, 10 \(\mu\)g of total proteins were applied onto 10% SDS-polyacrylamide gels copolymerized with 1 mg/mL gelatin and electrophoretically separated. Recombinant human MMP-2 and MMP-9 (Chemicon International, Inc, Temecula, CA) were used as gelatine zymography standards. The gels were subsequently washed twice for 20 minutes with 2.5% Triton X-100 to remove SDS and to allow the MMPs to renature before incubating in zymography buffer (20 mmol/L Tris, pH 8.0, 5 mmol/L calcium chloride, 0.02% sodium azide). The gels were incubated for 20 hours at 37°C with agitation. The zymographic gels were scanned using a Silverfast Expr1460 scanner equipped with SilverFast v5.52r09 software (LaserSoft Imaging AG, Kiel, Germany) and compared with each other.

**Statistical analysis**

Mean values and standard deviation (SD) are reported. Differences between 2 groups were analyzed using the Mann-Whitney \(U\) test. The Kruskal-Wallis test was used to test the differences between several groups. Significance level was set on \(P < .05\).

**Results**

**Preterm amnion mesenchymal cells contract collagen I matrices**

The use of amnion cell cultures in collagen I matrices was challenged by the extensive collagen I matrix contraction mediated by amnion mesenchymal cells. A significant decrease in the size of anchored collagen I matrices was detected over the course of 6 days (\(P < .001\)). In fact, the size of collagen I matrices was reduced by 40% at culture day 1 and 80% at culture day 5. This effect was independent of the presence of cocultured preterm amnion epithelial cells (Figure 2A). Furthermore, collagen I contraction could not be blocked by GM6001 (10 \(\mu\)mol/L) or aprotinin (50 \(\mu\)g/mL) (Table II). In the absence of amnion cells, no change in collagen I matrix size was detectable (Figure 2A). Additionally, migration of living preterm amnion mesenchymal cells out of zones of dense collagen I into zones of low collagen I concentrations was observed (Figure 2B).

In summary, collagen I matrices significantly contract with increasing culture time, and preterm amnion mesenchymal cells are capable of remodelling and migrating out of 3D collagen I matrices.

**Coculture of preterm amnion epithelial and mesenchymal cells within native and modified fibrin hydrogel matrices**

As illustrated in Figure 3A to C, fibrin matrix supplemented with fibronectin maintained the overall morphology of cocultured amnion epithelial and mesenchymal cells. The overall cell morphology was visualized by co-staining for both F-actin and cell nuclei. At culture day 2, amnion mesenchymal cells were sparsely distributed within the matrix and lacked prominent extensions (Figure 3A). In contrast, at culture day 8, amnion mesenchymal cells displayed formation of the typically migratory

---

### Table I Experimental groups of 3D collagen I matrices. Mesenchymal cells were embedded into collagen I matrices (Group I, III, V, VII) or mesenchymal cells were embedded and epithelial cells were seeded on top of these matrices (Group II, IV, VI, VIII). In some indicated cases the matrices were incubated with 10 \(\mu\)M GM6001 (Group III, IV, VII, VIII) and/or with 50 \(\mu\)g/ml aprotinin (Group V, VI, VII, VIII). Native cell-free collagen I matrices were used as control

<table>
<thead>
<tr>
<th>Groups of 3D collagen I matrices</th>
<th>Mesenchymal cells</th>
<th>Epithelial cells</th>
<th>GM6001 (10 (\mu)M)</th>
<th>Aprotinin (50 (\mu)g/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>II</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>III</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>IV</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>V</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>VI</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>VII</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>VIII</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Control</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
Figure 2  (A) Contraction of collagen I matrices mediated by preterm amnion mesenchymal cells. The size of collagen I matrices was reduced by 40% at day 1 and 80% at day 5. This effect was independent of the presence of co-cultured preterm amnion epithelial cells. In the absence of amnion cells, no change in collagen I matrix size was observed. Matrix sizes are expressed as percentages of initial matrix size at day 0 and are given as mean ± SD of three experiments. * indicates statistical significance of the condition ‘cell free-collagen I matrices’ vs condition ‘collagen I matrices containing mesenchymal and epithelial cells’ (p < 0.001). Scale bar: 1mm. (B) Migration of living preterm amnion mesenchymal cells (arrowhead) from a zone of dense collagen I matrix (arrow) into zones lacking collagen I after contraction of the matrix. The dashed line indicates, approximate, the border of contracted collagen I matrix. Living, migrating cells were stained with Dil-ac-LDL. Scale bar: 250 μm.
Table II  Contraction of collagen I matrices mediated by preterm amnion mesenchymal cells (Group I and II) could not be blocked by 10 μM GM6001 and/or 50 μg/ml aprotinin (Group III, IV, V, VI, VII and VIII). In the absence of amnion cells (Control), no change in collagen I matrix size was observed

<table>
<thead>
<tr>
<th>Groups of 3D collagen I matrices</th>
<th>Average of collagen I matrix size (%)</th>
<th>Matrix contraction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day 0</td>
<td>Day 1</td>
</tr>
<tr>
<td>I</td>
<td>99±1</td>
<td>58±5</td>
</tr>
<tr>
<td>II</td>
<td>99±1</td>
<td>57±11</td>
</tr>
<tr>
<td>III</td>
<td>99±1</td>
<td>66±5</td>
</tr>
<tr>
<td>IV</td>
<td>99±1</td>
<td>65±7</td>
</tr>
<tr>
<td>V</td>
<td>99±1</td>
<td>60±6</td>
</tr>
<tr>
<td>VI</td>
<td>99±1</td>
<td>61±8</td>
</tr>
<tr>
<td>VII</td>
<td>99±1</td>
<td>63±8</td>
</tr>
<tr>
<td>VIII</td>
<td>99±1</td>
<td>64±6</td>
</tr>
<tr>
<td>Control*</td>
<td>99±1</td>
<td>99±1</td>
</tr>
</tbody>
</table>

* Indicates values that are significantly different from cell containing collagen I matrices. Matrix sizes are expressed as percentages of initial matrix size at day 0, and are given as mean ± SD of three experiments.

phenotype with long cellular extensions (Figures 3B–C). As judged by DAPI staining, the number of mesenchymal cells was not significantly changed with increasing culture time (P = .556). At culture days 2 and 8, the mean number of mesenchymal cells was 135 ± 7 and 140 ± 14 cells/0.5 mm², respectively (Figure 3D). This finding indicates that preterm amnion mesenchymal cells in 3D fibrin matrices undergo differentiation rather than proliferation. Furthermore, fibrin matrices supplemented with fibronectin maintained the vitality of amnion cells. The vitality of amnion cells was not significantly changed over the time course of 8 days (P = .444). At culture day 2, 87% ± 9% of embedded mesenchymal cells (Figure 3D) and 89% ± 7% of epithelial cells were positive for DiI-ac-LDL uptake, determined as an indicator for cell-viability. Consistently at day 8, 84% ± 11% of mesenchymal cells (Figure 3D) and 86% ± 5% of epithelial cells displayed viability. The amnion epithelial cell layer was formed on top of 3D fibrin matrices supplemented with fibronectin. Amnion epithelial cells appeared similar to those observed in vivo (data not shown). Notably, no significant decrease in the size of fibrin matrices was detected over the course of 8 days (P = .222) (Figure 3E). Additionally, the overall morphology of amnion epithelial and mesenchymal cells cultured in fibrin matrices supplemented either with collagen I or laminin-1 was similar as in fibrin matrices supplemented with fibronectin (data not shown).

Taken together, preterm amnion epithelial and mesenchymal cell morphologies in 3D fibrin matrices were found to be similar as observed in vivo. Amnion cells are highly viable, and no significant decrease in the size of fibrin cell-matrix systems was observed.

Preterm amnion mesenchymal cell migration within fibrin matrices supplemented with fibronectin was associated with increased levels of activated MMP-9. This activity was reduced after incubation with 10 μmol/L GM6001. However, no activated MMP-9 was detected in cultures of preterm amnion mesenchymal cells within native fibrin matrices or fibrin matrices containing collagen I or laminin-1. The inactive proforms of MMP-9 and MMP-2 could be detected in all cultures (Figure 4).

In summary, fibronectin-filled fibrin matrices were associated with up-regulated MMP-9 activation. This seems to be specific for fibronectin as fibrin matrices supplemented with laminin-1 or collagen I did not up-regulate MMP-9 activation.

Comment

Our study was performed to investigate the behavior of human preterm amnion epithelial and mesenchymal cells in 3D collagen I and fibrin matrices, with the ultimate objective to develop a potential option for the treatment of PPROM patients based on tissue engineering. Tissue-engineered transplants can be composed of 2 components, tissue-specific cells placed into a biocompatible scaffold on/in which these cells can proliferate. The scaffold material provides initial mechanical support and a template for 3D organization.17 The approach pursued here relies on the design of a cell-containing matrix that mimics the architecture of native amnion. The selection of cellular components of 3D matrices needs to be tissue specific. In our previous study, we demonstrated that amnion epithelial and mesenchymal cells differ in their repair potential. Mesenchymal cells from preterm placenta healed more quickly than their term counterparts.11 Consequently, in this study, the cellular components of the here-developed cell-matrix transplants were amnion epithelial and mesenchymal cells from preterm placenta. Besides the selection of the cellular components that ultimately determine the...
function of such cell-matrix systems, an important feature of these systems is the possibility of in situ application. Moreover, concerning the aim of this study, it is important that such cell-matrix systems can be precisely administered via endoscopic methods to the ruptured site of the amnion membrane. Trying to meet this

Figure 3  Morphology of preterm amnion epithelial and mesenchymal cells in 3D fibrin matrices supplemented with fibronectin. Overall cell morphology at culture days 2 and 8 was visualized by staining for F-actin in combination with a DAPI nuclear stain. The reconstruction of the 3D system was performed using a stacked series of optical slices obtained by confocal laser scanning microscopy. (A) Amnion cell morphology visualized at culture day 2. Amnion mesenchymal cells were sparsely distributed without prominent extensions (M; arrow; star). The position of amnion epithelial cells is indicated (E; arrowhead). (B) Amnion cell morphology visualised at culture day 8. Amnion mesenchymal cells displayed a formation of typically migratory phenotype with cell extensions (M; arrow; star). The position of amnion epithelial cells is indicated (E; arrowhead). (C) Morphology of preterm amnion mesenchymal cells at higher magnification at culture day 8. Overall mesenchymal cell morphology was visualised by combining staining for F-actin (arrow) and nuclei (arrowhead). (D) Amnion mesenchymal cells were highly viable as determined by Dil-ac-LDL uptake, and the cell number determined by DAPI stain was not significantly changed over the time course of 8 days (p = 0.556). Data represent evaluation of three images per culture condition. (E) No significant decrease in the size of fibrin matrices was detected over the course 8 days (p = 0.222). Scale bars: A, 500 μm; B, 500 μm; C, 40 μm.
requirement, we have used 3D fibrin matrices as a scaffold matrix for isolated preterm amnion epithelial and mesenchymal cells. As outlined in this study, 2 interesting features of the 3D fibrin-matrix system were the refinement by the addition of extracellular matrix molecules that introduce additional adhesion sequences involved in cell adhesion and/or migration, and specific matrix degradation by cell-associated proteases such as MMPs or serine proteases, such as plasmin. We have demonstrated that preterm amnion epithelial and mesenchymal cells in modified 3D fibrin-based cell-matrix systems acquire their natural morphologies and displayed viability. In addition, in fibrin matrices supplemented with fibronectin, preterm amnion mesenchymal cells activated proteolytic programs of tissue repair, as specifically demonstrated by the activation of MMP-9. Moreover, fibronectin, here supplemented into native fibrin matrices, is an intrinsic component of clinical ‘fibrin glue’ formation. During clotting, plasma fibronectin incorporates into the fibrin network. Furthermore, the use of collagen I as a scaffold matrix was also investigated in this work. One limitation of this approach, as observed in this study, is contraction and, consequently, size reduction of collagen I matrices in vitro. In vivo, such a size reduction would inevitably induce failure of the implant followed by clinical consequences.

According to the literature, control of the contraction and organization of cells and matrix can be critical for successfully creating tissue engineered grafts.18 Unfortunately, these aspect cannot be easily predicated at the design of cell-containing collagen I matrices. In contrast to the collagen I matrix-cell systems, no decrease in the size of fibrin cell-matrix systems was observed over the time course of one week. Related to this data, one might speculate that such fibrin based cell-matrix systems may be useful in amnion cell transplantation. By virtue of fibrin’s biological characteristics, such an implant could seal the membrane leak instantly. Subsequently, the fibrin clot may be remodelled into native membrane tissue by amnion cells provided in the matrix as well as by surrounding amnion cells. These fibrin based cell-matrix systems should induce only a low immunological response. First, the fibrin matrix is physiologically and therapeutically relevant. It is widely applied in surgery as sealant and adhesive, and when formed from pure fibrinogen, represents a highly defined substrate.

Figure 4  Comparison of gelatinolytic activity of MMP-9 in cultures of preterm amnion mesenchymal cells cultured within 3D fibrin matrices. Fibrin matrices supplemented with fibronectin, collagen I or laminin-1 were assessed. Control cultures were prepared in native fibrin matrices. Increased levels of activated MMP-9 were detected in fibrin matrices containing fibronectin. Activity of MMP-9 was reduced after treatment with 10 μM GM6001. The positions of pro- and active forms of MMP-2 and MMP-9 are indicated as assessed by comparison with a MMP-2 and MMP-9 marker.
Second, human amnion epithelial cells do not express any of the leukocyte antigens (HLA)-A, -B, -C or –DR, suggesting amnion tissue is a promising cell source for a variety of tissue engineering applications including amnion cell transplantation. Recently, successful xenotransplantations have been reported with human amnion epithelial cells in the spinal cord of monkeys. Furthermore, different clinical studies confirmed that amnion material effectively facilitates epithelialization, can maintain normal cellular phenotypes, and reduce inflammation and scarring. Many studies are ongoing to investigate this unique feature of amniotic membranes as a prerequisite for tissue engineering.

Although the data presented here is promising, further studies must be undertaken to investigate the response of amnion cell-based tissue engineered grafts to more complex multifactorial in vivo environment.

Acknowledgment

We thank Michael Smith, PhD, from the Department of Materials, Swiss Federal Institute of Technology and Ajit Sankar Mallik, MD, from the Zurich University Hospital for their critical review of the manuscript.

References

EDUCATION

The American Association of Obstetricians and Gynecologists Foundation Scholars Program: Additional data on research-related outcomes

Georgine M. Pion, PhD, Charles B. Hammond, MD

Department of Psychology and Human Development, Vanderbilt University, Nashville, TN; Department of Obstetrics and Gynecology, Duke University, Durham, NC

Received for publication April 7, 2005; revised June 30, 2005; accepted July 13, 2005

KEY WORDS
Research Fellow AAOGF

Objective: The purpose of this study was to ascertain the progress of recipients of research training in obstetrics and gynecology in establishing an active research career in academic medicine.

Study design: Existing data were used to examine the extent to which 41 individuals who had received American Association of Obstetricians and Gynecologists Foundation (AAOGF) fellowships had achieved outcomes indicative of a career in academic medicine. Outcomes included employment as a full-time faculty member, receipt of NIH research funding, number of publications, the types of journals in which these articles had appeared, and the type of research (eg, basic vs patient-oriented).

Results: Among individuals who were awarded their fellowship between 1984 and 1997, 88% held faculty appointments, and 40% of these positions were in institutions that were more research-intensive that the medical degree–granting institutions of fellows. Slightly more than half of former fellows had successfully competed for NIH research funding, with 22% being awarded at least one R01 grant. Overall, fellows produced a total of 878 articles, one third of which appeared in clinical journals, 18% were in basic biomedical research journals, and 48% were in journals that published both types of research.

Conclusion: Previous AAOGF scholars have actively pursued research careers in academic obstetrics and gynecology. Their performance compares favorably with those of individuals receiving research training in other clinical specialties. A more complete understanding of their performance and the value added by the program would be possible if a core set of data on outcomes were available from other types of training efforts in both obstetrics and gynecology and other relevant disciplines.

© 2005 Mosby, Inc. All rights reserved.
careers in academic medicine. In terms of National Institutes of Health (NIH) initiatives, these include the Reproductive Scientist Development Program (RSDP), the Women’s Reproductive Health Recent Career Development Centers (WRHR), the Building Interdisciplinary Research in Women’s Health Program, and the Contraception and Infertility Research Loan Repayment Program (CIR-LRP). With regard to private support, one of the earliest initiatives aimed at developing academic leaders and strengthening research capability was the American Association ofObstetricians and Gynecologists Foundation (AAOGF) Scholars Program. Established in 1984, the annual number of fellowship awards has grown from 1 to 4, partly as a result of membership donations, major legacies of past members, and support from the Burroughs Wellcome Fund. As of 2004, a total of 57 fellowships had been awarded.

The early career outcomes of both AAOGF and RDSP fellows have been reported in 4 previous studies. Most frequently examined have been outcomes associated with the receipt of a faculty position and external research support—2 traditional indicators of a successful career in academic medicine. Although the cohorts studied, timing of data collection, and variables of interest differed, the results of these evaluations have been remarkably consistent. Between 90% and 95% of former fellows were in faculty positions, with between 58% and 80% also having attracted external research support at some point during their career.

This paper updates the career progress of former AAOGF scholars, focusing on those outcomes most indicative of an independent research career. In addition to faculty employment and receipt of research funding, attention is directed at examining fellows’ publication track records. Not only are data presented on traditional indicators used to measure research productivity (number of journal articles and the quality of these publications), but information also is presented as the research focus of this scholarship (ie, basic vs clinical research). This latter indicator was chosen, given the growing concern over the inadequate supply of physician investigators, particularly with regard to clinical and patient-oriented research.

Material and methods

In order to allow sufficient time for the outcomes of interest to occur, data collection focused on the 41 individuals whose first year of AAOGF funding began between 1984 and 1997. Rather than surveys of past fellows, primary reliance was placed on existing databases that contained information on all former fellows. Although this strategy limits the type of outcomes that can be examined, it imposes less respondent burden and is less vulnerable to any biases imposed by survey nonresponse. For background characteristics and academic employment, information was obtained from the AAOGF-maintained database on scholar recipients and supplemented by Web searches. The NIH Computer Retrieval of Information on Scientific Projects (CRISP) database was used for information on past and current NIH support. Publication data for each scholar were gathered from the Institute for Scientific Information’s (ISI) Web of Knowledge; for each individual, 2 searches were performed on the last name, 1 adding the first and middle initials (where applicable) and another adding only the initial of the first name. To avoid false positives, authorship of articles produced by these searches was verified by obtaining copies of each article and checking the name and institutional affiliation. Vanderbilt University’s Institutional Review Board approved the study.

To assist in interpreting the results, data from published outcome studies of other relevant research training programs are reported where available. These include: 1) a survey of 146 postdoctoral trainees who graduated from 1 of the National Research Service Award (NRSA) programs in primary care research between 1988 and 1997; 2) data compiled on 89 individuals who were supported between 1987 and 1996 by the Pediatric Scientist Development Program (PDSP), an NICHD-funded research training initiative in pediatrics similar to the RSDP in obstetrics and gynecology; and 3) 103 graduates of a fellowship program at Harvard Medical School aimed at training generalist researchers in internal medicine. Outcomes of individuals who received predoctoral training support from the Medical Scientist Training Program (MSTP) but who only earned the MD also are reported. Although these groups are far from ideal comparison groups for obvious reasons (eg, different aptitudes, training experiences, specialties, and research interests), they are useful in providing a context for evaluating the career outcomes of AAOGF scholars.

Because of the small sample sizes involved and resulting loss of statistical power, statistical tests of significance cannot serve as the sole criteria for identifying potentially meaningful subgroup differences. Thus, effect sizes were used to judge the magnitude of observed differences. The effect size is an index that is used to express, using standard deviation units, the difference between 2 groups on a particular measure. Accepted criteria exist for identifying the size of any group difference; an effect size between 0.20 and 0.49 is considered indicative of a “small” effect, an effect size of 0.5.0 to 0.79 is seen as a “moderate” effect, and one that equals or exceeds 0.80 is viewed as “large.”

Results

Background characteristics

Women accounted for approximately 49% of all 1984 to 1997 fellowship recipients; as might be expected from
trends in medical school enrollments over the past 20 years, their representation was somewhat greater among 1992 to 1997 scholars (52%) compared to the 1984 to 1991 cohort (45%). Fellows typically earned their MDs from research-intensive institutions. Approximately 20% graduated from US medical schools that were ranked in the top 10, with regard to total dollars awarded by the NIH to domestic institutions for research grants in 2000; this percentage climbed to 46% when the top 25 institutions were considered. Fifteen percent graduated from schools ranked between 26th and 50th in NIH research funds, and 27% graduated from lower-ranked medical schools. The remainder (12%) received their MD degree from a non-US medical school—a group that is excluded in the NIH rankings.

On average, scholars began their AAOGF-supported training about 7.5 years (SD 2.0) after completing medical school. This period has lengthened somewhat for more recent scholars. For individuals who began their postdoctoral study between 1984 and 1991, the average time elapsed between receipt of the MD and initiation of the AAOGF fellowship was 6.9 years; this figure increased to 7.5 for the 1992 to 1997 cohort (ES = 0.31). Women more often began their fellowships earlier than men (6.9 years vs 7.4 years, ES = 0.26). However, the inching upwards of the elapsed time between medical school and initiation of the Scholars Award occurred for both sexes. The average elapsed time for women was 6.6 years for the 1984 to 1991 cohort and 7.2 years for 1992 to 1997 cohort; the corresponding means for men were 7.1 and 7.8, respectively.

Faculty positions
Consistent with previous studies,1-4 nearly all (88%) of former fellows currently hold full-time faculty appointments in obstetrics and gynecology departments, with the remainder in clinical practice. This compares favorably to the available published data on graduates of research fellowship programs in other specialties (Figure). It is slightly below the 93% reported for graduates of the PSDP, but considerably higher than both the 64% reported for MSTP-supported trainees who only completed the MD (but not the PhD) and the 68% cited for former fellows of NRSA primary care research programs funded by the Health Resources and Services Administration.5,6,8 It also is comparable to the 83% reported for MSTP graduates who had earned both the MD and the PhD.8 Unlike former NRSA primary care fellows in which 55% remained affiliated with their fellowship training institutions, only one third of AAOGF scholars held faculty positions at the institution where they had completed their fellowship.5 However, 1992 to 1997 scholars were more likely than their 1984 to 1991 counterparts to remain at their postdoctoral training institutions (38% vs 25%, respectively).

NIH funding
More than half (56%) of 1984 to 1997 AAOGF fellows had been awarded 1 or more research grants by the NIH. This proportion is quite similar to that for MSTP trainees who earned only the MD (see Figure). It also is appreciably higher than that reported for primary care fellows; however, this difference is more difficult to interpret, given that the percentage reported for this latter group includes all types of federal research awards that were active at the time of the survey. Former AAOGF scholars have been less successful in securing NIH research awards than both pediatricians who graduated from the PSDP program and dual-degree holders who received MSTP support—a finding that is consistent with national figures.
on the overall performance of obstetrics and gynecology departments in obtaining NIH research dollars relative to their counterparts in pediatrics and the NIH success rates for MDs versus MD/PhDs.

Success in obtaining NIH support at some time in their post-fellowship careers was somewhat more likely for individuals in the early cohort. Whereas nearly two thirds (65%) had obtained NIH funding, this was true for 48% of the 1992 to 1997 cohort (ES = 0.34).

Former fellows in pediatrics did outperform their AAOGF counterparts in terms of having competed successfully for an R01 grant. Whereas 22% of AAOGF recipients had been awarded an R01 or R29 grant, this was true for 50% of PSDP fellows. Receipt of a career development award (a K award), however, was more similar (17% of AAOGF fellows vs 12 percent of PSDP fellows).

**Publication outcomes**

During their fellowship, the large majority (85%) of AAOGF scholars authored at least 1 publication. This was true of both cohorts, although those in the later cohort were slightly more productive. Whereas individuals who began their fellowship between 1984 and 1991 published, on average, 5.6 articles during their fellowship training, the corresponding figure was 6.5 for those in the 1992 to 1997 cohorts.

Similar to previous studies on gender differences for scientists, the publication productivity of men was noticeably higher than that for women scholars. Across all cohorts, women published an average of 3.2 articles during their AAOGF-supported fellowship training, whereas the average for men was 8.8 articles. These differential publishing records characterized both cohorts.

Despite some occasional fluctuations, a majority of scholars consistently had 1 or more papers published in any given year after their AAOGF-supported research training. For example, in the first year after their award had ended, 73% had published at least 1 paper on which they were the sole or coauthor. The corresponding percentages for years 3, 6, and 9 were 62%, 70%, and 76%, respectively. What this suggests is that the majority of scholars have continued to be active researchers throughout the first decade of their post-fellowship career.

**Quality of published research**

As shown in the Table, a noticeable proportion (33%) of scholars’ articles appeared in 1 of the top 10 journals in obstetrics and gynecology, as judged by the impact factor assigned to journals by the ISI (the frequency with which the “average article” in a journal has been cited in a particular year). If one focuses on the top 25 journals in biomedical and clinical research as judged by NIH intramural researchers, the percentage drops to 8%. The journals included in each of these 2 categories are listed in the Appendix.

It could be argued that the latter is a less appropriate measure for judging the quality of AAOGF fellows’ scholarship, given that this group of prestigious journals includes several that are unlikely outlets for communicating research in reproductive science and health. Thus, it is worth noting that about 45% of AAOGF fellows published at least 1 article in this set of highly regarded journals. This proportion is nearly identical to that reported for MDs who are Burroughs Wellcome Fund Career Awardees, a group of highly talented basic biomedical researchers who are awarded more support (up to 3 years of postdoctoral training and 3 years of faculty support).

**Types of journals in which publications have appeared**

One strategy to examine the type of research conducted by fellows is to consider the types of journals in which this scholarship was performed. Using categorizations assigned by the ISI, scholars’ articles were identified as appearing in: 1) a basic research journal if the journal primarily published research in the life or other natural sciences (eg, American Journal of Physiology, Endocrinology, and Molecular and Cell Biology); 2) a clinical journal if published in journals that primarily focused on clinical medicine (eg, Journal of Ultrasound in Medicine); and 3) a “mixed” journal that published both types of research (eg, JAMA and Obstetrics and Gynecology). This classification scheme had been used to describe the publications of former MSTP trainees and, thus, allowed a comparison between the performance of AAOGF scholars and MSTP trainees.

The most common publication channel for recipients of AAOGF training support consisted of journals that publish research in both the basic biomedical sciences and clinical areas (see Table). About 88% of fellows had published at least 1 article in these types of journals. Large majorities also had authored 1 or more articles in journals devoted to clinical medicine (73%) and journals devoted to communicating basic research (71%).

Looking at this in another way, AAOGF scholars published a total of 878 journal articles during the years following their fellowship. Nearly half (48%) of these articles appeared in journals responsible for publishing both basic and clinical research, 18% were in journals that focused on the basic biomedical sciences, and one third (33%) were aimed at readers of primarily clinical medicine journals. These percentages are fairly similar to those for scholars’ publications during their fellowship training; of the 249 articles authored by these individuals, 15% appeared in basic science journals, 29% in clinical medical journals, and 57% in articles that publish both types of research.
The distribution of publications across various journal types also was compared to that for the 1986 to 1990 cohort of MSTP graduates and MSTP MD only graduates—the group that most closely matched the AAOGF scholars in terms of career stage. In general, AAOGF fellows were more likely to publish their research in mixed and clinical medicine journals relative to both of these groups. Whereas 33% of scholars’ articles were in clinical publication outlets, this was true of only 4% and 8% of the articles authored by MSTP-funded MD/PhDs and MDs (ES = 0.82 and 0.65, respectively); these represent moderate to large differences. Similarly, 48% of AAOGF fellows’ work appeared in mixed journals as compared to 13% and 28% of the MSTP groups (ES = 0.79 and 0.42). Again, these are sizable differences.

Table Papers published by AAOGF fellows

<table>
<thead>
<tr>
<th>Outcome</th>
<th>n</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of publications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent of AAOGF fellows with one or more articles in journals that were among the:</td>
<td>34</td>
<td>83</td>
</tr>
<tr>
<td>Top 10 journals in obstetrics and gynecology</td>
<td>18</td>
<td>44</td>
</tr>
<tr>
<td>Highly prestigious biomedical research journals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent of articles by AAOGF fellows in journals that were among the:</td>
<td>320</td>
<td>33</td>
</tr>
<tr>
<td>Top 10 journals in obstetrics and gynecology</td>
<td>45</td>
<td>8</td>
</tr>
<tr>
<td>Highly prestigious biomedical research journals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Types of published research</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent of AAOGF fellows with one or more articles in journals devoted primarily to:</td>
<td>29</td>
<td>71</td>
</tr>
<tr>
<td>Basic research</td>
<td>30</td>
<td>73</td>
</tr>
<tr>
<td>Clinical medicine</td>
<td>Both types of research</td>
<td>36</td>
</tr>
<tr>
<td>Percent of articles by AAOGF fellows in journals devoted primarily to:</td>
<td>155</td>
<td>18</td>
</tr>
<tr>
<td>Basic research</td>
<td>296</td>
<td>33</td>
</tr>
<tr>
<td>Clinical medicine</td>
<td>Both types of research</td>
<td>426</td>
</tr>
<tr>
<td>Recent involvement in clinical research</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent of AAOGF fellows with one or more articles that represented:</td>
<td>17</td>
<td>42</td>
</tr>
<tr>
<td>Any type of clinical research</td>
<td>8</td>
<td>20</td>
</tr>
<tr>
<td>Patient-oriented research</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent of articles by AAOGF fellows that were:</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Patient-oriented research</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Mechanisms of human disease</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Clinical trials and interventions</td>
<td>Other types of clinical research</td>
<td></td>
</tr>
<tr>
<td>Development of new technologies</td>
<td>Epidemiologic research</td>
<td></td>
</tr>
<tr>
<td>Other types of clinical research</td>
<td>Health services and outcomes research</td>
<td></td>
</tr>
<tr>
<td>Epidemiologic research</td>
<td>Behavioral studies</td>
<td></td>
</tr>
<tr>
<td>Health services and outcomes research</td>
<td>Use of human tissue</td>
<td></td>
</tr>
</tbody>
</table>

* Percentages are based on a total of 41 fellows.

↓ Percentages are based on a total of 877 articles authored by fellows after completion of their fellowship.

↓ Percentages are based on a total of 77 articles that were published in 2002.

Involvement in clinical research

What the above analyses suggest is that the research of investigators trained with AAOGF funds has tended to lie closer to the clinical side of the basic-clinical research continuum—a pattern that is consistent with the arguments for the important role and contribution of MDs in conducting clinical research. However, the picture remains somewhat unclear, given that the “mixed” journals category is hard to interpret (ie, studies could be either basic or clinical research efforts). To explore this further, all articles published by AAOGF fellows in 2002, excluding reviews, letters, and editorial matter, were examined. Each of these 77 articles was classified with regard to their use of human subjects; studies that required IRB approval (and typically informed consent) were identified as human subjects research. For this cluster of studies, each was then categorized using a classification scheme developed by Kotchen et al in their study of NIH peer review for clinical research. A study could either be: 1) research that examined mechanisms of human disease; 2) clinical trials and other clinical intervention studies; 3) research focusing on development of new technologies; 4) epidemiologic research; 5) behavioral research; 6) health services and outcomes research.
research; 7) research that was based on using deidentified human tissue; or 8) other. These 8 categories are commonly regarded as clinical research. Categories 1 through 3 are typically used to describe the basic types of patient-oriented research—namely, research conducted with human subjects in which the researcher interacts with human subjects, as based on the definitions used by the NIH and Association of Patient-Oriented research.

Overall, 58% of fellows’ published articles were laboratory-based and did not involve human participants, 2 characteristics that are more common of basic research. The remaining 42% of published studies did involve human participants and fell into 1 of the 8 clinical research categories (see Table). Most common were studies that used deidentified human tissue (23%). The fractions that were epidemiologic, health services, or behavioral research were small (3-4%). Approximately 14% could be labeled as patient-oriented research, involving direct contact and interactions with patients as part of the study; 8% investigated at understanding the mechanisms of human disease; clinical trials and other intervention studies accounted for 5%; and 1% were focused on the development of new technologies.

Comment

Our examination of career outcomes corroborates the results of earlier evaluations of fellowship programs aimed at training physician-scientists in obstetrics and gynecology. The overwhelming majority (88%) have active careers in academic medicine. Slightly more than half have successfully competed for NIH funding at some point in their post-fellowship careers. Although proportion of fellows with current NIH funding is much lower (29%), it is most likely symptomatic of the difficulty of maintaining a continuous NIH-funded research program in an increasingly competitive environment. Furthermore, the pressures of clinical practice and the financial status of academic departments make it increasingly difficult for faculty to conduct significant research. Given the current environment, it is encouraging that 22% of fellows have been awarded at least 1 R01 grant—a percentage that is no different than that for new R01 investigators overall.

Approximately four fifths of all former fellows have had at least 1 publication in a leading obstetrics and gynecology journal, and nearly half have had their research published in 1 or more journals viewed as high quality by biomedical scientists. Compared to graduates of MSTP programs (even those who earned only the MD), AAOGF scholars’ research efforts have been more directed at clinical research questions. A snapshot of fellows’ published research in 2002 revealed that almost half (46%) of the articles matched the definition of clinical research developed by the NIH and others. Represented is the full range of clinical research, but the most frequent are studies that used human tissue (23%) and patient-oriented research (14%). This level of involvement in clinical research not only supports the distinct role of physician-scientists in reproductive health but also identifies the role that postdoctoral training programs, even ones that are modest in size, can have on increasing the pool of clinical investigators.

We believe the AAOGF scholars program has added to the research capability of academic departments of obstetrics and gynecology. However, in the face of changing times and options in funding, we believe that it is important to continue assessment of the program on a regular and ongoing basis. Further, conducting such assessments for other training initiatives in obstetrics and gynecology and in related clinical disciplines is worth considering for several reasons. Perhaps the most important concerns the added knowledge about what works in training physician-scientists that would be gained from having appropriate comparison groups and employing similar measures. For example, although little benefit would be obtained from comparing a core set of outcomes for MDs who have completed research fellows with those of MDs with no fellowship training, much could be learned from comparing the same outcomes for programs that provide different training experiences, levels of support, and so forth. We encourage the American Gynecological and Obstetrical Society, its Foundation (AAOGF), and other sponsors of research training in reproductive health to collaborate in the design and execution of evaluation studies that can further identify the effectiveness of such scientific training and how it applies to the long-term research goals of the discipline.

References

Appendix

Top 10 journals in obstetrics and gynecology

The 10 journals that were classified as obstetrics and gynecology and that had the highest impact factors as calculated by the Institute for Scientific Information were: Human Reproduction, Human Reproduction Update, Fertility and Sterility, Placenta, American Journal of Obstetrics and Gynecology, British Journal of Obstetrics and Gynecology, Menopause, Journal of the Society of Gynecological Investigation, Obstetrics and Gynecology, and Gynecologic Oncology.

Top 25 journals in biomedical research

CASE REPORTS

Ischiorectal abscess after sacrospinous ligament suspension

Michael Hibner, MD,a Jeffrey L. Cornella, MD,a,* Javier F. Magrina, MD,a Jacques P. Heppell, MDb

Department of Obstetrics and Gynecologya the Division of General Surgery,b Mayo Clinic, Scottsdale, AZ

Received for publication December 3, 2004; revised April 20, 2005; accepted May 12, 2005

KEY WORDS
Abscess
Bladder diseases
Prolapse
Uterine
Rectocele
Sacrospinous suspension
Streptococcus viridans

An ischiorectal abscess in a 66-year-old patient was determined to be an uncommon complication of sacrospinous fixation. The abscess was diagnosed 9 months after the patient had a sacrospinous ligament suspension. She was treated successfully with perianal incision, drainage, and intravenous antibiotics.

© 2005 Mosby, Inc. All rights reserved.

The sacrospinous ligament suspension is a common surgical procedure for treatment of vaginal wall prolapse. Known complications of this procedure include hemorrhage, buttock pain, nerve injury, rectal injury, vaginal stenosis, and recurrent anterior vaginal wall prolapse.1-3

Case report

A 66-year-old woman presented to our institution with a complaint of bloody vaginal discharge and pelvic pain. Nine months earlier, the patient had been diagnosed with a grade 3 pelvic organ prolapse and severe stress urinary incontinence.

The patient had a Burch colposuspension with anterior and posterior colporrhaphy and a bilateral sacrospinous ligament suspension. External operative note review showed that the sacrospinous ligament suspension had used 2 sutures of 0-polyester on each side. Sutures were placed using a suture-capturing device (Capio; Boston Scientific, Natick, MA). Further review of the operative report revealed that patient received 1 g of Cefotetan IV prior to the first incision. The vagina had been prepped with Betadine.

Four months after the operation, the patient noticed a watery, pink vaginal discharge and had pain with defecation. Three weeks before her presentation to our institution, the patient experienced substantial worsening of pain. She also noted a gush of malodorous, purulent material from the vagina. At the same time, she had a noticeable decrease in appetite, increased constipation, general malaise, chills, and a fever of 100.5°F.

On physical examination, the patient was found to have an extensive fullness in the left buttock. Skin over
the indurated area was red and tender. On vaginal examination, the mass was noted to extend to the apex of the vagina. A nonabsorbable suture was found in the apex on the left side. Some drainage of pus was evident beneath the suture. A rectal examination showed that the mass extended approximately 7 cm above the anal verge on the left side.

Computed tomography with intravenous contrast medium showed a large abscess on the left side of the rectum (Figure 1). The abscess seemed to have 3 loculations, with the lowest measuring 2.9 × 2.6 cm and abutting the lateral rectal wall. Superficial to the levator muscle was a larger collection measuring 7.1 × 5.1 cm. The most caudad collection measured 3.7 × 2 cm.

The patient was taken immediately to the operating room, where a rigid proctoscopy was performed under general anesthesia. The rectal mucosa was found to be normal with no inflammation or fistula detected.

A cruciate perianal incision was made, and a large amount of foul-smelling purulent material was drained. All the septa in the abscess cavity were disrupted with curved 15-cm Mayo-Stille scissors, and the remainder of the abscess was drained. The cavity was then irrigated with water and packed lightly with half-inch gauze. The sacrospinous suture was removed through a vaginal approach. The patient was given intravenous ciprofloxacin and metronidazole and was kept in the hospital 2 days longer for pain control. The culture from the cavity of the abscess grew Streptococcus viridans.

While the patient was in the hospital, her white blood cell count decreased from 20.9 × 10⁹/L preoperatively to 17.5 × 10⁹/L postoperatively. The patient was instructed to irrigate the abscess cavity with sterile water twice a day, using a 60-mL syringe and a short catheter. Two weeks later, the abscess cavity had completely healed.

Comment

In 1951 Amreich introduced the procedure by which the vagina was suspended to the sacral tuberosity. In 1968 Richter modified Amreich’s techniques, using the sacrospinous ligament rather than the sacral tuberosity as the point of fixation. This procedure was popularized in the United States by Randall and Nichols, who reported a 97% success rate with unilateral suspension.

Several early postoperative and long-term complications are reported. Cuff cellulitis, cystitis, and postoperative voiding difficulties are common. Other early complications include hemorrhage, nerve entrapment, and laceration of the rectum and bladder. Long-term complications include vaginal stenosis, stress urinary incontinence, shortened vagina, recurrent prolapse, and massive evisceration.

This ischiorectal abscess is the first reported after a sacrospinous ligament suspension. A Medline search was accomplished using the key words ischiorectal and abscess and sacrospinous. The time covered was from 1966 to the fourth week of January 2005. There were no articles describing formation of ischiorectal abscess after sacrospinous ligament suspension published in the journals covered by the Medline database.

Because S. viridans is usually not present in the rectum, the abscess may have been caused by contamination
from the skin of the buttocks or the mucosa of the vagina rather than from a perforation of the rectum. Streptococci are rarely resistant to Cefazolin, and most likely the cause of infection is inadequate vaginal prep or break in sterile technique.

All surgical sutures can be divided into monofilament and braided sutures. Braided sutures have capillary characteristics that allow absorption of water together with potential pathogens. Those pathogens can easily adhere to the large surface area of the multifilament suture. Because of all those characteristics of braided sutures, monofilament sutures are much more suitable in the areas in which the risk of infection is high.

To better visualize anatomical correlations of the abscess, vagina, and sacrospinous ligament, a 3-dimensional reconstruction was made (Figure 2). It clearly showed the abscess to be located between the left vaginal apex and the left sacrospinous ligament. This finding confirmed our suspicion that the suture between those 2 structures was the cause of the abscess.

References
Application of the three-dimensional maximum mode in prenatal diagnosis of Apert syndrome

Tilman Esser, MD, a Patrick Rogalla, MD, b Christian Bamberg, MD, a Karim D. Kalache, MD a

Departments of Obstetrics and Gynecology a and Radiology, b Charité University Hospital, Campus Mitte, Berlin, Germany

Received for publication February 11, 2005; revised June 6, 2005; accepted July 7, 2005

KEY WORDS
Apert Syndrome
Three-dimensional ultrasonography
Prenatal diagnosis

Case report

A 29-year-old woman, gravida 1 para 0, with an unremarkable medical history was referred for an ultrasound examination in our department at 22 + 0 weeks’ gestation because the extremities were not clearly seen at the referring institution. On 2-dimensional (2D) ultrasonography the second to fourth digit of both hands did not seem to move independently from each other, but the examination of the extremities was limited because of fetal position. The facial profile was normal, as was the shape of the head. There were no other obvious anomalies. The diagnosis of Apert syndrome was suspected because of the anomaly of the fingers, but could not be made with certainty at this stage of the diagnostic workup. A 3-dimensional (3D) ultrasound examination (Voluson 730-Expert, GE Healthcare, Milwaukee, Wis) of the fetal face was performed. The 3D data set was first manipulated in the multiplanar mode and then in the maximum mode (also called skeletal mode). Multiplanar imaging confirmed the lack of obvious abnormalities of the facial profile, whereas the maximum mode revealed a widely open metopic suture (Figure 1) with bilateral fusion of the coronal sutures (Figure 2). This information led us to repeat the examination of the hands, and syndactyly was confirmed. The patient was counseled and decided to terminate her pregnancy. Karyotyping was performed on the amniotic fluid as well as screening for the most common mutations of Apert syndrome. The diagnosis was confirmed by the identification of a heterozygote mutation Pro253Arg (C758G). Postmortem computer tomography confirmed the prenatal findings (Figures 1 , C, and 2, C).

Comment

The birth prevalence of Apert syndrome is 1:65,000. Characteristic features include coronal craniosynostosis, brachycephaly, midfacial hypoplasia, and symmetrical syndactyly of the hands and feet. Approximately one half...
of affected individuals are mentally retarded. An autoso-
mal dominant inheritance pattern has been observed,
but the majority of cases are sporadic as a result of a de
novo mutation. The known mutations that cause the
syndrome are 2 recurrent missense mutations of the
fibroblast growth factor receptor 2 gene (FGFR2) invol-
v ing 2 adjacent amino acids: S252W and P253R. It is
interesting to note that the mutation in our case
(Pro253arg) is associated with a better postsurgical out-
come for craniofacial appearance. This could explain
that the shape of the head was normal on 2D ultrasound.

Several cases of prenatal diagnosis of Apert syn-
drome that were based on the detection of abnormalities
of the head or extremities using 2D ultrasonography
have been reported. However, there is only 1 report of
the use of 3D ultrasonography for diagnosis of Apert
syndrome. In this report, 3D ultrasonography was not
crucial for the diagnosis and only refined it by reveal-
ning midfacial hypoplasia and downsloping palpebral
fissures.

The ultrasonographic abnormalities for this syn-
drome are easy to interpret in cases with a positive
family history. However, prenatal diagnosis of sporadic cases can be very challenging. In fact, the vast majority of sporadic cases have not been definitively diagnosed until the third trimester when the abnormalities of the skull shape related to craniosynostosis became obvious. As seen in our case, earlier in pregnancy the shape of the head can be normal and syndactyly, which is almost always present, may be missed.

Craniosynostosis is characterized by a premature fusion of 1 or more cranial sutures. In Apert syndrome, there is a widened metopic suture with the sagittal suture extending from the glabella to the posterior fontanelle caused by coronal craniosynostosis. Demonstration of the premature fusion of the coronal suture helps in differentiating Apert syndrome from other forms of craniosynostosis. Acrocephaly in Crouzon syndrome and Carpenter syndrome results from synostosis of coronal, sagittal, and lambdoid sutures, whereas in Pfeiffer syndrome, there is synostosis of the sagittal and coronal sutures. 3D maximum mode eliminates soft tissue echoes and prominently emphasizes high echogenic structures such as bones. We have shown that this mode is able to reveal direct signs of pathognomonic changes that occur in Apert syndrome that were not recognized by 2D ultrasound.

Acknowledgment

We acknowledge Dr Roberto Romero, Chief of Intramural Research, Perinatology Research Branch of the National Institutes of Health and Human Development, Bethesda, MD, for his valuable comments and suggestions.

References

Placement of a temporary vena cava filter during labor

Steven L. Clark, MD, Duane D. Blatter, MD, G. Marc Jackson, MD

St Marks Hospital, Hospital Corporation of America, LDS Hospital, and Department of Obstetrics and Gynecology, University of Utah School of Medicine, Salt Lake City, UT

Received for publication February 14, 2005; revised April 6, 2005; accepted April 16, 2005

KEY WORDS
Thromboembolism
Pregnancy
Vena cava filter

Placement of a vena cava filter for the prevention of pulmonary thromboembolism in select patients is a well established procedure in critical care medicine. We describe a case of placement and removal of a new removable vena cava filter in a pregnant patient, in this case during early labor. Vaginal delivery was accomplished without incident.

© 2005 Mosby, Inc. All rights reserved.

Case Report

A 21-year-old primigravid woman was admitted to the hospital at 38 weeks’ gestation with a diagnosis of extensive unilateral ileofemoral deep venous thrombosis. There was no clinical evidence of pulmonary embolism. She received a 5-day course of full anticoagulation with intravenous unfractionated heparin. Thereafter, the patient was discharged on a low-molecular-weight heparin, 1 mg per kg subcutaneously every 12 hours. Three days later the patient was readmitted to the hospital in early labor, with a cervix 3 cm dilated.

We felt it desirable to withhold full anticoagulation during labor; however, given the relatively recent nature of the deep vein thrombosis, we did not feel that sufficient time had passed to ensure full clot organization and adherence to the vessel wall. Because the clot was also extensive, we felt the patient was at risk of massive pulmonary embolization, either during the process of labor itself, or after sudden evacuation of the uterus at the time of delivery. With these concerns in mind, the patient was sent in the latent phase of labor to the interventional radiology suite for placement of a removable Tulip vena cava filter (Cook Incorporated, Bloomington, IN) (Figure).

The right internal jugular vein was cananlized and a 5 French pigtail catheter advanced under ultrasound guidance into the caudal inferior vena cava. Initial inferior venacavogram obtained in the anterior-posterior (AP) projection demonstrated complete effacement of the infrarenal inferior vena cava by the gravid uterus. Deployment of the filter into the vena cava under this degree of compression would have resulted in an unacceptable configuration with 2 attachment hooks engaged adjacent to each other in the posterior wall of the vena cava. Such configuration would result in limited effectiveness in trapping potential thromboemboli. The patient was placed in the left lateral decubitus position and repeat inferior venacavogram confirmed decompression of the vena cava; the filter was successfully deployed with a proper configuration in an infrarenal location. She labored in the left decubitus position, and underwent a spontaneous vaginal delivery of a viable male fetus who did well.

Anticoagulation was withheld for 12 hours and then reinstituted. Forty-eight hours after delivery, the inferior vena cava filter was retrieved, once again using a right
jugular vein approach after residual clot behind the filter had been ruled out by repeat venacavagram. She received postpartum anticoagulation with coumadin.

Comment

Classic indications for placement of a vena cava filter include pulmonary embolism despite adequate anticoagulation and a contraindication to anticoagulation in the presence of proximal deep vein thrombosis. While vena cava filters have occasionally been used in obstetrics for many years, clinicians have been reluctant to employ this device during pregnancy since the long-term sequela of a permanent vena cava filter inserted into a woman in her early 20s is not well studied; most permanent filters have historically been placed in older individuals. However, with the development of the newer temporary Tulip vena cava filter, such devices may be inserted for short periods of time and then removed when the risk of embolization is less.1 This makes them ideal for temporary use during circumstances in which full anticoagulation is relatively contraindicated, such as labor.

The unique features of this device include a hook which allows transvascular dislodgement, collapse, and retrieval. Retrieval is successful in over 95% of cases in which it is attempted. Most current studies suggest the sheath may be left in place for up to 14 days before removal; when clinical circumstances dictate the need for longer filter use, the device can be moved to a slightly different location within the vena cava, thus extending its use for an additional 14 days. The Tulip filter may also be left in place as a permanent filter, apparently without ill effect.

This is the first reported case of the placement and removal of a temporary vena cava filter in a pregnant woman in the obstetric or American literature. A previous case report from the UK described a similar woman who underwent delivery by cesarean section.2 Our report suggests that such women may undergo uneventful vaginal delivery shortly after filter placement without dislodgment or other adverse effects of the birthing process. In our patient, this was greatly facilitated by lateral decubitus positioning of the patient—this will probably be of value to other clinicians encountering the need for filter placement during late-term pregnancy.

References

Maternal death caused by midgut volvulus after bariatric surgery

Paul V. Loar, III, MD,a,* Luis Sanchez-Ramos, MD,a Andrew M. Kaunitz, MD,a Andrew J. Kerwin, MD,b Jesus Diaz, MDc

Departments of Obstetrics and Gynecology,a Surgery,b and Medicine,c The University of Florida Health Science Center Jacksonville, FL

Received for publication January 23, 2005; revised March 31, 2005


© 2005 Mosby, Inc. All rights reserved.

More than 100,000 gastric bypass surgeries were performed in the United States in 2003 alone, with up to 73% involving women. As a result, pregnancy after gastric bypass is becoming much more common.

Case report

A 31-year-old G4P2012 woman with a singleton pregnancy at 25 6/7 weeks and a history of laparoscopic Roux-en-Y gastric bypass was transferred to the University of Florida Health Science Center at Shands Jacksonville because of nonspecific abdominal pain and possible small bowel obstruction.

On arrival, her examination was nonacute and she was maintained on bowel rest with nasogastric decompression. A computed tomography (CT) found distension of the entire small bowel with air fluid levels and a distended colon from cecum to sigmoid. Ultrasound revealed a 784 g fetus in cephalic presentation and normal amniotic fluid. After several days with some clinical improvement, she began to complain of dark, black emesis and copious melanotic diarrhea. Tests for Clostridium difficile (C. diff) toxin were positive and treatment was begun. At the recommendation of the consultant general surgeon, flexible sigmoidoscopy was performed to exclude colon obstruction. There was no evidence of obstruction and the colon appeared normal. Symptom resolution after this procedure was attributed to therapeutic decompression. She continued to improve but later at 26 5/7 weeks’ gestation experienced preterm premature rupture of membranes with a prolapsed umbilical cord. An emergent cesarean delivery was performed via a Pfannenstiel incision, during which healthy loops of small bowel were visualized with no evidence of intra-abdominal infection. The infant’s Apgar scores were 1 at 1 minute, 6 at 5 minutes, and 8 at transfer to the neonatal intensive care unit (NICU).

On postoperative day 3 the patient was noted to be tachycardic, tachypneic, and hypotensive. She became increasingly unstable and was intubated after transfer to...
the medical ICU (MICU). Subsequently, septic shock, multiorgan failure, and abdominal compartment syndrome were diagnosed. This latter entity refers to massive intestinal edema associated with increased intra-abdominal pressures, which impair respiratory and renal function. During exploratory laparotomy, performed by the general surgeon at bedside in the MICU, copious amounts of feculent fluid were noted within the abdominal cavity. She had frank intestinal necrosis of the entire small bowel and right colon suggesting a vascular accident in the distribution of the superior mesenteric artery. The small bowel was twisted around the root of the mesentery creating a midgut volvulus. The timing of the intestinal ischemia and necrosis is enigmatic because the patient was noted to have normal healthy bowel at the time of her cesarean delivery. A perforation was noted in the distal small bowel with no focal point of obstruction. After consultation with her family, the patient was taken off of life support.

Autopsy findings included small bowel volvulus with associated ischemic necrosis, dilatation, and perforation. Multiple adhesions were seen in the root of the mesentery causing midgut volvulus. The cause of these adhesions is speculative. Certainly, the patient’s previous surgery is plausible. The effect of the pregnancies and metabolic changes associated with her gastric bypass are unknown.

The infant, reported as depressed and floppy at delivery, initially received blow-by, bag and mask ventilation, followed by endotracheal tube ventilation before transfer to the NICU. Initially diagnosed with infant respiratory distress syndrome and later with sepsis, hypotension, electrolyte abnormalities, seizures, hydrocephalus, anemia from chronic phlebotomy, and a grade IV intracranial hemorrhage, he received 2 doses of surfactant, fluid and pressor support, intravenous electrolyte repletion, anticonvulsants and antibiotics, serial lumbar punctures, and multiple blood transfusions. The infant was intubated shortly after delivery, followed by extubation to continuous pulmonary airway pressure and then to nasal cannula on postdelivery day 2. The above therapeutic interventions were continued at Shands until he was transferred to a nearby children’s hospital on day 33 with plans to follow-up with physical and occupational therapy as well as pediatric neurosurgery for shunt placement.

Comment

Publications that address pregnancy after gastric bypass surgery largely focus on postoperative malabsorptive disorders and describe decreases in obesity-related pregnancy complications. Limited information exists, however, with respect to adverse events after this procedure. In an autopsy study of 10 deaths after Roux-en-Y gastric bypass, half were due to technical complications directly related to the procedure. In addition, pulmonary embolism, both clinical and silent, was noted in 80% of these cases. Other complications include gastrointestinal hemorrhage, sepsis, abscess, obstruction, and sudden unexplained death. The International Bariatric Surgery Registry reports a mortality rate of 0.3% in the first 6 months. We found only 1 publication that reported maternal complications. This was a case report strikingly similar to ours, which described bowel infarction from an internal hernia through the mesentery after Roux-en-Y gastric bypass. The diagnosis was not made until exploratory laparotomy revealed 61 cm of ischemic small bowel, which was resected. Despite this, both mother and fetus died.

Our case, along with the previously reported case, underscores the need for increased vigilance on the part of both general surgeons and obstetricians in evaluating pregnant women with abdominal pain after bariatric surgical procedures for morbid obesity. General surgeons should have a high index of suspicion in these patients and a lower threshold for performing exploratory laparotomy to avoid missing intra-abdominal disease. Research assessing benefits and risks of bariatric surgery in reproductive age women, including in those who later conceive, is needed. Further study comparing outcomes among morbidly obese reproductive age women undergoing or not having bariatric surgery would be particularly useful.

References

Vulvar cellular angiofibroma: A case report

Ryan Kerkuta, MD*, Colleen M. Kennedy, MD, MS, Jo A. Benda, MD, Rudolph P. Galask, MD, MS

Departments of Obstetrics and Gynecology, and Pathology, Roy J. and Lucille A. Carver College of Medicine, University of Iowa Hospitals and Clinics, Iowa City, IA

Received for publication April 29, 2005; revised June 28, 2005; accepted August 8, 2005

KEY WORDS
Vulva
Cellular angiofibroma

Cellular angiofibroma is a benign growth initially described in 1997, with few reports to date. A 31-year-old woman presented with a 3-year history of a small left labial mass, which had recently increased in size to 5 cm, and was clinically thought to be a lipoma. A simple excision was performed. Histologically, the mass was consistent with a cellular angiofibroma. Ten months later, the growth has not recurred. Cellular angiofibroma is a rare, benign mesenchymal lesion typically occurring on the vulva, and should be considered in the differential diagnosis of a painless, soft, vulvar mass.

© 2005 Mosby, Inc. All rights reserved.

Cellular angiofibroma of the vulva is a rare benign growth of mesenchymal origin that was initially described by Nucci in 1997.1 There is little information about follow-up after excision. Previously described as a lesion of middle-aged women, we describe a case of a vulvar cellular angiofibroma that occurred in a 31-year-old woman.

Case report

A 31-year-old parous, white woman was referred for further management of an enlarging, nontender vulvar mass. The patient had first noticed a “walnut-sized” left labial growth about 3 years earlier and ultimately sought medical attention from her local physician when it enlarged to “lemon sized.” The physician referred the patient to a tertiary care center after attempted drainage of the mass was unsuccessful.

The patient had genital warts excised 15 years earlier and had no other history of sexually transmitted diseases. She used low-dose oral contraceptive pills and denied tobacco, alcohol, or illicit drug use. She had no personal or family history of breast or gynecologic cancer. Regular Papanicolaou test smears were all normal.

Examination revealed a 5-cm, soft, mobile, nontender mass involving the left labia majora (Figure 1). Clinically, the mass resembled a lipoma. Small bowel was not palpable in the mass. Groin lymph nodes were not enlarged on palpation. The remainder of the pelvic examination was otherwise unremarkable.

She underwent an uncomplicated simple resection of the mass in the operating room. Similar to lipoma mass...
excisions, the pseudocapsule was easily enucleated after the initial skin incision. The patient’s postoperative course was uncomplicated. At the time of follow-up 10 months later, there was no sign of recurrence.

Pathology

On gross examination, the mass was 5 × 4 × 3 cm, uniform, and fibrous with no fluid accumulation or cystic spaces. Microscopically, spindle- and oval-shaped stromal cells with bland nuclei and minimal mitotic activity were present. These were interspersed with thin bands of collagen fibers and rare adipocytes. The mass contained numerous small vessels and a focal perivascular lymphoid infiltrate. Mast cells were present (Figure 2). Immunohistochemical staining was positive for vimentin, estrogen receptor, and progesterone receptor, and negative for CD34, smooth muscle specific actin, desmin, muscle specific actin, and S-100. These findings are consistent with the diagnosis of cellular angiofibroma.

Comment

Mesenchymal tumors of the vulva include leiomyoma, hemangioma, and lipoma. Spindle cell tumors include neurofibroma, schwannoma, and smooth muscle sarcoma. More unusually encountered masses include angiomysolid, angiomysolid, and cellular angiofibroma. The distinction of cellular angiofibroma from these lesions is morphologic, although immunohistochemical studies are typically performed to confirm the diagnosis. Although cellular angiofibromas are benign mesenchymal tumors, exclusion of other vulvovaginal soft tissue tumors, including aggressive angiomyxoma and sarcoma, is essential to avoid unnecessary aggressive treatment. Clinically, an inguinal hernia should also be considered in the differential diagnosis of a vulvar mass, as the approach to a hernia repair differs from excising a simple mass.

Cellular angiofibroma has been identified in middle-aged women with a mean age of 48 (range 37-77) years. Clinically, cellular angiofibroma is often mistaken for a Bartholin gland, labial, or submucosal cyst. Although local excision with clear margins is the treatment of choice, there is little information about the long-term follow-up of this tumor. No cases with metastasis are reported in the literature. However, there is 1 report of tumor recurrence, in which a 49-year-old woman underwent a simple excision of a well-circumscribed 4-cm lesion and had recurrent swelling develop at the site of the previous excision 6 months later. A 6.5-cm well-circumscribed mass was excised without complication, with no further evidence of recurrence noted at 10 months. Both lesions were consistent with cellular angiofibroma.

Although similar lesions have been reported outside the vulva, cellular angiofibroma should not be mistaken for a nasopharyngeal angiofibroma, a polypoid intranasal growth found typically in adolescent boys. This lesion is life threatening and is characterized by a parallel arrangement of the collagen fibers.

Characteristic features of cellular angiofibroma are small lesions (typically less than 3.0 cm) with well-circumscribed margins, although extension into surrounding tissue was described in 1 case in which a

Figure 1  Vulvar mass.

Figure 2  Bland ovoid to spindled nuclei and absence of mitotic figures is present in this high-power view (original magnification: ×60).
46-year-old woman underwent initial enucleation, followed by 2 re-excisions and had no sign of recurrence at 19 months. The cells are spindle shaped, lying between bands of collagen. Hyalinized vessels, abundant mast cells, and scant adipocytes may be present. Few mitotic figures are identified. Cellular pleomorphism is rare, and tumor necrosis has not been described. Immunohistochemically, cellular angiofibromas are positive for vimentin and may express reactivity for CD34, whereas desmin, actin, S-100 protein, keratin, and epithelial membrane antigen are negative. In addition, the tumor has been found to be estrogen and progesterone receptor positive. However, the significance of the positive estrogen and progesterone receptors in angiofibroma is unknown. These receptors are normal in subepithelial mesenchymal cells of the lower female genital tract, so this may be a reflection of the cell of origin rather than an alteration in a neoplasm. Alternatively, it has been postulated that the expression of estrogen and progesterone receptors suggests a role in the pathogenesis of this tumor.

The patient presented in this report was younger and her mass larger than most others reported in the literature. Thus, cellular angiofibroma should be considered in the differential diagnosis of a soft, painless, vulvar mass even in younger women.

Acknowledgment

We thank Ingrid E. Nygaard, MD, MS, for her editorial assistance with the manuscript.

References

A tale of 2 pedunculated myomas

Ihab M. Usta, MD, Elie M. Hobeika, MD, Anwar H. Nassar, MD

Department of Obstetrics and Gynecology, American University of Beirut Medical Center, Beirut, Lebanon

Received for publication July 13, 2005; revised July 21, 2005; accepted August 11, 2005

KEY WORDS
Submucous myoma
Pedunculated

We present 2 unusual cases of prolapsed pedunculated submucous myomas. In 1 patient, the prolapsed part measured 12 cm, with a 64-cm intrauterine part. The second patient had prolapsed pedunculated submucous myoma, which subsequently retracted into the uterus. Gynecologists should be aware of unusual presentations of pedunculated submucous myoma to plan surgery.

Uterine leiomyomas are the most common benign tumors of the female genital tract; 5% are submucosal, and 1.3% to 2.5% are pedunculated.1,2 Of the pedunculated submucous myomas (PSMs), 19.2% to 26.1% measure >3 to 5 cm.2 PSMs rarely may induce labor-like pain, eventually dilating the cervix and prolapsing through it with possible avulsion of the twisted necrotic prolapsed portion.2

We present 2 cases of PSM; 1 case had an unusual appearance, and another case had unusual behavior.

Case 1

A 37-year-old woman (G3P3) had chronic foul-smelling discharge, acute bleeding, and labor-like pain that was followed by the protrusion of a mass on the perineum. A 12-cm fleshy, dark-red mass was seen originating from the uterus and prolapsing through a dilated cervix (Figure 1A). Ultrasound examination revealed a 17-cm uterus with a large myoma occupying the uterine cavity, which was continuous with the prolapsed part. Abdominal hysterectomy was performed. The myoma consisted of a 6-cm intramural round fundal fibroid tumor, with an elongated soft mid-portion coiled within the uterine cavity and a distal portion that was prolapsed partially through the cervix. The vaginal part was delivered vaginally; the other parts were removed abdominally. The length of the reconstituted myoma was 76 cm (Figure 1B).

Case 2

A 46-year-old grandmultiparous woman had a 4-month history of menorrhagia. Ultrasound examination revealed an irregular uterus that measured 10.6 × 6.8 × 7.8 cm, with a posterior 6.0 × 4.1 cm fibroid tumor. One week later, she experienced severe bleeding and labor-like abdominal cramps. Speculum examination showed a 5-cm pedunculated violaceous fibroid tumor that protruded through a 5-cm dilated cervix. The bleeding decreased, and the pain subsided 1 hour after admission. In the operating room, 7 hours later, a speculum examination by the same examiner revealed a closed cervix and retraction of the prolapsed myoma into the uterus. Intraoperatively, the myoma consisted of a whitish fundal lobe that was attached to the uterus >3.5-cm base and a violaceous lobe, which was the same part that had previously prolapsed through the cervix.

Reprints not available from the authors.

0002-9378/S - see front matter © 2005 Mosby, Inc. All rights reserved.
doi:10.1016/j.ajog.2005.08.026
Various imaging modalities can be used to diagnose and characterize prolapsed PSMs and to obtain information that might aid in the choice of the surgical approach. In our cases, only ultrasound examination was used, because both patients had acute vaginal bleeding.

Vaginal myomectomy is an option for treating small PSMs. Twisting, ligation and excision, and morcellation of larger tumors have been used. More recently, hysteroscopically assisted techniques have been used. However, the size of the myoma in the first case was prohibitive for the attempt at such techniques. The longest myoma ever reported measured 13 cm, which makes our case one of the largest and probably the longest myoma reported in the literature.

The second case represents the first report of the regression of a myoma into the uterine cavity after prolapse. Vaginal expulsion was not possible because no such expulsion was reported by either the nursing staff or the patient, and the pathologic specimen was the exact size and configuration of the image that was seen on ultrasound examination (Figure 2A and B). Manual replacement of prolapsed myomas was reported previously in 3 patients before hysterectomy to prevent the contamination of the abdominal cavity. In our case, replacement during vaginal scrubbing was possible; however, the cervix was closed, which suggested that the myoma had retracted earlier.

These cases illustrate the potential challenges gynecologists might face during the treatment of PSMs. Underestimation of the size of the myoma in the first case and attempts at vaginal myomectomy in similar situations might have led to disastrous results. Furthermore, a disappearing prolapsed myoma, such as in the second case, does not necessarily imply spontaneous expulsion; retraction of the prolapsed part should be kept in mind.

Figure 1  Case 1. A, The prolapsed part of the myoma. B, The myoma on a 90-cm table. The thin arrow shows the intramural part; the thick arrow shows the prolapsed part.

Figure 2  Case 2. A, The myoma 1 week before prolapse. B, Pathology specimen. The thin arrows show the base of the myoma; the thick arrows show the part that prolapsed and retracted.
References


Temporary balloon occlusion of the common iliac artery: New approach to bleeding control during cesarean hysterectomy for placenta percreta

Jin-Chung Shih, MD,\textsuperscript{a} Kao-Lang Liu, MD,\textsuperscript{b} Ming-Kwang Shyu, MD\textsuperscript{a}

Department of Obstetrics and Gynecology,\textsuperscript{a} and Department of Medical Imaging,\textsuperscript{b} National Taiwan University Hospital and National Taiwan University College of Medicine, Taipei, Taiwan

Received for publication June 1, 2005; revised July 25, 2005; accepted August 12, 2005

KEY WORDS
Placenta percreta
Balloon catheters
Common iliac artery

A case of placenta percreta was referred at 31 weeks’ gestation. We performed a cesarean hysterectomy preceded by placement of occlusive balloon catheters at bilateral common iliac arteries at 34 weeks’ gestation. This simple and safe technique provides satisfactory efficacy for control of profuse bleeding during operation, with blood loss estimated at 800 mL.

© 2005 Mosby, Inc. All rights reserved.

Placenta accreta is characterized by placental villi abnormally adherent to or invading the myometrium. Of the variants, the most severe form is placenta percreta, which indicates placenta penetrating through the entire uterine wall and even into the adjacent organs. Life-threatening hemorrhage often occurs in cases of placenta percreta during cesarean section, even with correct diagnosis and advanced preparation presurgery. The deaths were partly caused by inability to arrive at the correct diagnosis before surgery and, more importantly, catastrophic bleeding from abundant neovascularization and rich collaterals beyond the efficacy of hemostasis available using current surgical techniques.

We report our successful experience in a case of placenta percreta, where preoperative placement of occlusive balloon catheters in the bilateral common iliac arteries minimized blood loss during cesarean hysterectomy.

Case report

A 34-year-old woman, gravida 4, para 1, presented at our clinic at 31 weeks’ gestation because of placenta previa. She had given birth to 1 healthy term infant by cesarean section caused by prolonged labor. Other past history included 2 uterine curettages for miscarriages. Our initial ultrasonography revealed a male fetus weighing approximately 2300 g and presenting in the transverse position. Complete placenta previa with loss of the hypoechoic retroplacental zone led us to the diagnosis of placenta accreta. With weekly ultrasonographic investigation, progressive bulging of the placenta into the urinary bladder was observed and placenta percreta with bladder invasion was strongly suspected. Given the risk of spontaneous uterus rupture and an estimated fetal weight of 2500 g, an elective cesarean hysterectomy was proposed at 34 weeks’ gestation.

A multidisciplinary approach was scheduled at admission. Because of the inherent anastomoses between the internal and external iliac arteries, the radiologist

Reprints not available from the authors.
elected to use temporary occlusive balloons in the bilateral common iliac arteries (CIA) during our cesarean hysterectomy. The patient and her family were appropriately counseled with respect to the surgical and procedural risks incurred. On the day of the operation, both femoral arteries were punctured using the standard Seldinger technique, with placement of 8-F sheath introducers. The 5-F balloon catheters (maximal occlusive diameter 25 mm, Goodtec Co, Gifu, Japan) were then inserted and placed in the CIA. The bilateral CIA locations below the aortic bifurcation were easily identified using the guide wire without contrast-enhanced angiography. The fluoroscopy time and the dosimetry of fetal exposure were only 2.4 minutes and 1.8 mGy, respectively. Under endotracheal general anesthesia, cesarean section was commenced after the establishment of the arterial line and central venous route. After exploring the peritoneal cavity, tortuous cobweb-like vessels were identified covering almost the whole anterior uterine wall. The bulging placenta beneath the uterine serosa was readily identified. During application of the abdominal wall retractor, spontaneous uterine rupture and massive bleeding ensued at the implantation site. A uterine incision was immediately made over the higher uterine corpus to deliver the baby, with simultaneous inflation of bilateral CIA balloon catheters. The neonate weighed 2630 g and had Apgar scores of 7/9 at 1/5 minutes, respectively. Blood loss was dramatically decreased after balloon occlusion of the bilateral CIA. Subtotal hysterectomy was rapidly performed without removal of the placenta. Despite early initiation of the cesarean hysterectomy at this gestation, however, it was found that the adhesive placenta had invaded to the urinary bladder, resulting in unavoidable bladder injury. The bladder repair was performed smoothly by urologists after hysterectomy. The bilateral balloons were deflated before approximating the peritoneal cavity to ensure adequate hemostasis. An oximetry monitor was placed on the patient’s lower limb and the saturation maintained at around 85% to 92% during CIA occlusion. The arterial blood gas of lower limb was checked during this procedure, and the result yielded mild metabolic acidosis with respiratory compensation (pH 7.37, pCO2 30.8 mm Hg, pO2 105.9 mm Hg, HCO3- 17.4 mEq/L, BE – 6.6 mEq/L). The temporary CIA occlusion sustained 53 minutes, with blood loss estimated at 800 mL. The patient was transfused with 500 mL of packed red blood cells (RBC) only. The hemoglobin levels before and after the operation were 10.2 and 9.7 g/dL, respectively. The markers of reperfusion injury were checked on the following day and all demonstrated no elevation (CK 99 U/L, CK–MB 2.5 U/L, LDH 490 U/L). Histologic examination of the uterus revealed that the chorionic villi had penetrated the entire myometrium without intervening decidua, confirming the diagnosis of placenta percreta.

The operation was well tolerated by the patient, although recovery was complicated by temporary hematuria and ileus. Eventual discharge was uneventful 8 days after operation, with the Foley catheter removed 10 days later at the urology clinic. At 6 months after the operation, the mother and baby remained healthy.

Comment

Currently, the management options for placenta accreta include conservative and extirpative approaches. The conservative strategy entails leaving the placenta in situ during cesarean section without further hysterectomy. However, risk of sepsis and delayed hemorrhage is also incurred. The extirpative approach consists of immediate cesarean hysterectomy, avoiding placental removal during operation. Nonetheless, extirpative management is associated with significant risk of catastrophic bleeding. An adjuvant hemostatic technique is usually needed to minimize the risk of shock, multiple transfusion, and even maternal death. Preoperative embolization of the internal iliac artery (IIA) provides an excellent hemostatic effect in these cases; however, it is only suitable in cesarean hysterectomy with previable fetuses. Sometimes this procedure may elicit unexpected complications such as pelvic pain, sexual dysfunction, or bladder ischemia. The alternative feasible technique is bilateral IIA ligation during cesarean hysterectomy. Using this procedure, the pressures in the pelvic circulation are converted from arterial pressures to venous-like pressures, thus making branches in the circulation more amenable to hemostasis via simple clot formation. However, the failure rate for this may be as high as 60%.

A widely adopted alternative to IIA ligation is preoperative placement of occlusive balloon catheters in the bilateral IIA. This procedure appears to be superior because the obstetrician does not need to be familiar with the technique of IIA ligation, and can immediately occlude the bilateral IIA after cord clamping without the interference created by a small and bloody surgical field. However, the results of a number of trials appear to indicate that the technique is not as effective as expected, with 1 study even concluding no difference in outcomes comparing patients who underwent this procedure and those who did not.1 Failure of IIA ligation or balloon occlusion can be largely explained by the extensive anastomoses in the pelvic vasculature. The obturator artery serves as an important anastomosis bridging internal and external iliac arteries.2 Additionally, angiographic studies of patients undergoing IIA ligation have identified branches of the lumbar, sacral, rectal, femoral, and even internal thoracic arteries as the origin of collateral circulation, preventing pelvic ischemia. Collectively, these collateral pelvic-vessel networks may substantially decrease the efficacy of IIA ligation or occlusion.
To the best of our knowledge, the presented case is the first report using temporary CIA occlusion for management of placenta percreta during cesarean hysterectomy. Importantly, this approach appears to have some apparent advantages compared with existing techniques. Unlike placement of IIA balloon catheters, the CIA balloon catheter is easier to lodge in the proper location, especially for those cases with unusual vascular anatomy. The fluoroscopy exposure is decreased, and iodine-containing contrast medium is not required. Further, the procedure can be used to overcome collateral flow from the external iliac and femoral arteries, thus significantly decreasing hemorrhage during cesarean hysterectomy. However, prolonged occlusion of the CIA may be associated with reperfusion injury, thrombosis, and/or embolism of the lower extremities. Therefore, the occlusion time should be as short as possible. As the duration of aortic cross-clamping can be at least 1 hour in surgery for abdominal aortic aneurysm, our procedural time is relatively safe, especially with the presence of rich collaterals during pregnancy, and the blood flow cannot be totally occluded by balloon catheter. Further evaluation and comparison of the efficacy and outcome of this procedure is needed relative to the alternative techniques for management of advanced forms of placenta accreta.

References
Episiotomy dehiscence that required intestinal diversion

Carl H. Rose, MD,a Kristi L. Blessitt, MD,a Farshid Araghizadeh, MD,b John C. Morrison, MDa

Departments of Obstetrics and Gynecologya and Colon & Rectal Surgery,b University of Mississippi Medical Center, Jackson, MS

Received for publication May 12, 2005; revised August 4, 2005; accepted August 17, 2005

KEY WORDS
Episiotomy
Dehiscence
Infection
Ileostomy

Postpartum episiotomy dehiscence is a rare complication of vaginal delivery. Forceps-assisted vaginal delivery over mediolateral episiotomy was complicated by infection and dehiscence with rectal injury. A diverting ileostomy was used to permit healing. Episiotomy infection requires early recognition and thorough evaluation to exclude occult rectal injury.

A 40-year-old woman (G6P2032) was seen after premature rupture of membranes at 38 weeks of gestation. Her medical history was significant for obesity (body mass index, 29 kg/m²), chronic hypertension, and recurrent Grave’s disease that required radioactive iodine ablation before conception. Her first pregnancy culminated in a spontaneous vaginal delivery (2466 g) followed by a cesarean delivery for dystocia in her second pregnancy (2722 g). She was counseled and strongly desired vaginal birth after cesarean delivery for the current pregnancy. Oxytocin augmentation was begun, and labor progressed satisfactorily over the next 10 hours. However, fetal bradycardia ensued late in the second stage of labor; at +3 station, emergent outlet Simpson forceps-assisted delivery over second-degree mediolateral episiotomy was performed for a healthy 2265-g male infant with Apgar scores at 5 and 7 minutes of 9 and 9, respectively. On further inspection, the episiotomy was noted to extend into the rectal sphincter. Edges of the sphincter muscle were identified and reapproximated with the use of 4 interrupted 0 polyglactin sutures; another continuous 1-0 suture was used to rebuild the perineal body. The vaginal mucosa was closed with a smaller 2-0 suture that proceeded from the episiotomy apex to the hymenal ring. The bulbocavernosus stitches were pulled together, and the mediolateral skin incision was repaired with running 2-0 sutures. The patient’s postpartum course was complicated by recurrent hemorrhage because of uterine atony 3 and 5 hours after the delivery, which responded to 15-methyl prostaglandin given intramuscularly and misoprostol placed in the rectum. The episiotomy repair was noted to be intact during both episodes.

The next day she had a low-grade fever (99.8°F) and leukocytosis (white blood cell count 26,000/mm³); broad-spectrum antibiotics were begun. On the second day after the delivery, she complained of pain at the episiotomy site. Examination under anesthesia revealed partial dehiscence of the episiotomy repair with purulent drainage from an abscess cavity that communicated with the left ischiorectal fossa. A full-thickness injury that involved the anal sphincter complex was noted that extended inward 6 cm in depth and 1 to 2 cm in width from the anal verge into the rectum at the 6 o’clock position. The rectum mucosa was intact. The previous repair was completely opened to drain the abscess, and the devitalized tissue was debrided. The perineum and

Reprints not available from the authors.

0002-9378/S - see front matter © 2005 Mosby, Inc. All rights reserved.
doi:10.1016/j.ajog.2005.08.037
vaginal vault were irrigated, and the wound was packed with wet-to-dry dressings. A rectal tube was inserted, and anaerobic antibiotic coverage was added. Sequential postoperative blood glucose values were normal, and antinuclear antibody testing was negative. She remained afebrile throughout the remainder of her postoperative course.

Operative debridement of the perineal wound was repeated on postdelivery days 4 and 5 to ensure healthy, viable perineal tissues. Subsequently the colon and rectal surgical service was consulted for evaluation and treatment; on postpartum day 5, the clinical decision was made to undertake repair of the injury. Because the tissues continued to require debridement despite intensive surgical and antibiotic therapy, the decision was also made to divert the fecal stream simultaneously. Therefore, on postdelivery day 6, the injury was repaired primarily in 3 layers to restore annual rectal sphincter complex integrity. The subcutaneous wound above the repair, which extended into the ischiorectal fossa, was left open to heal by secondary intention. A diverting laparoscopic loop ileostomy was also performed to promote perineal wound healing. Her subsequent postpartum course was unremarkable, and she was discharged home on postdelivery day 8 with colon and rectal surgery follow-up visits. The perineal wound healed without event, as noted in the postpartum follow-up visits at 2 and 6 weeks. Five months later, she underwent ileostomy closure and had return of normal gastrointestinal function and fecal continence.

Comment

Infection and episiotomy dehiscence that involves the rectum is a catastrophic event. Permanent sphincter dysfunction with consequent fecal incontinence carries enormous hygienic and psychosocial consequences. Fortunately, dehiscence occurs in only 0.2% to 2% of all of episiotomies. Specific guidance for the treatment of concurrent proximal rectal injury is virtually nonexistent in the obstetric literature. Traditional doctrine has held that the failure of initial repair mandates a 2- to 3-month interval to permit thorough tissue revascularization. A preliminary report by Hauth et al\(^1\) described successful revision within 7 days of dehiscence in 8 patients, with only a single failure. Ramin et al\(^2\) have published the most extensive series to date regarding early revision. Early closure was successful in all except 2 of 35 cases (6%); both failures had pre-existing risk factors for poor tissue healing. All injuries were confined to the rectal sphincter and/or the lower 5 cm of rectum.\(^2\) The repairs were accomplished at a mean of 6.4 days (range, 3-13 days).

In most cases, diverting loop ileostomy or colostomy is not performed, and most uncomplicated wounds heal well, as noted by the published reports. Failure usually occurs when pre-existing risk factors such as diabetes mellitus, hypertension, obesity, or inflammatory diseases are present. In this case, because the patient was obese and because the abscess appeared early in the clinical course, the clinicians on both services believed that the likelihood of conservative treatment and delayed repair had a high risk of failure. Because of the extension of the abscess into the ischiorectal fossa, a diverting ileostomy was performed to deflect the fecal stream at the same time as the closure. Although immediate postpartum antibiotic therapy potentially could have changed the outcome in this case (absent factors other than obesity and emergent forceps delivery), there was no clinical indication to begin such treatment. Ileostomy was selected because it can be performed laparoscopically, is less prone to fistulae than colostomy, and requires much less patient care than traditional colostomy.

Infection appears to be the fundamental cause in most repair failures. After dehiscence, the incidence of failed secondary repair does not appear to differ significantly between midline (69%) versus mediolateral (86%) incisions. Other reported risk factors include chronic corticosteroid use, refractory cough, inflammatory bowel conditions, immunologic deficiencies, obesity, or collagen-vascular disease. Recommended laboratory evaluation includes complete blood count, antinuclear antibody, human immunodeficiency virus, and hepatitis profile. Depending on individual history, further evaluation may include an upper gastrointestinal series, barium enema, computed tomography scan, magnetic resonance imaging, or colonoscopy.

To date, this is the only reported case of episiotomy dehiscence that required temporary intestinal diversion that we were able to find in the literature. A diverting loop ileostomy technique was selected because the effluent quality and stomal care are reduced significantly compared with conventional colostomy.

References

Endometriotic umbilical port site metastasis after laparoscopy

Umut Barbaros, MD,a,* Ahmet Cem Iyıbozkurt, MD,b Mıne Gulluoglu, MD,c Merve Barbaros, MD,d Yesım Erbil,a Vahit Tunali,e Selcuk Mercan a

Departments of General Surgery, a Gynecology and Obstetrics, b and Pathology, c Istanbul Medical School, Istanbul University, Istanbul, Turkey; Department of Gynecology and Obstetrics, Private Hospital of Göksoy, Istanbul, Turkey d; Istanbul Training and Research Hospital, Samatya, Istanbul, Turkey e

Received for publication November 17, 2004; revised May 12, 2005; accepted May 12, 2005

KEY WORDS
Endometriosis
Laparoscopy
Endometriotic implant
Umbilical port site metastasis

Endometriosis, the presence of functioning endometrial tissue in anatomic locations other than the uterine cavity, still perplexes practitioners by its rare presenting forms. Endometriosis has a long history. It was described as adenomyomata of the rectovaginal septum in early reports.1,2 This was similar to uterine adenomyomata of the uterus.3 Although extrapelvic endometriosis may be present in up to 12% of women with endometriosis,3 umbilical endometriosis is uncommon, with an estimated incidence of 0.5% to 1.0% of all patients with endometrial ectopia.4

The majority of cutaneous endometriosis develops on surgical scars of the abdomen of young women in the reproductive period.5 Recently with more frequent use of a less invasive technique like laparoscopy for surgical treatment of endometriosis, more cases of endometriosis at port sites are encountered. From 1976 to August 2004, 6 cases of endometrioma of port sites after laparoscopy were published: The first report occurred after a sterilization procedure, published in 1990.6 Two cases were after laparoscopic-assisted subtotal hysterectomy,7 2 were after laparoscopy for endometriosis5,8 and 1 occurred after bilateral hernia treated by laparoscopy that was converted to laparotomy.9 In this report, we present a patient with endometriotic umbilical port site metastasis who previously had laparoscopic resection of ovarian endometriotic cysts.

Case report

A 40-year-old woman presented with a recent history of irregular umbilical mass and an associated severe cyclic pain during her menses. The mass was dark colored, hard, and measured 2 to 3 cm in diameter and was located on the previous laparoscopic umbilical port site scar (Figure 1). Ultrasonography of the lesion revealed a hypoechoic 29 × 21 × 26 mm lobulated lesion located...
at the subcutaneous tissue just below the umbilicus with intense arteriolar network image seen during Doppler ultrasound. The patient’s history was significant for laparoscopic removal of a multilocular endometriotic cyst from the left ovary. During this laparoscopic procedure, the cyst wall had been removed without using a bag through an inguinal port. The operative site had been irrigated using around a liter of ringer’s lactate solution. The rest of the patient’s history was unremarkable.

During the operation, the hard mass located just beneath the umbilicus above the rectus sheath was excised completely along with the adhered fascia (Figure 2). The umbilicus appeared normal. The defect was primarily closed. The pathology report confirmed that the mass excised was an endometrioma (Figure 3). She had an uneventful postoperative course.

Comment

Although umbilical endometriosis is uncommonly reported, it is usually easily diagnosed because of the cyclic character of the pain it elucidates, as in our subject. Simple excision of the umbilical endometrioma is the treatment of choice.10

The histogenesis of endometriosis is still unknown, although there has been a lot of research as to its origin. In the case of an umbilical port site endometrioma, the lesion may have developed from peritoneal seeding of cells because of pneumoperitoneum or from direct contact of the excised lesion with the abdominal wall. It is also proposed that cutaneous endometriosis may arise from endometrial tissue that is transported via lymphatics or vascular channels.11 Because it is known in the previous laparoscopy that the endometriotic lesion was removed through the sheath of the instruments, our case lends support to the aerosolization theory, the concept that pneumoperitoneum influences the implantation of free intraperitoneal endometriotic cells.
Routinely applying techniques like introduction and removal of instruments and the excised lesions within the sheath to avoid contact with the abdominal wall, using a bag during removal of endometriotic lesions and exsufflating the abdomen while the ports are in place may help to prevent endometriotic implants. Carefully planned case-control studies may help to yield more information about the etiopathogenesis of umbilical endometriosis.

References

On the Sites of the Negative and Positive Feedback Actions of Estradiol in the Control of Gonadotropin Secretion in the Rhesus Monkey

Y. Nakai, T. M. Plant, D. L. Hess, E. J. Keogh, Ernst Knobil

Endocrinology, 1978;102:1008-1014

FIG. 3. Restoration of LH and FSH secretion by a chronic intermittent iv infusion of LHRH (1μg/min for 6 min every hour) initiated on day 0 in an ovariectomized rhesus monkey in which gonadotropin secretion had been abolished by a hypothalamic lesion (see text).

FIG. 4. Cessation of tonic gonadotropin secretion in an ovariectomized rhesus monkey after placement of bilateral RF lesions in the medial basal hypothalamus on day 0. Gonadotropin secretion was restored by a chronic intermittent iv infusion of synthetic LHRH initiated on day 68 (1μg/min for 6 min every hour). EB injected sc on day 79, when 25% of the prelesion gonadotropin levels had been restored, resulted in an abrupt decline of LH and FSH levels followed by a discharge of these hormones. EB given before LHRH administration (day 13) and an estradiol-containing Silastic capsule inserted during a post-LHRH control period (day 106) were ineffective in eliciting gonadotropin discharges. The vertical lines beneath the data points indicated values below the level of RIA sensitivity.
Commentary by Lawrence D. Longo, MD

In the mid 1940s, Geoffrey W. Harris\(^1\) (1913-1971) posited the idea that hormones, as yet unidentified, regulated the function of the pituitary gland. Harris’ group later produced ovulation in rabbits by intrapituitary infusion of an extract of bovine hypothalamus.\(^2\)

Following considerable effort, Roger Guillemin and Andrew V. Schally independently demonstrated that hypothalamic peptides stimulated the release of follicular stimulating hormone (FSH) and luteinizing hormone (LH) both in vitro and in vivo, and in 1971, each reported the amino acid sequence of the decapeptide that became known as gonadotropin releasing hormone (GnRH).\(^3,4\)

Subsequently, in a series of studies in *Macaca mulata*, Ernst Knobil (1926-2000) and colleagues demonstrated that GnRH was released in a pulsatile manner and, carried to the adenohypophysis via a portal vascular network, regulated the release of FSH, LH and, thus, ovarian function.\(^5,6\) Following bilateral radiofrequency (RF) lesions placed in the arcuate region of the medial basal hypothalamus of an adult ovariectomized Rhesus monkey, a continuous infusion of exogenous LHRRH failed to restore gonadotropin levels. In contrast, and as shown in Figure 3, pulsatile LHRH infusion (1 \(\mu\)g/min for 6 minutes every hour) resulted in sustained restitution of LH and FSH secretion. Figure 4 illustrates the decrease in blood plasma gonadotropin levels following the RF lesions, and the failure of estradiol benzoate (EB) administration (42 \(\mu\)g/kg body weight) to restore these. Again, gonadotropin secretion was restored in response to chronic intermittent LHRH infusion.

At the University of Pittsburgh, Knobil established the Center for Research in Human Reproduction, where over several decades he made many seminal contributions to understanding the function of the hypothalamus, the pituitary gland, gonads, and the biology of reproduction. These contributions to an understanding of ovulatory function have resulted in a significant impact of health care. Knobil was elected to the National Academy of Sciences, and the American Academy of Arts and Sciences, and received numerous international awards and honors.

References


Discovery of the hypothalamic gonadotropin-releasing hormone pulse generator and of its physiologic significance*

Ernst Knobil, MD

Shortly after a specific radioimmune assay for monkey luteinizing hormone (LH) became available in the late 1960s, we began to investigate the dynamics of its secretion in a variety of physiologic and experimental circumstances. One such early study was designed to determine whether a circadian rhythm in the plasma

SELECTED MANUSCRIPTS FROM THE 2005 CREOG & APGO ANNUAL MEETING

The Council on Resident Education in Obstetrics and Gynecology (CREOG)
The Association of Professors of Gynecology and Obstetrics (APGO)

2005 CREOG & APGO ANNUAL MEETING
March 2-5, 2005
The Grand America Hotel
555 South Main Street
Salt Lake City, UT 84111

APGO Web site: www.apgo.org

Patrick Duff, MD, CREOG Chair
Department of Ob-Gyn
University of Florida
P.O. Box 100294
1600 SW Archer Road
Gainesville, FL 32610
Phone: (352) 392-3222
Fax: (352) 392-2808

William N. P. Herbert, MD, CREOG Vice Chair
Department of Ob-Gyn
University of Virginia
Box 800712
Hospital Drive, 3rd Floor Private Clinics
Charlottesville, VA 22908
Phone: (804) 924-9937
Fax: (804) 982-1840

Haywood L. Brown, MD, CREOG Program Co-chair
2005 CREOG & APGO Annual Meeting
Department of Ob-Gyn
Duke University
DUMC 3084
203 Baker House
Durham, NC 27710
Phone: (919) 668-3948
Fax: (919) 668-5547

Dee E. Fenner, MD, APGO President
Department of Ob-Gyn
University of Michigan
1500 E. Medical Center Drive
L4100 Women’s Hospital
Box 0276
Ann Arbor, MI 48109
Phone: (734) 763-6670
Fax: (734) 647-9727

J. Mac Ernest, MD, APGO President-elect
Department of Ob-Gyn
Wake Forest University
Medical Center Boulevard
Winston-Salem, NC 27157
Phone: (336) 716-1025
Fax: (336) 716-6937

J. Mac Ernest, MD, APGO Program Co-chair
2005 CREOG & APGO Annual Meeting
Department of Ob-Gyn
Wake Forest University
Medical Center Boulevard
Winston-Salem, NC 27157
Phone: (336) 716-1025
Fax: (336) 716-6937
“Every time I look back, I meet the eyes of my foreparents looking steadfastly forward.”

Professor Roger Shinn

Concerns abound with regard to the future of our specialty of obstetrics and gynecology. In his presidential address to the American College of Obstetricians and Gynecologists, Dr John Gibbons summarized the issues that dominate discussion: (1) recruitment of medical students; (2) change in residency programs; and (3) the new face of private or academic practice. The first is highlighted by the decrease in the percentage of graduating US seniors choosing a career in obstetrics and gynecology and the associated decrease in residency slots filled by US graduates. The marked decrease in the number of men entering the specialty is noted with concern, and the shift of student interest to emergency medicine, radiology, ophthalmology, anesthesia, and dermatology noted with envy. It must be said that a decade ago both anesthesia and radiology sounded their own alarms, so the astute observer must suppose that there are powerful cyclic forces operating.

The concern about residency training focuses on the ever broadening scope of requirements, something that as a member of the Residency Review Committee I am well aware of, but the breadth of medical knowledge presents a challenge to all undergraduate and graduate medical education programs. The 80-hour work week has surely placed new challenges on our residency programs. As to practice, increased debt burden, administrative burdens, liability concerns, and family and personal interests all contribute. Dr Gibbons speaks enthusiastically and hopefully about the potential to “refocus our specialty.”

One only need cast a brief glance at history to realize that similar concerns are not new for our specialty. Two of our most insightful and visionary specialty leaders, both my predecessors as chair at the University of Michigan, were important observers of this phenomenon.

In a 1920 address to the American Medical Association, Dr Reuben Peterson, chair from 1901 to 1931, outlined “The Future of Obstetrics and Gynecology as a Specialty.” This was a time when obstetrics and gynecology were considered widely by many as separate specialties, and “gynecologists threw themselves into the development of pelvic surgery to such an extent that things outside the operative field failed to interest them. … The result is that the ground work absolutely essential to one aspiring to devote himself to one division of obstetrics and gynecology is lost sight of and there is a wild scramble for a short cut to fame and fortune, usually through the dexterous handling of the scalpel.” He then argued for the unification of obstetrics and gynecology and for an emphasis on broad education as the basis for both specialists and subspecialists.

Dr Peterson saw merit in a suggestion that only medical school gynecologists and psychiatrists should select those most suited for gynecologic surgical careers, for “if by any chance (they) were to run across an
applicant who had ambition to become an expert in everything, physical and functional, pertaining to the genital tract of women, it would be comparatively easy to have such a person become either a temporary or a permanent occupant of the psychiatric clinic.”

He envisioned strong hospital-based training programs for both medical students and postgraduates, not yet widely called “residents.” His description of the ideal department of obstetric and gynecology to offer training for “young graduates” fits very well with the academic obstetrics and gynecology residency program that we would all recognize today. Peterson wrote that “each man would work out the hardships of being up all night with a confinement case and then be obliged to do difficult hysterectomies the next day” and hoped the trainee would “keep faith and find enough in his specialty to interest him so he will do his regular work well and add a little something to the sum of human knowledge.”

In another major address, published in the American Journal of Obstetrics and Gynecology as “Obstetrics-Gynecology: A Time for Change,” Dr. J. Robert Willson, my chairman who held his position from 1962 to 1979, raised not dissimilar issues: “changes in attitude and education of medical students; a proliferation of scientific and technical information to be incorporated into clinical practice; increasingly burdensome controls by burgeoning institutional, professional, and governmental bureaucracies; changing patient attitudes; changing physician attitudes.” Dr. Willson was an early acceptor and proponent of primary and preventive care and felt that “too little emphasis is placed on preparing house officers for contemporary practice.” He was one of the first to understand that we must be responsive to society and our patients as we develop the curricular content of our educational programs. Dr. Willson’s detailed critique of residency training remains insightful and provocative and resonates in the current discussions at the Accreditation Council for Graduate Medical Education (ACGME) and the Residency Review Committee (RRC).

As to medical students, he comments “that we now have as many or more applicants than first-year positions is considered by some to be an indication of success … but are we being responsible? … How many obstetrician-gynecologists trained in the present model will be content to spend almost all their time counseling adolescents, the elderly, and others and providing contraceptive and other routine ambulatory services and how well will they do these things?”

Dr. Willson then advocated changes and proposed eliminating ineffective programs, approving only programs that provide comprehensive training, competency-based evaluations of trainees, regular program reviews, appropriate specialty examinations, and recertification. He saw the American Board of Obstetrics and Gynecology, American College of Obstetricians and Gynecologists, Association of Professors of Gynecology and Obstetrics, the Council on Resident Education in Obstetrics and Gynecology, and the Residency Review Committee all as important participants in this change, and it is remarkable that virtually all of his recommendations have been successfully implemented, if not as quickly as he would have liked or in time for him to see.

So if similar concerns have existed for more than a century and previous leaders have predicted, advocated, and directed change, I believe we must follow their lead, face the challenges, and find the solutions that will chart the course for the optimal wellness and health of our patients and our specialty.

I believe, as did Dr. Willson, that change must come from our leaders and leadership organizations. I have demonstrated that as a specialty we have faced similar—remarkably similar—challenges before and have responded constructively.

Reuben Peterson and J. Robert Willson were right that our specialty will flourish if defined broadly: obstetrics and gynecology, surgery and primary care, specialty and subspecialty. We must address the concerns of those of who advocate for attention to surgical training, but breadth and balance must be maintained. The role of primary and preventive care was emphasized by Dr. Willson, and its importance was recently reaffirmed by specialty leaders in an “academic blueprint.”

Dr. Peterson also raised the issue of time, specifically night call. We need to heed the lesson of the 80-hour work week. After the ACGME mandate, the University of Michigan Medical School implemented an 80-hour work week for medical students. It did not take long, nor should it have, for the faculty to ask, “How about us?” Dr. Gibbons is correct that we must address the challenges not by focusing on residents and residency education or on medical students and medical education, but we must focus on practitioners and their practice: the very specialty of obstetrics and gynecology. Student interest groups will not change the reality of practice. Resident professionalized in an 80-hour work week will bring new expectations to their professional lives. We must refocus, redefine, if not reinvent our specialty. Already, as Dr. Gibbons notes, there is movement toward larger practice groups, groups that can absorb part-time doctors and offer reasonable call schedules.

Other specialties have developed hospitalists, full-time hospital-based clinicians. Weinstein has proposed the role of “laborist,” someone whose focus is management of laboring patients. One of the risks of some of these innovations is loss of continuity, but ultimately we must assure quality care and patient safety while maximizing satisfaction for all participants in the health care system. Recently our Maternal Fetal Medicine faculty have experimented with a night float system, taking the lead from residents and medical students for whom night float has become a national standard. In our
department, despite occasional logistic challenges, third- and fourth-year residents and medical students have well developed and enhanced night float systems. More junior residents are experimenting with options on how best to adapt it to their educational needs and work hour constraints.

We need to develop programs for reentry and retraining because well-prepared colleagues who have taken time out for family or life issues can be invaluable when they are ready to return to full or limited specialty practice. Family medicine has recently reported a comprehensive prescription for the successful future of that specialty. I believe strongly in a model in which the academic training programs, both university and non-university based, serve as change agents. These departments can be laboratories to test laborist, generalist, night float, part-time, reentry, and other options. And those models that are most successful will and must serve as exemplars for others to adapt in their own local, historical, and cultural context. These departments and their leaders must take the risk and create the innovation to be laboratories and exemplars of what obstetrics and gynecology as a specialty can be, should be, and will be.

There are several recognizable explanations for some of the current challenges, some already clearly described Dr Gibbons. To these, I want to highlight a few:

### Generational issues

I do not believe we have paid adequate attention to generational issues, which are becoming more relevant and important than they have ever been before. Our business school colleagues have come to accept these fundamental differences among my generation, the baby boomers, and generation X. And it looks like generation Y than generation X has been for the boomers. We need to understand each other’s values and work together better as we all move forward.

<table>
<thead>
<tr>
<th>Table I</th>
<th>Generations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional generation, 1925-1940</td>
<td>Traditional obligation to conform</td>
</tr>
<tr>
<td>Baby boomers, 1945-1963</td>
<td>Radicals</td>
</tr>
<tr>
<td></td>
<td>Individuals</td>
</tr>
<tr>
<td></td>
<td>Dual careers</td>
</tr>
<tr>
<td></td>
<td>Self-gratification</td>
</tr>
<tr>
<td>Generation Xers, 1965-1982</td>
<td>Latchkey kids</td>
</tr>
<tr>
<td></td>
<td>Extended adolescence</td>
</tr>
<tr>
<td></td>
<td>Values</td>
</tr>
<tr>
<td></td>
<td>Work</td>
</tr>
<tr>
<td></td>
<td>Org. culture</td>
</tr>
<tr>
<td>Duty</td>
<td>Self-fulfillment</td>
</tr>
<tr>
<td>Self-fulfillment</td>
<td>Opportunity for advancement</td>
</tr>
<tr>
<td>Opportunity for advancement</td>
<td>Work is leisure</td>
</tr>
<tr>
<td>Work is leisure</td>
<td>Command and control</td>
</tr>
<tr>
<td>Command and control</td>
<td>Consensus</td>
</tr>
<tr>
<td>Consensus</td>
<td>Participation</td>
</tr>
<tr>
<td>Participation</td>
<td>Idealists</td>
</tr>
<tr>
<td>Idealists</td>
<td>Self-directed</td>
</tr>
<tr>
<td>Self-directed</td>
<td>Competence</td>
</tr>
<tr>
<td>Competence</td>
<td>Realists</td>
</tr>
<tr>
<td>Realists</td>
<td>Free agents</td>
</tr>
<tr>
<td>Free agents</td>
<td>Self-reliant</td>
</tr>
</tbody>
</table>

There are several recognizable explanations for some of the current challenges, some already clearly described Dr Gibbons. To these, I want to highlight a few:

### Table II | Assets and liabilities

<table>
<thead>
<tr>
<th>Baby boomer assets</th>
<th>Generation Xer assets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team player</td>
<td>Independent/individualistic</td>
</tr>
<tr>
<td>Driven</td>
<td>Adaptable</td>
</tr>
<tr>
<td>Service oriented</td>
<td>Techno-literate</td>
</tr>
<tr>
<td>Willing to go the extra mile</td>
<td>Creative</td>
</tr>
<tr>
<td>Baby boomer liabilities</td>
<td>Unintimidated by authority</td>
</tr>
<tr>
<td>Not budget minded</td>
<td>Generation Xer liabilities</td>
</tr>
<tr>
<td>Self-centered</td>
<td>Impatient</td>
</tr>
<tr>
<td>Judicial</td>
<td>Poor people skills</td>
</tr>
</tbody>
</table>

A series of authors have identified and classified generations as the traditionalist or silent generation (1925-1940), the baby boomers (1945-1963), the generation Xers (1963-1982), and generation Y (1985-). With respect to health care and medicine, “the silent generation” saw medicine as mysterious and doctors as god; “baby boomers” craved convenience and control and saw health problems as major life events; and generation X, Y, and onward are highly educated, highly uninsured, and Internet savvy and think they can diagnose their own disease. For boomers, seminal events are Vietnam, civil rights, birth control pills, and Woodstock and for generation Xers, personal computers, AIDS, the energy crisis, and now 9/11.

These life experiences lead to different values (Table I) and different assets and liabilities (Table II). To simplify, boomers are workaholics and generation Xers strive to “get a life.” These differences lead to different approaches to work-life conflicts and low morale. Generation Xers will be more direct and outspoken, giving the impression that they are self-centered. In fact, expectations about work hours, time spent at work, and productivity as well as financial remuneration can be very different. Mentors need to understand these generational differences and use this knowledge to best advise and assist students, residents, and young colleagues.
Gender issues

We must also understand that in addition to generational issues, gender inevitably will play a role in all medical practice, not just obstetric and gynecologic practice in the future.\(^{21}\) With more than 50% of medical students now women, with more than 75% of all obstetrics and gynecology residents now women, “attention must be paid” to gender: women’s roles in society, families, and professions are inevitably going to influence change—practice style and tenure, to mention just a few—and this change will affect and benefit women and men in the years to come.

Differences that have been described, and I am grossly oversimplifying, show that women and the groups they form tend to use expressive skills; show compassion; and value relationships, interpersonal connections, and interactions as compared with traditional male individualism, separation, and independence. So if you add family and personal concerns, the biomedical information explosion, administrative burdens, debt and liability, and the 80-hour work week to generational issues and gender, we have a perfect storm that for our specialty seems sometimes to rival the most destructive natural disaster.

So how do we move forward? I believe we need to be bold and aggressive. Peterson and Willson both were, but I am concerned that recently we have seen too much hand wringing, too much backward looking and longing for the past, too few innovative ideas and innovations, too little execution. In one of my favorite recent books, *Execution: The Discipline of Getting Things Done,* \(^{22}\) 2 business leaders speak of the importance of execution:

> “Everybody talks about change. In recent years, a small industry of changemeisters has preached revolution, reinvention, quantum change, breakthrough thinking, audacious goals, learning organizations, and the like. … But unless you translate big thoughts into concrete steps for action, they’re pointless. Without execution, the breakthrough thinking breaks down, learning adds no value, people don’t meet their stretch goals, and the revolution stops dead in its tracks. What you get is change for the worse because failure drains the energy from your organization. Repeated failure destroys it.”

So I advocate for action and execution in the clinical training and academic laboratories that I described above. This active transformation of our specialty, which will quickly and inevitably spread, will serve as an important signal to our medical students and residents that what we are doing has much to offer for them, for Bossidy and Charan also write:

> “The beliefs that influence specific behaviors … need changing. These beliefs are conditioned by training, experience, what people hear … about prospects and perceptions about what leaders are doing and saying. People change them only when new evidence shows them persuasively that they’re false. … We don’t think ourselves into a new way of acting; we act ourselves into a new way of thinking.”

We must share in the criticism Bossidy and Charan level at those who have spent too much time on strategic planning and managing operations—clinical operations for many us—and not enough on execution. Gibbons, Gabbe, and others have shown us the challenges and suggested solutions, American College of Obstetricians and Gynecologists task forces are giving us plans and opportunities: We need to move forward. We need to try new ways of doing things: laborists, generalists, computerized continuing medical education, simulated surgical training, night floats, shared practices, liability reform, preferably in departments that can share their examples of best practice with others. This is what our predecessors did. They are looking for us to do the same.

In closing, I would like to adapt and invoke the challenge and spiritual commitment of my favorite hymn: “Let there be thoughtful, progressive transformation of the specialty of obstetrics-gynecology, and let it begin with us.”

References

The essential elements of undergraduate medical education in obstetrics and gynecology: A comparison of the Association of Professors of Gynecology and Obstetrics Medical Student Educational Objectives and the National Board of Medical Examiners Subject Examination

Maya M. Hammoud, MD,a Susan M. Cox, MD,b Barbara Goff, MD,c Alice Goepfert, MD,d Aggie Butler, PhD,e David B. Swanson, PhD,e Kathleen Z. Holtzman, BS,e Krista Allbee, BA,e Nadine T. Katz, MD,f Sonya S. Erickson, MDg

University of Michigan Medical School, Ann Arbor, MI; University of Texas Southwestern Medical Center, Dallas, TX; University of Washington, Seattle, WA; University of Alabama at Birmingham, AL; National Board of Medical Examiners, Philadelphia, PA; Albert Einstein College of Medicine, New York, NY; University of Colorado Health Sciences Center, Denver, CO

Received for publication February 24, 2005; accepted August 8, 2005

KEY WORDS
Association of Professors of Gynecology and Obstetrics
National Board of Medical Examiners
Undergraduate medical education
Educational objectives
Obstetrics and gynecology subject examination

Objective: The objective of this study was to investigate whether the essential elements of the Association of Professors of Gynecology and Obstetrics (APGO) Medical Student Educational Objectives were adequately represented on the National Board of Medical Examiners (NBME) obstetrics and gynecology subject examination, and that the topics questioned on that examination were covered by the APGO objectives.

Study design: The Undergraduate Medical Education Committee of APGO and the NBME staff separately reviewed the same 2 NBME obstetrics and gynecology subject examinations. The questions were mapped to the 15 essential elements of the APGO educational objectives and comparisons were made to check how well they matched.

Results: All the essential elements of the educational objectives were covered by the NBME subject examination. Of the questions on the examination, 99% were deemed appropriate for medical students with 70% of the questions mapping to “Priority 1” objectives.

Conclusion: The NBME examination provides an appropriate assessment of mastery of what a medical student should learn, as represented by the APGO Medical Student Educational Objectives.

© 2005 Mosby, Inc. All rights reserved.
Medical education is currently undergoing numerous changes at both the graduate and undergraduate levels. With the potential effects on student teaching of residents' 80-hour work week restriction and increasing clinical demands on faculty, as well as the decentralization of students to multiple sites during the obstetrics and gynecology clerkship, it has become more challenging than ever for clerkship directors to ensure a uniform quality educational experience for all their students. Therefore, clear learning objectives to guide faculty and students about what students should learn become increasingly important, so that learning outcomes do not fluctuate widely as a function of the site and/or clerkship in Obstetrics and Gynecology (Ob-Gyn) is the major source of curriculum guidance for the third-year students. The NBME Subject Examination is generally given at the end of the rotation to test applicants' fund of knowledge in the field of Ob-Gyn and presumably represent what is important to be covered in the obstetrics and gynecology clerkship. This examination is typically given at the end of the clerkship or unit of instruction. The majority of third-year Ob-Gyn clerkships use the National Board of Medical Examiners' (NBME) Ob-Gyn Subject (shelf) Examination as an objective assessment.

The majority of third-year Ob-Gyn clerkships use the National Board of Medical Examiners' (NBME) Ob-Gyn Subject (shelf) Examination as an objective assessment. According to the NBME, in 2004, the Ob-Gyn Subject Examination was used by 118 of the 125 US medical schools accredited by the Liaison Committee on Medical Education. Subject examinations are primarily designed to provide institutions with effective evaluation tools and useful student performance data that can be compared with a large representative group of students at the same stage of training. This examination is generally given at the end of the rotation to test applicants' fund of knowledge in the field of Ob-Gyn. It typically accounts for about 30% of the final grade with a range of 25% to 70%. The NBME Subject Examination questions are written by experts in the field of Ob-Gyn and presumably represent what is important in the subject for an undifferentiated medical student to know. However, agreement between the published APGO objectives and the test items on the NBME subject examination has never been previously evaluated in any systematic fashion.

The primary objective of this study was to investigate whether the essential elements of the recently revised APGO Medical Student Educational Objectives, 8th edition, were adequately represented on the Ob-Gyn Subject Examination. Likewise, we wanted to validate that the topics questioned on the NBME Ob-Gyn Subject Examination were covered by the APGO educational objectives.

Materials and methods
The APGO Medical Student Educational Objectives, 8th edition
The UMEC of the APGO prepared the 8th edition of the Medical Student Educational Objectives in 2003 and 2004. To develop this major revision, the UMEC edited the previous edition of the educational objectives for the medical school clerkship to ensure coverage of significant additions and changes in the field and to better define the essential knowledge and experience recommended for medical students in Ob-Gyn. In doing the work, UMEC used the process described by Roy M. Pitkin, MD, in his presidential address to the American Gynecological and Obstetrical Society in September 1996, titled “But Isn’t That Your Job, Son?” Doctor Pitkin broke down the then existent 60 APGO Medical Student Educational Objectives, 6th edition, into 3 categories (priorities):

Priority 1: Objectives that he believed were of such importance that all medical students must learn and master them;
Priority 2: Objectives that he believed students should be expected to learn; and
Priority 3: Objectives that he believed medical students can be expected to learn.

He described this nomenclature as, “Need to know, nice to know, and nuts to know.” In his analysis, Dr Pitkin grouped the “need to know” objectives into 7 essential elements, which he proposed should be the “essence of the medical student experience in Ob-Gyn.”

The UMEC followed Dr Pitkin's format and identified 15 essential elements that encompass 37 of the 58 educational objectives as shown in Table I. The APGO UMEC proposes that these 15 elements, to use Dr Pitkin’s words, represent the current “essence of the medical student experience in obstetrics and gynecology.”

The NBME Ob-Gyn Subject Examination
Since 1989, the Ob-Gyn Subject Examination has been 1 of the 7 clinical science discipline-based examinations provided through the NBME Subject Examination Program. The examination is a 100-item multiple-choice question examination designed to be administered at the end of the clerkship or unit of instruction. The majority
of questions on the examination have been used previously on the USMLE Step 2 Clinical Knowledge (CK) Examination. Before scored use on the USMLE (and subsequently the subject examination), all questions must receive 2 separate approvals: first by content experts (Step 2 Obstetrics and Gynecology Test Material Development Committee) and second by an Interdisciplinary Committee composed of a surgeon, a psychiatrist, a pediatrician, an ambulatory medicine physician, an internist, and an obstetrician/gynecologist. The purpose of the Ob-Gyn content on Step 2 CK is to assess knowledge of common Ob-Gyn problems that a generalist would need to know, rather than focusing on what a specialist would need to know. For example, questions on gynecologic cancers deal with recognition of the cancer rather than management.

NBME staff developed draft forms of the subject examination based on content and statistical specifications. Each form must meet a specific statistical target to ensure that the difficulty of new examinations remains

<table>
<thead>
<tr>
<th>Table I</th>
<th>The essential elements and their corresponding educational objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essential element</td>
<td>Related educational objective(s)</td>
</tr>
</tbody>
</table>
| 1. Clinical skills in history-taking and physical examination | 1. History  
2. Examination  
4. Diagnosis and management plan  
5. Personal interaction and communication skills  
3. Pap smear and cultures |
| 2. Collect and interpret a cervical cytology—first line disposition, limitations of cervical cytology | 33. Contraception and sterilization  
15. Ectopic pregnancy |
| 3. Thorough grounding in modern contraceptive technology | 36. STIs and UTIs  
8. Maternal fetal physiology |
| 4. Differential diagnosis of the “acute abdomen”—pelvic infection, ectopic pregnancy, adnexal torsion, appendicitis, diverticulitis, renal calculi | 10. Antepartum care  
9. Preconception care |
| 5. Physiologic adjustments that accompany normal gestation—especially laboratory test results | 6. Legal and ethical issues  
5. Preventive care and health maintenance  
16. Spontaneous abortion  
33. Contraception and sterilization  
57. Sexual assault  
58. Domestic violence  
34. Abortion  
43. Amenorrhea  
45. Normal and abnormal uterine bleeding  
46. Dysmenorrhea  
47. Menopause  
38. Endometriosis  
48. Infertility |
| 6. Embryonic and fetal development—what does and does not affect it, what is and is not teratogenic | 11. Intrapartum care  
14. Lactation  
40. Disorders of the breast  
35. Vulvar and vaginal disease  
56. Sexuality and modes of sexual expression |
| 7. Health and well-being of populations—social and health policy aspect of women’s health, ethical issues, sterilization, abortion, domestic violence, adolescent pregnancy, access to health care | 17. Medical and surgical conditions in pregnancy  
18. Preeclampsia/eclampsia  
23. Third-trimester bleeding  
30. Postterm pregnancy  |
| 8. Menstrual cycle, including menopause | 51. Vulvar neoplasms  
52. Cervical disease and neoplasia  
53. Uterine leiomyomas  
54. Uterine carcinoma  
55. Ovarian neoplasms |
| 9. Infertility | |
| 10. Intrapartum care  
11. Intrapartum care |
| 12. Vaginal and vulvar disorders | |
| 13. Sexuality—patient and physician | |
| 14. Common problems in obstetrics | |
| 15. Screening for reproductive cancers | |
constant across examination forms within and across years. Table II shows the distribution of questions as assigned by the NBME staff. Each draft examination form undergoes a stringent review by 4 to 5 Ob-Gyn clerkship directors to ensure quality and appropriateness. Each question is graded for appropriateness and content and the entire examination is evaluated for deficiencies or redundancies of topics. Once the reviews are complete, NBME staff collates grades and comments for each item for review by the chair of the Step 2 Ob-Gyn Test Material Development Committee. The chair reviews all comments and replaces questions deemed inappropriate or too difficult on the basis of reviewer input. In general, only 4% to 8% of questions are replaced.

Comparison of the APGO Objectives and the NBME Subject Examination

To investigate the congruence between the topics of the recently revised objectives and the NBME subject examination, in 2004 the APGO UMEC and the NBME staff separately reviewed the same 2 NBME Ob-Gyn shelf examinations. The UMEC mapped each question to the specific learning objective(s) under 58 educational topics in the APGO Medical Student Educational Objectives, 8th edition. Because the majority of the NBME test items involve some component of diagnosis and management, the UMEC chose not to code to this topic (educational topic 4).

The NBME staff mapped the questions to the learning objectives using the question’s diagnosis or disease as a guideline for choosing the appropriate learning category. The NBME staff considered this task to be challenging because the learning objectives provided a 1-dimensional way of coding items and NBME uses a multidimensional coding system that maps questions based on diagnosis and physician task (eg, health and health maintenance, mechanism of disease, diagnosis, and management) as outlined in Table II. The NBME process is illustrated in the Figure. As shown, the NBME staff did map to “diagnosis and management” for each of the test items that fit into this category. Both the UMEC and NBME mapped test items to multiple educational topics when indicated and the results were compiled under each of the 15 essential elements of the educational objectives.

This study did not involve experimentation on human or nonhuman subjects and was exempt from Institutional Review Board approval.

### Table II  Content outline for NBME subject examination

<table>
<thead>
<tr>
<th>Gynecology</th>
<th>Range</th>
<th>Obstetrics</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promoting/Maintaining Health</td>
<td>5%-10%</td>
<td>Prenatal care</td>
<td>5%-10%</td>
</tr>
<tr>
<td>Postmenarchal/reproductive</td>
<td></td>
<td>Assessing at-risk pregnancy</td>
<td></td>
</tr>
<tr>
<td>Peri- and/or postmenopausal</td>
<td></td>
<td>Intrapartum care</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Contraception</td>
<td></td>
</tr>
<tr>
<td>Mechanisms of disease</td>
<td>15%-20%</td>
<td>Placental dysfunction</td>
<td>10%-15%</td>
</tr>
<tr>
<td>Infections</td>
<td></td>
<td>Pregnancy and labor</td>
<td></td>
</tr>
<tr>
<td>Neoplasms</td>
<td></td>
<td>Postpartum disorders</td>
<td></td>
</tr>
<tr>
<td>Breast disease</td>
<td></td>
<td>Fetus and newborn</td>
<td></td>
</tr>
<tr>
<td>Urinary</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Menstrual</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endocrine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infertility</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Establish diagnosis</td>
<td>15%-20%</td>
<td>Normal pregnancy</td>
<td>15%-20%</td>
</tr>
<tr>
<td>Infections</td>
<td></td>
<td>Nonobstetric complications of pregnancy</td>
<td></td>
</tr>
<tr>
<td>Neoplasms</td>
<td></td>
<td>Obstetric complications</td>
<td></td>
</tr>
<tr>
<td>Breast disease</td>
<td></td>
<td>Complications of puerperium</td>
<td></td>
</tr>
<tr>
<td>Urinary</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Menstrual</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endocrine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infertility</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Principles of management</td>
<td>5%-10%</td>
<td>Pharmacologic treatment</td>
<td>10%-15%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Treatment/diagnosis steps</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nonpharmacologic treatment</td>
<td></td>
</tr>
<tr>
<td>General principles</td>
<td>—</td>
<td></td>
<td>1%-5%</td>
</tr>
</tbody>
</table>
A previously healthy 23-year-old woman comes to the physician because of pain and burning with urination for 3 days and loss of urine several times over the past 24 hours. She has not had fever or chills. Examination shows no flank or costovertebral angle tenderness. Urinalysis shows leukocytes too numerous to count and numerous rod-shaped bacteria. Which of the following is the most likely mechanism of this patient’s urinary symptoms?

(A) Decreased bladder muscle tone
(B) Delayed neuromuscular maturation
(C) Ectopic insertion of the ureters
* (D) Trigonal inflammation
(E) Uninhibited detrusor muscle conditions

APGO Classification: 36 (STIs and UTIs)
NBME Classification: Gynecology–Mechanisms of disease–Infection
% Answering correctly: Ob-Gyn Subject Examination–90%; Step 2–81%.

A 23-year-old primigravid woman at 42 weeks’ gestation is admitted in labor 6 hours after spontaneous rupture of the membranes. Contractions occur every 5 minutes. The cervix is 2 cm dilated and 90% effaced; the vertex is at –2 station. The estimated fetal weight is 3500 g (7 lb 11 oz). Her temperature is 37.2°C (99°F), blood pressure is 110/74 mm Hg, and pulse is 88 beats per minute. External fetal monitoring shows a fetal heart rate of 156 beats per minute with good variability. Which of the following is the most appropriate next step in management?

* (A) Observation
(B) Biophysical profile
(C) Administration of magnesium sulfate
(D) Administration of oxytocin
(E) Cesarean delivery

APGO Classification: 11 (intrapartum care), 4 (diagnosis and management)
NBME Classification: Obstetrics–Management–Treatment/Diagnosis steps
% Answering correctly: Ob-Gyn Subject Examination4–71%; Step 2–84%.

*Correct answers

Figure  Sample items from the NBME Subject Examination in Ob-Gyn showing NBME mapping to the APGO objectives.
Results

The results of the 2004 review of the 2 NBME examinations (n = 200 questions) by the UMEC and the NBME staff are shown in Table III.

According to the APGO UMEC review, nearly 100% of the questions were viewed as appropriate on the basis of the competencies contained in the APGO 8th edition Medical School Educational Objectives. A total of 245 entries were mapped to the objectives because several questions were mapped to more than 1 educational topic. Of these, 171 (70%) were Priority 1 (need to know), 72 (29%) were Priority 2 (nice to know), and 2 (1%) were Priority 3 (nuts to know). Only the 171 Priority 1 items are listed in Table III. The NBME mapping generated 246 entries because many items were double coded to topic 4 (Diagnosis) under essential element 1 in addition to the actual diagnosis. The UMEC did not code to this topic. Of the 58 APGO educational topics, 56 were represented on the NBME Subject Examination. No test items were coded to topic 34 (Abortion) or topic 58 (Domestic Violence).

In general, there was high consistency between the APGO UMEC coding and the NBME coding (80% consistent). However, there were some differences that were most likely due to the NBME coding scheme not explicitly addressing history, examination findings, or obstetric procedures. As mentioned previously, the NBME coder made the decision to map questions using NBME’s primary coding scheme and not code to multiple categories. This strategy diverged from that of the APGO UMEC.

Comment

There have been many publications in the literature regarding the contribution of the subject examination to students’ final grades. To the best of our knowledge, this is the first time an NBME subject examination in a clerkship has been systematically mapped to well-established clerkship educational objectives. Our findings indicate that the NBME subject examination matches well with the APGO objectives. This reflects the similar philosophy underlying the subject examination development and the APGO objectives: ensuring that students learn general Ob-Gyn principles, which they will need, independent of what field they choose.
All medical students ask the question, “What do I need to know?” The APGO UMEC proposed the essential objectives to serve as a guide for curriculum design and content as well as a study guide to prepare for national licensure tests. Most medical schools across the country use the NBME subject examination as an objective assessment of performance of their medical students. The use of a written examination in this fashion presents some challenges as many believe that the examination begins to drive the curriculum. However, the use of a standardized examination such as the NBME subject examination provides an easily obtainable measure of knowledge and allows for a comparison across different groups, such as comparing groups at different times of the year or comparing groups between different sites. In addition, it allows the schools to examine how their students perform compared with the national norm.

Our study findings should assure Obstetrics and Gynecology Clerkship Directors that the NBME examination provides an appropriate assessment of mastery of what a third-year medical student should learn as represented by the APGO Medical Student Educational Objectives.

Acknowledgments

We gratefully acknowledge the significant contributions of the Undergraduate Medical Education Committee (UMEC) of APGO in reviewing the NBME examinations: Jessica Bienstock, MD, Susan M. Cox, MD, Eve Epsy, MD, Alice Goepfert, MD, Maya M. Hammoud, MD, Nadine T. Katz, MD, James J. Neutens, PhD, Edward Peskin, MD, Elizabeth Puscheck, MD, Paul Krueger, DO, Sonya S. Erickson, MD.

References

Medical students self-reported work hours: Perception versus reality

Colleen Casey, MD,a Sangeeta Senapati, MD,a Casey B. White, PhD,b Larry D. Gruppen, PhD,c Maya M. Hammoud, MDa

Department of Obstetrics and Gynecology,a Office of Medical Education,b Department of Medical Education,c University of Michigan Medical School, Ann Arbor, MI

Received for publication February 21, 2005; revised July 3, 2005; accepted August 8, 2005

Objective: The objective of this study was to compare the students’ actual work hours with their self-reported work hours during the obstetrics and gynecology clerkship, and to determine whether the number of hours worked correlate with the amount of “scut” reported or students’ rating of the quality of the clerkship.

Study design: Students self-reported work hours were compared against their actual scheduled hours over 2 different academic years. Pearson’s correlation was performed to correlate the actual hours with the amount of reported “scut” work and the overall rating of the quality of the clerkship.

Results: The actual hours per week worked by students averaged 59 hours in 2003 and 48 hours in 2004. Students overestimated their work hours both years. Students who worked more hours rated the clerkship lower and the quality of the clerkship significantly improved from 2003 to 2004 (4.2 vs 3.8 P < .03).

Conclusion: The majority of third-year students overestimate their work hours in obstetrics and gynecology. The rating of the overall quality of the clerkship increase significantly with fewer hours worked, and it is not affected by the amount of “scut” work.

The adoption of the 80-hour work week has resulted in major changes in residency programs in the United States. Although the airline, trucking, and railway industries have a longstanding history of having work hours regulated by the government,1 until recently work hour restrictions in the medical field were not mandated. However, on June 11, 2002, the Accreditation Council for Graduate Medical Education (ACGME) passed universal standards limiting resident work hours to 80 hours per week. These new regulations became effective July 1, 2003.

In February 2004, the Liaison Committee on Medical Education (LCME) officials added an annotation regarding student work hour regulations to its accreditation standards.2 They recommended that in addition to monitoring the amount of classroom time and examination frequency, medical schools should pay attention to the hours that medical students work during the clinical years and the educational value of their clinical activities. Student duty hours should be set taking into account the effects of fatigue and sleep deprivation on learning and patient care, and medical students should not be required to work longer hours than residents. Soon after this decision by the LCME, the Medical
Student Section (MSS) of the American Medical Association (AMA) passed a resolution (MSS Resolution 16: Medical Student Clinical Training and Education Condition),\(^3\) asking the AMA to encourage the LCME to follow the lead of the ACGME and include specific limits on the workload of students in its accreditation standards. Namely, the resolution for medical student education called for an 80-hour work week, a 24-hour limit on shifts, and on call no more than once every 3 nights.

As early as 2001, 17 of 124 US medical schools reported having formal policies on medical students hours;\(^5\) however, little is known about the number of actual or reported student work hours. Burke et al published self-reported students’ work hours during their third-year required clerkships; in that study students indicated an average between 47 to 87 hours per week depending on the clerkship, with 2 clerkships reaching more than 80 hours. A study by Chen et al at Harvard Medical School, which asked medical students to track their daily hours for 1 week, indicated that students spent a mean of 74 hours per week in the hospital, with some reporting as many as 106 hours per week. Mean work hours were dramatically higher on inpatient rotations (90 hours per week) than on outpatient rotations (45 hours per week). Students on surgery rotations worked the longest hours, with a mean of more than 94 hours per week spent in the hospital during an inpatient rotation.

At the University of Michigan Medical School, when the 80-hour work week regulation was adopted for residents, medical students began to report anecdotally an increase in their work hours. In addition, curriculum leaders considered adopting similar restrictions on medical students work hours, so starting in 2003, students on all required clerkships (internal medicine, surgery, obstetrics and gynecology, pediatrics, family medicine, neurology, psychiatry) were asked at the end of the clerkship to report how many hours they were required to be in the hospital or clinics on a 4-point categorical scale as follows: 1 = less than 60 hours; 2 = 61 to 70 hours; 3 = 71 to 80 hours; 4 = more than 80 hours.

To reduce the number of medical student work hours, a student “night float” system was initiated in 2004, where students’ on-call responsibilities consisted of 3 consecutive 12-hour night shifts and a 12-hour shift on the weekend. This replaced the regular once-a-week call from the previous year. For the purposes of this study, we considered each year separately because of the curriculum change. After institutional review board exemption was obtained, all the students rotating at the University Hospital (70 in 2003 and 65 in 2004) were considered for the study. Although data were available on the reported work hours for most of the students because more than 95% filled out a clerkship survey, we were unable to retrieve all the actual work schedules that resulted in available data for 35 students from 2003 and 42 students from 2004. In calculating the actual hours, we added the number of hours scheduled each day to calculate average hours per week. The hours per week were then combined and divided by 6 to calculate average hours per week worked.

To compare students’ scheduled or actual work hours with their reported or perceived work hours, the students were separated into 4 groups on the basis of their reported work hours category 1 to 4 (1 = <60 hours; 2 = 61-70 hours; 3 = 71-80 hours; 4 = 80 hours). These were compared against the actual work hours we calculated for them. In addition, Pearson’s correlation was performed to correlate the actual hours with the amount of reported scut work and the overall rating of the quality of the clerkship. The \(\chi^2\) analysis was performed to check if the amount of scut work was different between 2003 and 2004, and an analysis of variance was performed to compare the overall quality of the clerkship between the 2 years.

**Results**

The actual hours per week worked by students were higher in 2003 (59 hours, range: 52-68) than in 2004 (48 hours, range: 43-52). The estimated hours worked per week were also much higher in 2003 than 2004 as

**Methods**

At the beginning of the 6-week Ob-Gyn clerkship, all students rotating at the University Hospital were given a work schedule. The schedule specified their daily assignments including clinics, inpatient services, and on-call responsibilities. At the end of the clerkship, all students completed a confidential online survey rating the quality of their experiences on a 5-point Likert scale from 1 to 5 (1 = poor; 5 = excellent). The questionnaire covered several areas of their experiences including the overall quality of the clerkship. They were also asked to estimate the amount of time spent in academically unproductive work on a scale of 1 to 5 (1 = too little; 3 = just right; 5 = too much), and the average number of hours per week they were required to be in the hospital or clinics on a 4-point categorical scale as follows: 1 = less than 60 hours; 2 = 61 to 70 hours; 3 = 71 to 80 hours; 4 = more than 80 hours.
shown in Table I. At least 71% of the students overestimated their work hours in 2003 compared with 53% of the students in 2004. More interestingly, although the actual hours worked was only about 10 more hours in 2003 than it was in 2004, 71% of the students reported working more than 70 hours per week in 2003 and only 12% reported working that many hours in 2004. Students reporting less than 60 hours per week were the most accurate, whereas those reporting more than 80 hours were the least accurate (Figure).

The actual hours worked correlated negatively with the overall rating of the quality of the clerkship for 2004 and for both years combined. Students who worked more hours rated the clerkship lower (Table II). In addition, the actual number of hours worked correlated with the amount of scut work for both years combined. This relationship did not hold true when considering each year separately. There was no significant correlation between the rating of the overall quality of the clerkship and the amount of scut work. There was no significant distribution difference in the amount of scut work reported between 2003 and 2004; however, the overall quality of the clerkship did significantly improve in 2004 (4.2 vs 3.8 $P < .03$).

## Comment

The reported hours and the actual hours worked by medical students decreased in 2004, which reflected a restructuring of the clerkship. Students no longer had regular on-call duties, which eliminated 24- to 36-hour shifts, so the actual number of hours worked per week decreased from 59 hours in 2003 to 48 hours in 2004. Although a large number of students were still overestimating their hours in 2004, they represented a smaller percentage than the students who overestimated their hours in 2003. These findings indicate that students tend to overestimate their work hours more when they are actually working more. In addition, the more hours students worked, the more scut they reported. This could indicate that with longer hours, students are either spending more time in academically unproductive activities or the students potentially perceive less quality to their learning experience because they are tired.

Our data show that decreasing the number of work hours and providing a structured schedule contributed to higher student ratings of the Ob/Gyn clerkship. Several studies support the notion that a structured work schedule results in increased student satisfaction with the clerkship.7-10 Considering that the Ob/Gyn clerkship has been the lowest rated clerkship by medical students,11 these findings support the importance of developing a structured clerkship with reasonable hours to improve the quality of the educational experience for third-year medical students.
It is interesting that third-year medical students reported spending more hours at the hospital than what they were scheduled to work. We thought one explanation for this might be that some students elected to stay and work beyond their scheduled hours. However, a discussion with students and residents indicated this was not the case. Although there were occasional instances where a student stayed late for an interesting case, students on gynecology call were also often sent home ahead of schedule. Students were also often excused early and they were not expected to be present for gynecology afternoon rounds. In addition, when we calculated their hours, we tried to err on the side of overestimating. For example, we added extra hours for students on gynecologic oncology, assuming they stayed late every day although this only happened on average 2 times a week. We also assumed students were at the hospital 1 hour before rounds, which is often not true. In any case, even if there were instances where a student might have worked more hours than scheduled, it was impossible to account for all of the extra reported hours, which in some cases reached more than 20 hours per week.

A second possibility might be that students simply are not able to estimate their work hours accurately. It has been suggested that medical students’ estimations of their work hours may actually be a reflection of residents’ work hours observed by the students. In addition, previous studies in medical and higher education have shown that students are not very accurate in self-assessing their performance. When looking specifically at medical student self-assessment, investigators have found that the less experience a student has (ie, transitioning from the familiarity of the classroom to the clinical clerkships), the more self-assessment accuracy and performance suffers. Student work hours can potentially fall into this category where third-year medical students are simply not experienced enough to accurately assess their hours.

A third possibility for these findings might be that students think they are working more hours because they are working hard or because of the many variations in their daily schedule. The Ob-Gyn clerkship is a challenging clerkship. Students rotate through so many different services including the operating room, labor and delivery, clinic, or night float. Some of these services can be very demanding and stressful and students might tend to recall these experiences more than others. This makes it difficult for students to accurately recall the hours by simple reflection. It would be interesting to perform similar analyses for other clerkships, specifically, some of the less demanding ones with more outpatient setting and no on-call duties to check if there would be similar findings.

Whatever the causes are, this perception by medical students that they are working longer hours than they actually are can have serious implications. The perceived work hours required in surgical residencies play a part in a decreased interest in those specialties. For example, Miller et al surveyed 134 third- and fourth-year medical students to determine factors that influenced their decision in choosing a specialty. They found that 38% would not consider a residency program with a reputation for long work hours. In a separate study, Miller et al looked at how students entering surgical fields developed their personal rank list. They found that 40% of the applicants ranked higher those programs that strictly adhered to work hour limitations. Thus, as students estimate long working long hours in obstetrics and gynecology they might also consider those hours negatively as they consider their residency choices. This is of special importance to the Ob-Gyn specialty because of the decreased number of students choosing to enter the field.

The adoption of the 80-hour work week regulation mandated by the ACGME has resulted in a restructuring of several residency programs nationwide. Much discussion has been generated about its impact on residents’ experience and job satisfaction. When considering restricted work hours for medical students we must remind ourselves of the role of the medical student in the hospital—medical students are in the hospital to learn. It is our duty to create an ideal environment for them to learn by maximizing good educational opportunities and minimizing “scut” work. Furthermore, we have to consider how those perceptions of long hours might be impacting students’ career choices especially in obstetrics and gynecology.

References


Development and assessment of a Web-based evaluation and management coding curriculum for residents

Paul M. Lemen, MD*

Department of Obstetrics and Gynecology, Medical College of Wisconsin, Milwaukee, WI

Received for publication March 2, 2005; revised June 8, 2005; accepted August 1, 2005

**Objective:** The purpose was to assess a Web-based learning module for residents on evaluation and management coding, and to determine whether using a point system based on reimbursement is an effective way to measure improvement in performance.

**Study design:** Nineteen residents at a university-based residency completed an online module on evaluation and management coding. Pretest and posttest were administered consisting of 10 actual patient notes from which to abstract the level of service. Tests were scored by percent correct as well as assigning points to correct responses equal to 2004 Medicare reimbursement for that item's level of service. Incorrect responses were debited points equal to the absolute difference in reimbursement between the incorrect and correct evaluation and management level.

**Results:** Average percent correct was 44.2 on the pretest and 45.3 on the posttest ($P = 1.0$), with only 47% of subjects improving their performance. The mean point score was 513.44 on the pretest and 555.87 on the posttest ($P = .02$), with 84% improving. Most subjects (84%) rated the module equal or superior to a didactic format and felt that their knowledge was improved by the exercise.

**Conclusion:** An electronic learning module on evaluation and management coding for residents, using scoring based on reimbursement schedules, is effective for instruction and assessment and well accepted by learners.

Evaluation and management (E&M) coding, part of the American Medical Association’s current procedural terminology system, serves as the method by which physicians bill patients and third-party payers for many clinical services. As such, understanding of the E&M coding system and the necessary documentation that accompanies it are an important component to the current practice of medicine in this country. Improper coding and documentation can lead to either overbilling for services (which may constitute fraud) or undercompensation for services rendered. Despite this fact, there are numerous studies that indicate that physician competence in the application of E&M coding is poor.

To date, there are no studies that assess an educational method for instructing residents in this important subject. At the Medical College of Wisconsin Affiliated Hospitals, obstetrics and gynecology residents have traditionally received an annual 60-minute lecture from a departmental billing professional. This study was designed to construct and assess an interactive Web-based electronic module to educate residents in E&M coding.
Chart documentation:

CHIEF COMPLAINT: Vulva check.

HPI: Patient is a 62-year-old G3, P2 white female many years postmenopausal. She was last seen in this clinic approximately two years ago for an annual exam. Patient was seen by Dr. M in September for routine exam. Pap smear performed at that time was normal, but patient had significant changes of the vulva and multiple complaints, and she was referred here for evaluation. Patient states she has significant vulvar itching, occasional burning. She notices frequent abrasions with slight bleeding, especially after strenuous activity. She does use occasional topical over-the-counter steroids with some relief, because if she does not, she finds herself scratching at night. She is not on any hormones. She is not currently sexually active.
PAST MEDICAL HISTORY: Otherwise unchanged.

PHYSICAL EXAM:

GENERAL: A thin white female in no acute distress.
VITAL SIGNS: Her weight is 125, blood pressure 124/60.
EXTERNAL GENITALIA: Anatomically intact. She does have significant lichenification of the labia and perianal area. There is a loss of labial integrity around the clitoris. The labia are somewhat excoriated, very thin, and pale. The vagina is pale and very thin. Urethra, as well, shows some petechial formation.

IMPRESSION: Vulvovaginal atrophy and significant lichen sclerosis.

PLAN: I have recommended that the patient use vaginal estrogen in the form of estradiol tablets (0.025 mg) twice weekly and that she start a high-potency steroid. She was given a prescription for 0.05% clobetasol to be used twice daily for 2 weeks, once daily for 2 weeks, and then once or twice a week as needed for control. She was shown with a mirror where to apply the ointment to make sure that she gets all of the areas, perianally, especially. The vaginal tablet she will use twice a week as well as I think the excoriations are coming from significant atrophy. Patient will return in 6 to 8 weeks for repeat assessment, and if there is not significant improvement, we will do a biopsy. Otherwise, she will call with problems in the meantime.

Abstraction explanation:

This documented encounter contains a chief complaint, brief HPI, and problem-pertinent review of systems. The History element achieves a Level 3 service. The Physical exam element is problem-focused, so achieves a Level 2 service. The Medical decision making is overall of low complexity, achieving a Level 3 service for this element. For an established patient, the level of service met or exceeded by two of the three elements dictates level of service, so this encounter is coded as an established patient, level 3 (99213).

Figure Example item from pretest and posttest consisting of documentation of patient encounter (with brief explanation of coding abstraction).

coding and documentation and gauge resident acceptance of the format.

Material and methods

The study population consisted of obstetrics and gynecology residents at the Medical College of Wisconsin Affiliated Hospitals. Residents were assigned an electronic online teaching module on the topic of E&M coding and documentation in the outpatient setting. The author and the departmental professional billing staff composed the learning module. It was conducted using the ANGEL electronic course management system (CyberLearning Labs, Indianapolis, IN) that is in use at the Medical College of Wisconsin. The ANGEL platform allows secure access from any Internet-enabled computer. Users can log in at the time of their choosing and work through the module at their own pace in 1 or multiple sittings.

The exercise consisted of a pretest, the learning module, a posttest, and a user survey. The pretest consisted of 10 actual outpatient notes, each from a different patient (Figure). Professional full-time billing personnel in the Department of Obstetrics and Gynecology had previously abstracted the notes to determine the correct level of service. The notes were from a mix of new and established patients who had varying levels of service provided for gynecological complaints. Notes were sanitized to delete patient name and any other Health Insurance Portability and Accountability Act-defined patient identifiers. Residents were asked to determine the level of service provided based on the documentation in the note.
The pretest was scored in 2 ways. The first way consisted of a simple percent correct. The second consisted of a point system based on 2004 Medicare reimbursement schedules (Table I). Using this system, a correct answer earned points equal to the dollar amount reimbursed by Medicare for that level of service, for a maximum score of 709.81 (Table II). An incorrect response caused a subtraction in points equal to the absolute difference in the reimbursement between the correct level and the incorrectly answered level. For example, if the documentation for the visit supported a new patient, level 3 code, but the resident responded new patient, level 2, their total score would be debited 38.47 points (absolute value of the difference in reimbursement, 130.93 – 92.46). The rationale for derivation of the point system was based on the concept that with this subject matter, simple scoring of items correct did not adequately assess either the module itself or the learner.

With E&M coding, not all errors are created equal. If a learner abstracted a chart note incorrectly and coded an “established patient, level 2” because they misunderstood what constitutes a complete review of systems, it represents a relatively minor error. However, if several errors were made in recognizing the levels of history, examination, and medical decision-making components or did not recognize that the patient was established, the errors compound and should be graded more harshly. Using the point system based on reimbursement schedules, larger errors would be penalized more than small ones. Medicare reimbursement schedules were used because of their relatively universal use nationwide to aid in applicability and interpretation.

The learning module consisted of 7 short (1 to 3 pages) sections of informational material that could not be accessed until completion of the pretest. Most sections contained 1 to 2 review questions with explanations of correct and incorrect answers. These questions were used for review to highlight important information and were ungraded. Primarily, however, they served as internal checkpoints for the ANGEL platform. Residents could not access a subsequent section until the questions at the end of the current section were completed.

At the conclusion of the module, a posttest was administered identical in content to the pretest with the item order scrambled. The posttest could not be accessed until all sections of the module were completed. The scoring of the posttest was likewise identical to the pretest. In addition, a short survey was conducted at the end of the exercise to obtain information on resident attitudes and opinions on the exercise. The effectiveness of the learning module was measured by assessing the improvement (or difference) in resident performance on the pretest versus the posttest.

This study was considered exempt from review by the Medical College of Wisconsin Institutional Review Board. All residents in the program were asked to complete the module. However, data were analyzed only for those who consented to participate in the study.

### Statistical analysis

The difference in percent correct and points scored was analyzed using the sign test for median. This is a nonparametric test, which makes no assumptions about the data having a normal distribution. Analyses were performed with Minitab statistical software (Minitab, Inc, State College, PA).

### Results

Twenty-one residents completed the online module. Of these, 19 consented to have their data used for analysis. The mean score on the pretest was 44.2% (range 10% to 70%) using percent correct scoring and 513.44 (range 274.00 to 620.89) using point scoring. The mean score on the posttest was 45.3% using percent correct scoring and 555.87 (range 409.74 to 633.36) using point scoring. Only 47.4% of residents (9 of 19) improved their performance after completion of the module when assessed by percent correct. However, when evaluated using the point score, 84.2% of residents (16 of 19) improved their performance. Using percent correct score, the mean score on the posttest improved 1.1% ($P = 1.0$), whereas using the point score the mean score improved 42.43 ($P = .02$).
With regard to the types of errors committed, there were differences noted between the pretest and posttest (which resulted in the improved point score on the posttest). Among all learners, equal numbers of errors (20) were made on both the pretest and the posttest with regard to misclassifying a new patient as an established one or vice versa. Errors in the level of service coded were significantly different between pretest and posttest, however. Of the 102 level-of-service coding errors on the pretest, 62 (61%) were overcoded. In contrast, of the 93 level-of-service errors on the posttest, 61 (66%) were undercoded. More importantly, 35 (34%) of these errors on the pretest were miscoded by 2 or more levels of service. On the posttest, only 11 level-of-service errors (12%) were miscoded by 2 levels and none by greater than 2.

A resident acceptance survey was conducted at the completion of the module. All 19 subjects completed the survey for a 100% response rate. Sixteen subjects (84%) felt that the module had met the objective of improving their familiarity with E&M documentation and coding. Eleven (58%) felt the electronic format was superior to a traditional lecture, with an additional 5 (26%) indicating that the 2 formats were equivalent. Two respondents (11%) felt the lecture was superior, and 1 (5%) answered “not sure.”

With regard to the actual system used (the ANGEL course management system), 15 (79%) thought the platform was worthwhile as a learning tool, although 8 of the 15 indicated that the subjects able to be taught with it were limited. Three respondents (16%) felt it was “probably not” worthwhile, and 1 (5%) was “not sure.” Nine of 19 (47%) felt that homework assignments to complete on their own time were appropriate in resident education, with only 2 indicating that they strongly agreed with their use. Six of the 19 (32%) were neutral, whereas 3 (16%) disagreed with their use and 1 (5%) strongly disagreed. The total time reported as necessary to complete the instructional module alone averaged 32 minutes (range 15 to 65 minutes, median 35 minutes), whereas the entire exercise including pretest and posttest averaged 56 minutes (range 15 to 120 minutes, median 60 minutes).

Comment

Understanding proper coding and documentation helps to develop practice habits that save time, ensure proper reimbursement, and protect both from fraud allegations and medical malpractice claims. To the practicing physician, the importance of understanding E&M coding and documentation is obvious. Despite its importance, physicians commonly report insecurity in their own competence in this area. The literature bears this out, showing marked discrepancies between levels of service coded versus that documented.

The inherent difficulty in learning this system is compounded when the learner is a resident. Not only is the system perceived to be confusing and complicated, but the relevance to the physician in training can be difficult to grasp. Recently the Accreditation Council for Graduate Medical Education put in place competency guidelines for the training of all residents. One of the core areas that all physicians should be competent in is systems-based practice. This competency requires residents to “demonstrate an awareness of and responsiveness to the larger context and system of health care and the ability to effectively call on system resources to provide care that is of optimal value.” Understanding and applying E&M coding would be an important component of this competency area. Despite this mandate and the obvious importance in their future practice of medicine, there is a dearth of literature about resident performance, comprehension, or education concerning E&M coding and documentation. Ng and Lawless found that pediatric residents miscoded problems-oriented visits almost two thirds of the time, usually by undercoding. Interestingly, the year of training had no impact on accuracy in coding.

Although our program has attempted to consistently instill the importance of proper documentation and the fundamentals of proper coding, residents have universally reported feeling less than competent in the area. We have traditionally taught this subject in didactic sessions, with learners, faculty, and billing instructors all unsatisfied at the result. It was felt that a self-paced learning module that demanded interaction and objectively graded individual performance would be a better educational experience in the teaching of this subject. Although its preparation is slightly more labor intensive than preparation of a lecture, it was felt to be more effective than a lecture format and much more efficient than multiple small group or individualized sessions for residents with the billing staff. With recent limitations on duty hours and constant expansion of the skills and knowledge necessary for residents to learn in training, time is arguably the most valuable asset in medical education. The ability to assign educational material to be completed off duty is therefore also attractive.

The effectiveness of the exercise was to be judged by improvement in learner performance on a set of abstraction exercises. It was felt that a simple scoring of right answer versus wrong answer was inadequate for this subject. The scoring system detailed in Material and Methods was devised to overcome this problem. Because not all E&M coding errors are equivalent, the system allowed greater penalty for answers that were farther off the mark than others.

It was hypothesized that regardless of percent correct on posttest versus pretest, learner performance would nonetheless improve. Put another way, even if mistakes were still made, they would be of smaller magnitude.
Obviously the goal is objectively correct coding without any errors at all, but small errors are more desirable than large ones. The results of the data seem to confirm this hypothesis. Although the overall change in percent correct was negligible, there was a significant improvement in point score on the posttest versus the pretest. When judged by their percent correct, approximately half (47%) of the subjects actually performed more poorly on the posttest versus the pretest. However, two thirds of this cohort still improved their performance overall using the point scoring. Although residents still answered many of the items incorrectly, they answered “smarter” and their mistakes were less significant. This is confirmed again in considering how many errors were off by only a single level of service on the posttest (88%) versus the pretest (66%).

The survey results indicated favorable acceptance of the electronic module as a teaching tool for the subject matter. The majority of subjects (84%) felt the learning objectives were met, and an identical number felt the format to be equal or superior to the lecture format that each had experienced several months earlier in the academic year. The entire exercise took no longer (on average) than a traditional lecture, but this included pretest and posttest activities that would not necessarily be included in future applications of this exercise. The actual learning module alone took slightly more than half as long as a traditional lecture. A majority of subjects (79%) likewise felt that the ANGEL platform would be worthwhile for other subjects. Interestingly, a minority of resident subjects (47%) indicated agreement with the concept of homework assignments in graduate medical education, with only 2 (11%) agreeing strongly.

There are a number of potential flaws and criticisms to this study. First and foremost is the actual method of assigning points based on reimbursement schedules. The nature of the reimbursement schedules is such that the interval differences in dollars (or points) between consecutive levels of service are not equal. If the correct answer to a given question should have been “established patient, level 3,” then that answer would have earned 50.59 points. Consider 2 subjects who make similar mistakes. The first answers “established patient, level 2,” but the second answers “established patient, level 4.” They have both made essentially identical errors in the spirit of this assessment: They are off by a single level of service and have correctly identified the patient as established. However, because they are penalized the difference in reimbursement, the first subject is penalized only 14.41 points, whereas the second is penalized 28.52 points.

Consideration was given to standardizing the penalty, such that each level of service had consistent difference between them or assigning a set penalty (eg, 50 points) for each level of service by which an answer was mistaken. The point system remained as is, however, because it was felt that it would provide a more tangible sense of the consequence of miscoding if it was tied directly to actual reimbursement schedules.

Comparing the performance of subjects after completing the module with a control group that receives a didactic lecture was not feasible for this study because of the small overall sample size, but future studies should consider this. It would be interesting to determine whether resident level (postgraduate year) made a difference in performance, but the sample sizes in this study were again too small to make such comparisons.

Although I feel that the assessment using this point system indicates that the learning module was effective, these results need to be validated. Validated assessment will come with determining whether coding and documentation actually improve in the clinical setting after participation in the exercise.

Acknowledgment

I thank Donna Kroening, CCS, for her invaluable help in the preparation of the E&M coding module as well as William Herbert, MD, and the rest of the faculty of the Association of Professors of Gynecology and Obstetrics/Solvay Educational Scholars Development Program for substantive guidance and advice.

References

5. The Accreditation Council for Graduate Medical Education. ACGME Outcome Project General Competencies. Available at: www.acgme.org/outcome/comp/compfull.asp.
Teaching residents coding and documentation: Effectiveness of a problem-oriented approach

Sawsan As-Sanie, MD, MPH,a,b Denniz Zolnoun, MD, MPH,a Mary Ellen Wechter, MD,a Georgine Lamvu, MD, MPH,a Frank Tu, MD, MPH,a,c John Steege, MDa

Objective: We sought to assess the effectiveness of a problem-oriented approach to teaching residents accurate coding and documentation of ambulatory gynecology visits.

Study design: This was a pilot before-and-after study. Nine resident volunteers underwent 4 individual instructional sessions on coding and documentation with a trained faculty member over 6 weeks. Outcomes were assessed by comparing the appropriateness of procedure and diagnostic codes billed in participant continuity clinic prior to and in the 6 to 9 months following the intervention.

Results: Following the intervention, participants demonstrated an increase in the accuracy of coding the correct category of the evaluation and management service, an increase in the appropriate use of modifiers, and a decline in undercoding errors.

Conclusion: Problem-oriented interactive learning appears to be an effective method of teaching residents proper coding and documentation.

Accurate and consistent coding and documentation of outpatient office visits is the foundation of operating a compliant and productive office practice. Unfortunately, many US residents in obstetrics and gynecology, as well as those in other subspecialties, graduate with a limited understanding of proper coding and documentation.1,2

To address this ubiquitous problem, the Accreditation Council for Graduate Medical Education includes system-based education as 1 of 6 general competencies required for resident development.3 System-based education expects that residents demonstrate “an awareness of and responsiveness to the larger context and system of health care.” As such, residents should be knowledgeable about coding, reimbursement, and the management of a medical practice. Identification of these general competencies was the first step in the Accreditation Council for Graduate Medical Education’s long-term effort to emphasize outcome assessment in resident education and the accreditation process.

Despite the obvious need for an effective education paradigm, there are no published guidelines on teaching or assessing the competency of coding and

KEY WORDS
Resident education
Coding
Documentation
Practice management
documentation in an ambulatory gynecology setting. Thus, the objectives of this study were: (1) to develop a problem-oriented approach to teaching proper coding and documentation of ambulatory gynecology visits; and (2) to assess the feasibility and effectiveness of such an approach by evaluating changes in documentation competencies and billing patterns.

Material and methods

This was an institutional review board approved pilot before-and-after study conducted at the University of North Carolina (UNC) Department of Obstetrics and Gynecology between January 2003 and February 2004. Nine obstetrics and gynecology residents, postgraduate years (PGY) 2 to 4, were asked and consented to participate in this study. Each resident attended 4 individual instructional sessions with 1 of 4 trained faculty members over a 6-week period.

Each session was an individual 1-hour learner-based module that was based on a standard curriculum, drawing examples of successes and errors from that individual resident’s dictations in the prior 3 months. The curriculum was designed to address the most common coding and documentation errors identified in a sample of resident dictations audited prior to the educational intervention. Prior to meeting with the residents, each participating faculty member underwent the hospital’s mandatory coding and documentation course and an additional 1-hour training course, which reviewed the study’s curriculum.

Session 1 provided an overview of Medicare standards for coding and documentation. This focused on the usage of Current Procedural Terminology (CPT) Evaluation and Management (E/M) codes and their linkage to the International Classification of Diseases (9th revision) diagnostic codes. The medical history, physical exam requirements, and complexity of medical decision making for each level of service provided were reviewed. The appropriate use of modifiers-24, -25, and -51, which are used to assure proper reimbursement of a payable service provided, was highlighted in this session. For example, modifier-24 is used when providing an unrelated E/M service in the postoperative period of a major surgical procedure. Modifier-25 is used when the same physician performs a significant and separately identifiable E/M service on the same day as a minor procedure or a preventive health maintenance examination. Modifier-51 is used when multiple but separate procedures are performed on the same visit.

Session 2 focused on the differences between the 1995 and 1997 Medicare guidelines for the documentation of the physical examination. During this encounter, the faculty member reviewed the 1995 requirements of a multiorgan approach as well as the 1997 allowance for a focused genitourinary examination. This session also provided an overview of 2 methods of choosing the level of service: billing according to evaluation and complexity versus billing for time-based counseling.

Session 3 reviewed the recommended components of an annual health maintenance examination, based on the American College of Obstetricians and Gynecologists guidelines on preventive care. The emphasis was on women ages 18 to 39 years, the most prevalent age group in the resident continuity clinic. A discussion of the North Carolina preventable causes of morbidity and mortality served as a reminder for residents to look beyond the breast examination and Papanicolaou smear when providing preventive care. Individual resident dictations were then reviewed on the appropriateness of their coding and documentation for an annual health maintenance visit.

Finally, session 4 was used as an opportunity to review 5 audited resident dictations for errors defined in the first 3 sessions. Residents were asked to provide feedback on their dictations and coding and received case-specific guidance from the instructor. At the end of the session, a posttest was administered, followed by immediate feedback on the answers.

Coding and documentation competencies were the primary outcome evaluated. Four general categories of competencies were developed based on the UNC Teaching Physicians Oversight Committee guidelines for computing a compliance error rate: class 1 (overcoding) errors, class 2 (undercoding) errors, appropriate use of modifier-25, and correct coding of a preventive medicine visit.

Class 1 errors are overcoding errors that cannot be billed as reported by the physician. The most common class 1 errors are category errors in which the wrong category of visit is chosen and overcoding by 1 or more levels of service. These errors are considered the most egregious because they bill for a higher fee than the documentation supports. For example, billing CPT code 99385 (preventive medicine, new patient 18-39 years old) when the appropriate code is CPT 99395 (preventive medicine, established patient 18-39 years old) would be considered a class 1 error.

Class 2 errors are undercoding errors that may be billed as coded but are also serious because they generally undercode, causing lost revenues. For example, billing CPT code 99213 (expanded problem-focused visit, established patient) when the documentation supports the level CPT 99214 (detailed problem-focused visit, established patient) is a class 2 error.

Coding competencies were assessed according to standard Medicare guidelines. A UNC billing compliance auditor randomly selected 10 outpatient gynecology visits provided by study participants: 5 dictations in the 3 months prior to the intervention and 5 dictations between 6 and 9 months following the completion of the intervention. To assure adequate diversity in the selected
dictations, 2 of the 5 dictations in each time period were selected from preventive medicine visits. The remaining 3 were problem-oriented visits. Within these 2 categories, dictations were randomly selected from the appropriate time period. Each dictation was audited to identify errors in the coding competencies of interest. Also, a billing and coding database was queried for all outpatient gynecology visits provided by study residents in the same period. Diversity in levels of service coded as well as appropriate coding of annual health maintenance visits were examined using this database. Descriptive and bivariate statistics, including the Wilcoxon signed-rank test, were performed using STATA 8.0 (STATA Corp, College Station, TX).

**Results**

Eleven residents volunteered to participate in this study; 2 residents were excluded because they completed the educational sessions within 1 month of graduation. Among the remaining 9 residents, 5 were PGY2 and 4 were PGY3. The average postintervention examination score was 94% correct.

After the educational intervention, residents showed a decline in overcoding errors, primarily because of an improvement in assigning the correct category of service rendered (Table). Nearly a third of audited dictations prior to the intervention were coded in the wrong category of service; this declined to 13% following the intervention ($P = .05$). There was also a trend toward a decrease in coding errors that result in underbilling, including meaningful improvements in billing for procedures provided and increases in the appropriate use of modifier-25. For example, 2 of 4 procedures (50%) performed in the preintervention period were not billed; this declined, although not statistically significant, to a failure to bill for 3 of 8 procedures (38%) in the post-intervention period.

Based on the preintervention audit of resident dictations, the most common error when coding an annual health maintenance exam was linking the International Classification of Diseases (9th revision) code V72.3 (annual visit with Papanicolaou smear) to a problem-oriented CPT E/M code (99201-99205 or 99212-99214) rather than to a preventive medicine code (99385-99387, 99395-99397). This coding error was billed in 54% of the preventive medicine visits prior to the intervention and declined to 37% of preventive medicine visits in the 6 to 9 months after the intervention ($P = .32$).

To assure adequate diversity in the types of visits reviewed, 40% of audited dictations were preventive medicine visits; the remainder was problem-oriented visits. Among the problem-oriented visits, the most common diagnoses were pelvic pain (30%), abnormal uterine bleeding (22%), uterine fibroids (13%), contraception (9%), and first-trimester abortion (9%). There was no relationship between the primary diagnosis and frequency of coding error.

The diversity of E/M levels of service coded appeared to improve following the intervention. For example, among all outpatient problem-oriented gynecology visits provided by participating residents in the study periods, the use of upper-level E/M codes (99204, 99214, 99244) increased from 11.7% to 20.6% of visits ($P = .05$). High-level E/M codes (99205, 99215, 99245) also increased, from 0% to 4% ($P = .03$), whereas use of low-level E/M codes (99201, 99202, 99211, 99212, 99241, 99242) declined from 27.9% to 15.3% ($P = .21$). Reflecting this change in the diversity of E/M levels of service billed, the average charge billed per encounter rose from $109.48 to $120.29.

**Comment**

Despite the importance of mastering accurate coding and documentation, there are currently no accepted standards

<table>
<thead>
<tr>
<th>Coding Error</th>
<th>Preintervention</th>
<th>Postintervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1 errors (unbillable errors)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overcoded due to wrong category billed</td>
<td>14/45 (31.1)</td>
<td>6/45 (13.3)</td>
</tr>
<tr>
<td>Overcoded by 2 or more levels of service</td>
<td>0/45 (0)</td>
<td>0/45 (0)</td>
</tr>
<tr>
<td>Overcoded by 1 level of service</td>
<td>2/45 (4.4)</td>
<td>4/45 (8.9)</td>
</tr>
<tr>
<td>Class 2 errors (billable errors)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consult opportunity missed</td>
<td>0/1 (0)</td>
<td>0/3 (0)</td>
</tr>
<tr>
<td>Undercoded by 1 level of service</td>
<td>6/45 (13.3)</td>
<td>2/45 (4.4)</td>
</tr>
<tr>
<td>Undercoded by 2 or more levels of service</td>
<td>0/45 (0)</td>
<td>0/45 (0)</td>
</tr>
<tr>
<td>Procedure not billed</td>
<td>2/4 (50.0)</td>
<td>3/8 (37.5)</td>
</tr>
<tr>
<td>Modifier-25 omitted</td>
<td>4/5 (80.0)</td>
<td>6/12 (50.0)</td>
</tr>
</tbody>
</table>

* Total number of errors among all residents per opportunity to make error.
† Wilcoxon signed rank test.

| | Preintervention | Postintervention |
| | No. errors* | % | No. errors* | % | $P$ value |
| Class 1 errors (unbillable errors) | | |
| Overcoded due to wrong category billed | 14/45 | 31.1 | 6/45 | 13.3 | .05 |
| Overcoded by 2 or more levels of service | 0/45 | 0 | 0/45 | 0 | — |
| Overcoded by 1 level of service | 2/45 | 4.4 | 4/45 | 8.9 | .16 |
| Class 2 errors (billable errors) | | |
| Consult opportunity missed | 0/1 | 0 | 0/3 | 0 | — |
| Undercoded by 1 level of service | 6/45 | 13.3 | 2/45 | 4.4 | .16 |
| Undercoded by 2 or more levels of service | 0/45 | 0 | 0/45 | 0 | — |
| Procedure not billed | 2/4 | 50.0 | 3/8 | 37.5 | .93 |
| Modifier-25 omitted | 4/5 | 80.0 | 6/12 | 50.0 | .39 |
for teaching these skills in residency programs. In fact, many residents in obstetrics and gynecology, as well as those in other subspecialties, graduate with a limited understanding of proper coding and documentation. In a survey of obstetrics and gynecology residents attending a regional research conference, only 29% of residents felt confident in coding problem-oriented visits and 43% felt confident in coding preventive care visits.2

Understanding how to correctly use modifiers was equally lacking in this group of residents: only 29% felt confident using modifier-25, 14% understood the use of modifier-24, and no residents were familiar with modifier-51. This trend is not limited to residents in obstetrics and gynecology. In a large survey of emergency medicine residents, the majority of residents rated their confidence in their ability to accurately code as “minimal” (26%) or “not at all” (42%).1

To our knowledge, this is the first study that attempts to address this deficit in resident education in obstetrics and gynecology. We found that after individual instruction, university-based residents demonstrated meaningful improvement in the knowledge and application of coding principles. Residents tended to show declines in the instances of noncompliant billing, both in assignment of E/M category and level of service provided. There was a trend toward a decline in undercoding errors as well as an improvement in billing for procedures performed. The appropriate use of modifiers to accurately bill for services provided seemed to increase, as did the appropriate coding of annual Papanicolaou smears during preventive medicine visits.

Residents in this pilot study also showed improvement in increasing the diversity of the E/M levels of service coded, which may be another marker for improved facility in coding for different levels of service. As a result, the average charges billed per patient seen also increased. Assuming correct application and documentation of the corresponding code for each visit, this suggests that trained residents demonstrate improved ability to use their coding and documentation skills to reflect the complexity of services rendered. Although this rise in average charge per patient seen was approximately $10.81, this modest amount could translate into substantial increases in billing for the same number of services provided. For example, in a residency program with 24 residents, providing 10 outpatient gynecology visits per week for 47 weeks of the year, this increased diversity in coding could result in the capture of lost revenues conservatively estimated at $120,000 per year. Indeed, this value understimates the net benefit of coding education because it does not account for other improvements in coding knowledge, such as increased use of modifiers to appropriately avoid denials of claims.

It is important to note that this training program involved a small number of participating residents, and the number of audited dictations in this study was few. Thus, we had limited ability to detect statistically significant improvements. Also, the follow-up was relatively short and the 5 dictations may have been uncharacteristic of the resident or patient population as a whole. In such a small study, results cannot necessarily be generalized to other resident or patient populations. The participating residents were volunteers and may have represented a particularly motivated sample of physicians committed to improving their coding knowledge. The long-term effectiveness of this training program should be evaluated in future studies.

Although there was substantial time investment in face-to-face teaching sessions with faculty instructors on 4 separate occasions, preliminary results of our previous work teaching a 1-hour problem-oriented module suggest that this approach may be effective in improving knowledge of coding and documentation.2 Thus, this study, in addition to our previous work, is valuable in that it demonstrates the feasibility of improving resident coding skills. It also illustrates the potential valuable financial gains that improved coding could have for the department and for the individual physician.

Today in a climate of escalating costs, falling reimbursement, and legal scrutiny, skills for proper coding and documentation are vital to the financial endurance of health care providers and institutions. Indeed, coding skills are becoming as essential to the success of the graduating resident as any other skill acquired in residency. With cooperation and communication among teaching programs, successful training in coding skills should be rapidly developed and implemented in obstetrics and gynecology training programs.

Acknowledgment

The authors acknowledge Dr Robert Cefalo, Ms Heather Scott, and Ms Christy Davis for their contribution to the design and execution of this study.

References

Implementation and evaluation of a genetics curriculum to improve obstetrician-gynecologist residents’ knowledge and skills in genetic diagnosis and counseling

Charles J. Macri, MD, Nancy D. Gaba, MD, Lauren M. Sitzer, MD, Lisa Freese, MS, Susanne L. Bathgate, MD, John W. Larsen Jr, MD

Department of Obstetrics and Gynecology and the Wilson Genetics Center, The George Washington University School of Medicine and Health Sciences, Washington, DC

Received for publication February 25, 2005; revised June 29, 2005; accepted August 1, 2005

KEY WORDS
Genetics curriculum
Adult learners
Standardized patients
Obstetrician-gynecologist resident education
Web-based education
Objective Structured Clinical Examination

Objective: This study was undertaken to develop, implement, and evaluate a genetics curriculum for obstetrician-gynecologist residents.

Study design: We prospectively evaluated the effect of a genetics curriculum on obstetrician-gynecologist residents’ knowledge and skills. Residents completed a needs assessment and pretest. Educational intervention included 2 3-hour didactic sessions with 1 hour of lecture followed by case discussion and 1 3-hour session of experiential learning using standardized patients who evaluated residents’ knowledge and skills in taking family history, drawing genetic pedigrees, and counseling patients. Posttest scores were compared with pretest scores.

Results: Needs assessment was completed by all 40 obstetrics and gynecology residents and identified limited and variable genetics education in medical school. Twenty-eight of 40 residents attended the entire educational intervention and completed the pretest and posttest, and 25 of 28 showed improved test scores. Residents stated that they were more confident in their ability to take a family history, record a 3-generation pedigree, and counsel patients about genetic conditions after completion of the genetics curriculum.

Conclusion: This multifaceted genetics curriculum improved residents’ knowledge of genetics as well as their confidence in applying genetic concepts as assessed by the pretest and posttest and by their comments in the debrief session.

© 2005 Mosby, Inc. All rights reserved.
genetics and genetic tests was increasing among physicians, particularly among recent graduates and physicians who are exposed to patients with genetic problems in their practices but that deficiencies still remain. This group stated that although “medical school courses in genetics may improve knowledge, it was not sufficient.” Rather, they determined that greater emphasis is needed at all levels of medical education.

Obstetrician-gynecologist (OBGYN) residents are expected to have a broad working knowledge of genetics to provide patients with accurate and up-to-date care. This depends on an understanding of preconception and prenatal risks; family history and cancer risks; and disease risk for adult-onset disorders such as osteoporosis, diabetes, hypertension, and stroke. OBGYN residents are also expected to teach medical students, physician assistant students, and junior residents. A resident spends 20% to 25% of the average week at work supervising, evaluating, or teaching others.

Most OBGYN residents have limited experience with genetics because of fragmented medical school curricula, rapidly changing knowledge about genetics, and the previous view of genetics as pertaining only to rare disorders. Curriculum change in OBGYN requires evaluation of the learners, the teachers, and the program as noted by Metheny et al. Specifically, the “purpose of the evaluation is to determine the impact of the curriculum on and to evaluate the learner.” We have incorporated several evaluation techniques in our genetics curriculum for OBGYN residents including multiple-choice question examinations, direct observation with checklists, and objective structured clinical examination (OSCE).

### Background

In 2001, a genetics curriculum for OBGYN residents was presented by one of the authors (C.J.M.) as an Association of Professors of Gynecology and Obstetrics (APGO) workshop. This proposed curriculum included genetic counseling, family history, 3-generation pedigrees, multifactorial inheritance, mendelian and nonmendelian inheritance, chromosome abnormalities, single-gene disorders, population genetics, screening for at-risk populations, cancer genetics, molecular diagnostics, and prenatal screening and testing. Prior to the implementation of this study, OBGYN residents at this institution learned genetics by individual study and periodic lectures about genetics topics and as part of clinical experience.

The purpose of this study was to test the effectiveness of a structured multifaceted curriculum on the test scores of OBGYN residents at 1 large, urban, university-based residency program. We used the core competencies for genetic knowledge, skills, and attitudes for all health care providers established by the National Coalition for Health Professional Education in Genetics to develop the learning objectives. Although not every patient or family will require chromosome analysis or molecular testing, it has been noted that important information about disease risk may be identified from the family history, especially when recorded in the 3-generation pedigree format using standard pedigree nomenclature.

### Standardized patients

A standardized patient (SP) experience was used to help teach and evaluate the OBGYN residents’ knowledge and skills in taking a family history and drawing a 3-generation pedigree. The use of SPs in medical education dates back to the 1960s. Standardized patients are trained to portray a specific patient role and to report objectively on what takes place in the encounter. Furthermore, the use of SPs and simulations can provide a means by which defined performance criteria can immediately be evaluated by faculty and peers to permit feedback and practice in a realistic but less threatening educational environment.

In 1982 Laube et al reported on the effectiveness of a clinical skills instruction program for learning about the acute abdomen.

It has been demonstrated that OSCEs are psychometrically stable and valid as evaluation tools. Because of their validity and reliability, international and national licensure boards have adopted OSCEs as a method to evaluate clinical skills of candidates. In 2003 the National Board of Medical Examiners initiated the use of SPs and a multiple OSCE as part of the step 2 clinical skills examination. Because of the usefulness of these methods of instruction and evaluation in education of medical students and residents and the importance of the step 2 clinical skills examination as part of the licensing examination, a clinical learning and simulated skills center was established at the George Washington University School of Medicine and Health Sciences.

### Material and methods

This educational research project was approved by the Institutional Review Board at the George Washington University School of Medicine and Health Sciences (GWU). The curriculum presented at APGO was adjusted to fit into the OBGYN competency skills program at GWU, and a needs assessment survey was conducted addressing demographic information, prior educational experience in genetics, advanced degrees, and a Likert rating scale (ranging from 1 to 5) of residents’ knowledge and comfort level about genetic diagnoses and testing.

The faculty consisted of medical geneticists, a genetic counselor, maternal-fetal medicine specialists, and general obstetrician/gynecologists.

Following completion of the needs assessment, all residents completed a 40-item, multiple choice question
pretest with questions modified from standard review sources in OBGYN and medical genetics. These questions assessed knowledge about genetic issues in preconception counseling, prenatal diagnosis, spontaneous abortion, and recurrent pregnancy loss as well as molecular genetics and DNA diagnostic techniques. Pretest scores were recorded for each resident completing this exercise.

Learning objectives were provided to each resident. The educational intervention included 2 3-hour didactic sessions and 1 3-hour session for history taking and pedigree drawing using SPs in the GWU clinical learning and simulated skills center. The content of the first didactic session included lecture-based topics (introduction to mendelian inheritance [20 minutes], family history and genetic counseling [20 minutes], and pedigree analysis and construction [20 minutes]).

This was followed by small-group, case-based sessions, in which the residents were provided cases illustrating the concepts covered in the lectures, including pedigree analysis for autosomal dominant disorders (breast cancer, BRCA1), autosomal recessive disorder (cystic fibrosis), and a population-based genetic risk assessment (sickle cell anemia). The small groups were provided with a written case, learning objectives, resource materials, and a series of questions to answer. Each group had a faculty member who served as a facilitator. The group members selected a spokesperson to keep notes and present their groups’ questions and answers to the entire group in the third hour. Residents were encouraged to use online medical genetics resources when needed. We created an informational Web site (www.gwumc.edu/edu/obgyn/genetics) at our institution to facilitate standardized access to information by all residents and students.²¹⁻²³

The second session focused on advanced topics in medical genetics including nonmendelian inheritance (mitochondrial inheritance, imprinting, uniparental disomy, and mosaicism), chromosomal disorders (trisomy 21, 18, and 13, monosomy X), and screening in at-risk populations (Tay Sachs disease, sickle cell anemia, thalassemias.) The cases in this second session included a family history of mental retardation consistent with Fragile X syndrome (triplet repeat disorder), a baby born with Prader-Willi syndrome (imprinting and uniparental disomy), and Turner mosaicism.

The third session included an experiential learning environment using SPs who presented cases in medical genetics to the residents. The residents were expected to take a family history, draw a pedigree, and counsel the patient about the particular genetic condition. Teaching cases included patients with elevated maternal serum alpha fetoprotein level in the second trimester, breast cancer risk with 3 family members with premenopausal breast cancer in an Ashkenazi Jewish family, and cystic fibrosis screening in a preconception counseling patient visit. Both residents and faculty observed the residents counseling the SPs and used a checklist to evaluate and provide feedback about their history taking and counseling. The SP gave feedback to the residents about their professionalism, history taking, and counseling.

This session was followed by a half-hour debrief session about the genetics curriculum and the SP experience. After the educational intervention, the residents completed the 40-question posttest. Posttest scores were compared with pretest scores using paired Student t-test.

Results

Forty residents completed the needs assessment and 28 completed the pretest, educational intervention, posttest, and SP exercise.

Needs assessment

Forty residents completed the needs assessment survey. Prior to the educational intervention, 33 of 40 residents stated they were “not comfortable” identifying medical genetics resources on the Internet and had limited experience using Web-based resources such as On-line Mendelian Inheritance in Man. Thirty-five of 40 stated that they were “not comfortable” taking a detailed family history, recording a 3-generation pedigree, and counseling patients about genetic conditions.

Exam scores

Of the 28 residents who completed all parts of the education intervention, 25 of the 28 residents improved on the posttest. The pretest and the posttest included the same 40 questions. Of the 28 residents, the pretest scores ranged from 20 of 40 to 32 of 40 correct answers, with a mean of 26.89, and the posttest scores ranged from 23 of 40 to 38 of 40 correct answers, with a mean of 32.00, which was statistically significant (P = .001).

Standardized patient session

Feedback was given informally from SPs and faculty to residents. The evaluation aspect of this process is under development and will be included in future projects.

Debrief session

The evaluation of this education program in genetics by the residents included both verbal feedback and written comments in the debrief session. Most residents rated the educational experience highly and stated that they felt more comfortable taking a detailed family history, recording a 3-generation pedigree, identifying Web-based resources for genetic conditions, and counseling patients after completion of the program.

After talking with SPs, the residents expressed a better understanding of what questions might arise with
regard to family members and management decisions. Interacting with SPs allowed the residents to practice formulating a mental script to use with actual patients in the future.

In the needs assessment, some residents indicated that role playing and small group discussion would not be helpful to improve their knowledge and understanding of genetics topics; however, after participating in the curriculum, they acknowledged the value of the small group discussions and SPs.

Conclusions

The genetics curriculum improved residents’ knowledge of genetics as assessed by the pretest and posttest. Residents expressed more confidence in their ability to take a detailed family history, record a 3-generation pedigree, use Web-based resources for genetic conditions, and counsel patients after completion of the program.

Comment

The contemporary practice of medicine requires health care providers to have knowledge about genetics and possess the skills necessary to obtain, interpret, and communicate information about genetics. The importance of hereditary contribution to disease is often underappreciated by the general public and many health care providers.

Through training, primary care providers including obstetrician-gynecologists and OBGYN residents can be prepared to take a role in advocating for patient needs in the area of genetics. Current medical education programs using lecture format may not be the most effective method for providing health care personnel with the requisite knowledge and skills in genetics. Innovative educational modalities, experiential in nature, including SPs, may better serve to educate adult learners.

Acknowledgment

For assistance with statistical analysis of pretest and posttest scores, you may contact Dr Quannetta T. Edwards, FNP, WHCNP, Chair, Graduate Nursing Program and Coordinator, Family Nurse Practitioner Program, Howard University, College of Pharmacy, Nursing and Allied Health Sciences, Division of Nursing, Washington, DC.

References

Objective: The objective of the study was to identify factors associated with satisfaction in the Indiana University School of Medicine Obstetrics-Gynecology residency program and the residents’ and faculty’s perception of whether these factors were extant in our program.

Study design: Residents and faculty at the Indiana University School of Medicine Obstetrics-Gynecology program were surveyed using an instrument based on prior primary care specialty investigations. Multivariate regression evaluated the impact of various factors on resident satisfaction.

Results: Seventy-seven percent (44 of 57) and 100% (35 of 35) of faculty and residents, respectively, completed the survey. Relevant training, collegiality, adequate resources, workload, care continuity, supportive coworkers, learning environment, autonomy, role ambiguity, and supportive faculty were significantly associated with resident satisfaction. Care continuity, role ambiguity, and learning environment were the areas of largest faculty/resident disagreement.

Conclusion: Relevant training and collegiality were most strongly linked to resident satisfaction. Three areas of dissatisfaction were identified, and we will seek to remedy these areas.

The percentage of US medical students choosing obstetrics and gynecology has decreased from its peak of 7.5% in 1993 to 5.5% in 2004. In addition, for 2004, 16% of the 245 obstetrics and gynecology programs went unfilled through the National Residency Matching Program. Student clerkship dissatisfaction, lifestyle issues, gender issues, and high professional liability premiums have been cited as causes for this decline in student interest in obstetrics-gynecology. Residents play an active role in the education of medical students during their obstetrics-gynecology clerkship.

One potential reason for student clerkship dissatisfaction resulting from resident-student interactions may be that residents themselves are unhappy or dissatisfied. Evaluations of residents’ satisfaction with their residency program have been conducted in internal medicine and family practice. Factors associated with higher satisfaction were those that enhanced learning, such as interaction with attending physicians and participating in patient rounds and seminars. Although residents noted they received more of their learning from other residents, the contact they had with attending physicians was most predictive of satisfaction. Additional factors found to affect satisfaction included measures of continuity of care, autonomy, collegiality, work that encourages professional growth, work group loyalty, and role conflict.
Our primary objective was to assess the impact of these previously identified factors on obstetrics and gynecology residents’ satisfaction. Because contact with faculty plays a major role in resident satisfaction,7 we postulated that a disparity between residents’ and the faculty’s perceptions might serve as a guide to focusing residency program improvement efforts. Hence, our secondary objective was to determine the degree of agreement between faculty and the residents as to whether the satisfying factors were present within our program.

Materials and methods

Residents and faculty involved with the Indiana University School of Medicine Obstetrics-Gynecology program were asked to complete a survey instrument on the ANGEL Web-based system (http://www.angellearning.com; Indianapolis, IN). ANGEL is an interactive Internet curriculum tool available at the study institution. The resident survey instrument was adapted from instruments developed from those used by other primary care specialties.7,8 Items were added to address issues specific to the study residency program. Residents were asked to respond on a Likert-type scale to such items as “I influence things that affect me on the job.” The faculty instrument was created by modifying the items from the resident instrument such that they were asked to respond to a similar item such as “Residents influence things that affect them on the job.” Residents and faculty were also asked to respond to items to measure demographic characteristics.

The survey items assessed several domains including: autonomy, work that encourages professional growth, continuity of care, work cohesion, workload, role ambiguity, role conflict, adequate resources, loyalty to coworkers, loyalty to the program, supportive coworkers, supportive faculty, task significance, relevant training, collegiality, and learning environment. These domain categories, with the exception of learning environment, were used in the previous studies and were found to possess psychometric reliability.7,8 We asked about satisfaction 11 different ways, 3 of which were used by Randall et al7 and considered as possible dependent variables to be used in the regression analysis. The results of the 3 approaches were similar. The dependent variable chosen for the multiple regression analysis presented was “I feel satisfied with this residency program.”

Ten regression analyses were conducted using as independent variables the residents responses to the questions used in prior studies,7,8 with additional questions that were found to be significantly related to satisfaction in this study within each of the 10 domains. Table I lists the number of questionnaire items included in the regression analyses for each of the 10 domains, the sources of the items and Cronbach’s alpha values for the group of variables.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Items, no.</th>
<th>Source(s)</th>
<th>Cronbach’s alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevant training</td>
<td>3</td>
<td>Randall et al7</td>
<td>0.765</td>
</tr>
<tr>
<td>Collegiality</td>
<td>3</td>
<td>Randall et al7</td>
<td>0.764</td>
</tr>
<tr>
<td>Adequate resources</td>
<td>3</td>
<td>Randall et al7</td>
<td>0.637</td>
</tr>
<tr>
<td>Workload</td>
<td>3</td>
<td>Randall et al7</td>
<td>0.746</td>
</tr>
<tr>
<td>Continuity of care</td>
<td>2</td>
<td>Randall et al7</td>
<td>0.490</td>
</tr>
<tr>
<td>Supportive coworkers</td>
<td>3</td>
<td>Randall et al7</td>
<td>0.748</td>
</tr>
<tr>
<td>Learning environment</td>
<td>5</td>
<td>From Seelig,8</td>
<td>0.907</td>
</tr>
<tr>
<td>Role ambiguity</td>
<td>3</td>
<td>Randall et al7</td>
<td>0.789</td>
</tr>
<tr>
<td>Autonomy</td>
<td>3</td>
<td>Randall et al7</td>
<td>0.523</td>
</tr>
<tr>
<td>Supportive faculty</td>
<td>5</td>
<td>From Randall et al7</td>
<td>0.563</td>
</tr>
</tbody>
</table>

Institutional review board approval was obtained prior to administration of the surveys. The instrument was available online for completion from January 5, 2004, through February 16, 2004. The residents and faculty were contacted via e-mail and fax with instructions indicating where the questionnaire was available. A “whodunit” agent available on ANGEL was used to identify those who had not completed the survey to whom reminder e-mails and faxes were sent.

A commonly used computer software package, SPSS version 11.5.0 (http://www.spss.com Chicago, IL) was used to conduct the data analyses. Multiple regression analysis was used to determine how strongly the factors were associated with resident satisfaction. The dependent variable used for the multiple regression analyses was the residents’ responses to the statement, “I feel satisfied with this residency program,” coded from 1 to 5 where 1 was “strongly disagree” and 5 was “strongly agree.” The Pearson χ² test was used to compare resident and faculty responses to key items in each of the 10 factors that were found to be associated with resident satisfaction. To satisfy the cell size limitation of the χ² test, the “strongly agree” and “agree” responses as well as the “disagree” and “strongly disagree” responses were combined. The neutral responses were kept as a separate category. P values less than .05 were considered statistically significant.

Results

Seventy-seven percent (44 of 57) and 100% (35 of 35) of faculty members and residents, respectively, completed the survey. Females made up 80% of the residents and 54.5% of the faculty respondents. Satisfaction with the program was noted by 68.6% of the residents, which was not significantly higher than...
the 61.4% of the faculty who believed the residents were satisfied with the program ($P = .394$). Enthusiasm for the program was reported by 60.0% of the residents, which was similar to the 56.8% of faculty members who thought the residents were enthusiastic about the program ($P = .869$). Dissatisfaction with medical profession in general was noted by 28.5% of the residents and 21.0% of the faculty. As shown in Table II, relevant training, collegiality, adequate resources, workload, continuity of care, supportive coworkers, learning environment, role ambiguity, autonomy, and supportive faculty were all significantly associated with resident satisfaction.

The agreement between the responses of the faculty and residents on the satisfaction factors is shown in Table III. There was consistent agreement between the faculty and residents that 4 of the 10 factors influencing resident satisfaction were present in our program: collegiality, supportive coworkers, manageable workload, and relevant training. Residents and faculty agreed that there was collegiality as demonstrated by an excellent amount of interaction and ideas exchanged in the study department. Supportive coworkers were thought extant because the majority of residents and faculty agreed that the residents listen to each other’s problems and help each other when needed.

One notable concern was the majority (88.6%) of residents professed that residents gossip about other residents; however, such behavior had been witnessed by only 54.5% of the faculty ($P < .005$). Although 88.2% and 85.7% of residents and faculty, respectively, thought that residents received adequate training in procedural skills; only 62.9% of residents stated that this program gave them the opportunity to learn the cognitive skills needed to successfully treat patients, whereas 88.1% of faculty believed that residents had opportunities to learn these cognitive skills ($P = .015$). A majority of both residents (85.7%) and faculty (68.2%) believed that residents had to work very fast to keep up with their work. Patient load was felt to be about right by 54.3% of the residents and 61.3% of the faculty.

There was disagreement on the other 6 factors including adequate resources, continuity of care, role ambiguity, autonomy, supportive faculty, and learning environment. However, adequate resources, continuity of care, and role ambiguity were areas in which the residents felt the factors were present in the program to a greater degree than the faculty. In the clinics, 57.1% of residents felt they have enough support services, whereas only 27.3% of faculty thought that the residents found the services to be adequate ($P = .010$). Residents overall perceived a great deal of continuity of care in our program, with 82.9% responding that they were satisfied with this aspect of their practice. The faculty may not be cognizant of the amount of continuity of care built into the program in that only 34.1% of the faculty members thought the residents were satisfied with continuity of care ($P < .005$). Nearly two thirds of residents (65.7%) stated they knew their work responsibilities, but only 34.1% of the faculty thought they did ($P = .017$). Nearly all residents (94.3%) believed they knew how to get their work done, whereas only 54.5% of the faculty agreed ($P < .005$).

The other three factors (autonomy, supportive faculty and learning environment) were perceived as lacking by the residents to a significantly greater degree than the faculty, suggesting that resident needs may not have been sufficiently met in these areas. Although the majority of residents (71.4%) reported that they can influence things that affect them on the job, nearly all (93.2%) of the faculty felt residents have this influence ($P = .011$). Three quarters of the residents (77.1%) believed they had input in deciding how patient care demands were shared, which was significantly lower than the 93.1% of the faculty who held this view ($P = .008$). Commitment to teaching by faculty was reported by 57.1% of residents, whereas 86.4% of faculty believed this commitment existed ($P = .003$). Almost half of residents (48.6%) noted that they had been spoken to inappropriately or demeaned by a faculty member more than once, whereas only 18.2% of faculty admitted to engaging in such behavior ($P = .016$). More than half of residents (62.9%) agreed that there is a staff member they felt was easily approachable with a personal or work-related problem; 95.4% of the faculty thought they could be entrusted with this role ($P = .001$).

The area of most consistent disagreement between the residents and the faculty was the learning environment. Only half of the residents (51.4%) felt that resident education is a high priority of this program, whereas 90.9% of faculty felt this way ($P < .005$). Although three quarters of faculty (77.3%) felt they are teaching based on current medical literature only half of residents (47.1%) felt this was the case ($P < .005$). Appropriate and timely feedback from faculty was perceived to be available by 34.3% of residents, whereas 65.9% of

<table>
<thead>
<tr>
<th>Table II: Impact of Specific Factors on Resident Satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor</td>
</tr>
<tr>
<td>Relevant training</td>
</tr>
<tr>
<td>Collegiality</td>
</tr>
<tr>
<td>Adequate resources</td>
</tr>
<tr>
<td>Workload</td>
</tr>
<tr>
<td>Continuity of care</td>
</tr>
<tr>
<td>Supportive coworkers</td>
</tr>
<tr>
<td>Learning environment</td>
</tr>
<tr>
<td>Role ambiguity</td>
</tr>
<tr>
<td>Autonomy</td>
</tr>
<tr>
<td>Supportive faculty</td>
</tr>
</tbody>
</table>
faculties that gave such feedback ($P = .002$). Also of note, only 20.5% of residents reported receiving sufficient career counseling from faculty, whereas 52.3% of faculty reported that they thought the residents were receiving adequate career counseling ($P = .009$).

**Comment**

This study is the first of its kind to look at resident satisfaction in an obstetrics and gynecology residency program and also the first to compare resident and faculty perceptions of factors that affect resident satisfaction. The dissatisfaction rate with the medical profession of 28.5% of the residents and 21.0% of the faculty is similar to that from the Community Tracking Study physicians’ survey, which showed that in 1996 to 1997, 24.2% of obstetrics-gynecology physicians were very or somewhat dissatisfied. It would appear that our residents and faculty are as satisfied as other obstetrician-gynecologists; however, the Community Tracking Study reported that obstetrics-gynecology was the specialty with the lowest percentage of physicians who report being “very satisfied.” The identification of domains most relevant to resident satisfaction in which the residents perceive weaknesses in the educational process, but the faculty may not, has potential to identify opportunities for improvements. Because faculty contact can

<table>
<thead>
<tr>
<th>Table III</th>
<th>Level of agreement between the responses of the residents and faculty members on the survey items and factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factors in which there was agreement</td>
<td></td>
</tr>
<tr>
<td>Collelgiality</td>
<td>Residents, %</td>
</tr>
<tr>
<td>A lot of interaction</td>
<td>85.7</td>
</tr>
<tr>
<td>Residents participate in discussions</td>
<td>40.0</td>
</tr>
<tr>
<td>Supportive coworkers</td>
<td></td>
</tr>
<tr>
<td>Residents listen to each other</td>
<td>94.3</td>
</tr>
<tr>
<td>residents help each other</td>
<td>88.6</td>
</tr>
<tr>
<td>residents gossip about each other</td>
<td>88.6</td>
</tr>
<tr>
<td>Relevant training</td>
<td></td>
</tr>
<tr>
<td>residents learn procedural skills</td>
<td>88.2</td>
</tr>
<tr>
<td>residents learn cognitive skills</td>
<td>62.9</td>
</tr>
<tr>
<td>program provides good preparation</td>
<td>71.4</td>
</tr>
<tr>
<td>Workload</td>
<td></td>
</tr>
<tr>
<td>residents have to work very fast</td>
<td>85.7</td>
</tr>
<tr>
<td>residents workload is about right</td>
<td>54.3</td>
</tr>
<tr>
<td>Factors in which residents thought the domain was present more than faculty</td>
<td></td>
</tr>
<tr>
<td>Adequate resources</td>
<td></td>
</tr>
<tr>
<td>clinic provides sufficient support</td>
<td>57.1</td>
</tr>
<tr>
<td>Continuity of care</td>
<td></td>
</tr>
<tr>
<td>residents are satisfied</td>
<td>82.9</td>
</tr>
<tr>
<td>residents provide most of care</td>
<td>91.5</td>
</tr>
<tr>
<td>Role ambiguity</td>
<td></td>
</tr>
<tr>
<td>residents know their responsibilities</td>
<td>65.7</td>
</tr>
<tr>
<td>residents know how to get work done</td>
<td>94.3</td>
</tr>
<tr>
<td>Factors in which residents thought domain was less present than desired</td>
<td></td>
</tr>
<tr>
<td>Autonomy</td>
<td></td>
</tr>
<tr>
<td>residents can influence things</td>
<td>71.4</td>
</tr>
<tr>
<td>residents have input on how patient care demands are shared</td>
<td>77.1</td>
</tr>
<tr>
<td>Supportive faculty</td>
<td></td>
</tr>
<tr>
<td>faculty are committed to teaching</td>
<td>57.1</td>
</tr>
<tr>
<td>faculty repeatedly demean residents</td>
<td>48.6</td>
</tr>
<tr>
<td>staff are approachable with concerns</td>
<td>62.9</td>
</tr>
<tr>
<td>Learning environment</td>
<td></td>
</tr>
<tr>
<td>resident education is a high priority</td>
<td>51.4</td>
</tr>
<tr>
<td>faculty use current medical literature</td>
<td>47.1</td>
</tr>
<tr>
<td>timely constructive feedback exists</td>
<td>34.3</td>
</tr>
<tr>
<td>adequate career counseling exists</td>
<td>20.5</td>
</tr>
</tbody>
</table>
positively affect resident satisfaction\(^7\) and residents and faculty positively affect student interest,\(^4\) making these improvements in residency programs could positively affect our specialty as a whole.

Ten factors were found to be associated with resident satisfaction in our study, whereas Randall et al\(^7\) found only 6 factors to be correlated with resident satisfaction. Also, learning environment was added as a variable in this study, which was not included in the earlier study of primary care residents. The common factors between their study and this one are continuity of care, autonomy, and collegiality. The study by Randall et al\(^7\) used responses from 119 internal medicine, family medicine, and pediatrics residents. The difference in factors identified as important to satisfaction level may be due in part to the differences between primary care residency programs and surgical residency programs.

Although the majority of the survey instruments items were derived from those used in studies in which internal reliability was shown.\(^7,8\)\(^\) 1 limitation of this study is that these instruments were not independently validated. It must also be acknowledged that this is a single institution study and all the factors associated with satisfaction were not the same as those in the study of primary care residents.\(^7,8\) Also, this study was based on self-reported data and subject to the well-known limitations of such. Residents’ opinions often vary from rotation to rotation, although at a given point in a year, this sample should reflect opinions from all rotations at our program. The faculty instrument asked the faculty members to make judgments on residents’ opinions. This may be somewhat difficult, given varying resident contact and familiarity from faculty member to faculty member. The faculty frequently chose the neutral response option on several questions, which illustrates this concern.

The 3 domains associated with resident satisfaction, including autonomy, supportive faculty, and learning environment, were not present to the extent desired and this apparent deficit appeared to be underappreciated by the faculty members. Although our residents perceive adequate instruction in the how-to aspects of clinical and surgical practice, their concerns about deficits in cognitive skills teaching would imply that the faculty has insufficiently emphasized the why of what we do for our patients. We believe learning environment might be construed as a sum of the other domains, making it perhaps the most important factor in this study. Faculty should be instrumental in shaping the learning environment, specifically giving more timely constructive feedback, explicitly delineating expectations at the start of each rotation, emphasizing evidence-based medicine, and formalizing career counseling to improve resident satisfaction.

The American College of Obstetricians and Gynecologists Task Force on Residency Issues included many of the domains associated with satisfaction in this study (John R. Musich, personal communication, February 21, 2005). Similar recommendations were made more than 15 years ago and illustrate the never-ending need for educators to commit themselves to improving the residency learning environment.\(^10\) New issues face today’s educators; however, the increased focus on providing service to patients at multiple sites and producing clinical revenue, takes time away from teaching.\(^11\)

Some of the recommendations have already been implemented on some services including appointment of a resident advocate, an informal effort to match interested resident mentees to faculty members with skill sets that the resident hopes to acquire, resident-to-resident mentorship, and a midrotation feedback. However, for change to be successful, there must be total commitment to these changes by those directing the program along with the entire faculty.\(^9\) Change is necessary not only to improve resident morale and learning environment and thus medical student recruitment but also to improve patient outcomes and satisfaction because prior studies have shown that physician satisfaction correlates with patient compliance with treatment\(^12\) and patient satisfaction.\(^13\)

Future studies involving other programs might delineate additional improvements that those program directors could share in the literature and at meetings as part of best practices to enhance the learning environment for residents.

References

Residency attrition rate in obstetrics and gynecology: Are we losing more postgraduates today?

Maria Manriquez Gilpin, MD*

Morgan Hill Women’s Health Care, Morgan Hill, CA

Received for publication February 24, 2005; revised June 4, 2005; accepted July 22, 2005

Objective: The purpose of this descriptive study was to determine the attrition rate in 2003 and to establish where residents matriculate after leaving an obstetrics and gynecology residency program.

Study design: A questionnaire was sent by e-mail to all program directors in obstetrics and gynecology residencies in the United States. The questionnaire asked for the number of residents who had left a program, what year of training the resident was in, and whether the departure was a transfer, withdrawal, or dismissal. It asked whether a transfer was to an obstetrics and gynecology residency program or to another specialty; if the resident transferred to another specialty, which specialty did the resident choose.

Results: Two hundred nineteen of 253 programs responded (86.5%). Of residents who left programs, 49% left in the first year of training; 34% left in the second year of training; 13% left in the third year of training, and 4% left in the fourth year of training. The reason for attrition was that 75% of the residents transferred to another residency program; 16% of the residents withdrew from training, and 8% of the residents were dismissed. Of the transferring residents, 60% remained in obstetrics and gynecology.

Conclusion: Although resident attrition was higher than in 1992, more residents remained in obstetrics and gynecology.

© 2005 Mosby, Inc. All rights reserved.

Recruitment and retention of residents is fundamentally important to obstetrics and gynecology. Although the former challenge has attracted considerable attention,1 the latter has not.

In 1992, the attrition rate was determined to be 3.47% or 299 residents over 2 years.2 In that study, most residents who left obstetrics and gynecology programs joined other specialties. The data used to evaluate the rate of attrition in obstetrics and gynecology in 2003 was the Accreditation Council for Graduate Medical Education (ACGME) Accreditation Data System (ADS). The ADS is a collection system that contains data for all ACGME-sponsoring institutions and programs. The attrition rate in obstetrics and gynecology for 2003 was 4.5% or 210 of the 4665 residents. Of those 210 residents, 9 residents were dismissed; 132 residents transferred to other programs, and 69 residents withdrew.

To date, no study has examined the career path of residents who leave their obstetrics and gynecology
One study suggests that, when all residencies are examined, residents who changed residencies rarely left a controlled living (CL) specialty (eg, anesthesiology, dermatology, emergency medicine, pathology, ophthalmology, and radiology) for a non-CL (NCL) specialty (eg, family practice, internal medicine, pediatrics, obstetrics/gynecology, and surgery). During the time of that study, there was a net gain in CL specialties and a net loss in NCL specialties.

The objective of this study was to determine the subsequent specialty training of the residents who left their residencies in obstetrics and gynecology. It was assumed that most residents who left an obstetrics and gynecology residency would enter a controlled lifestyle specialty.

**Material and methods**

I obtained an e-mail list for all the United States obstetrics and gynecology residency programs and sent the survey by e-mail to each of the 253 program directors or coordinators. A second and third request was sent from the ACGME Residency Review Committee for Obstetrics and Gynecology. The survey asked for the following data on the residents who had left their residency programs in the past 3 years: the number of residents who left and the postgraduate year they were in when they left the residency program. The survey also sought to identify whether the departure was a transfer, a withdrawal, or a dismissal. The survey inquired whether a transfer was to another obstetrics and gynecology program or another specialty; if the transfer was to another specialty, the survey wanted to know which one. This survey did not identify the programs or residents that were involved, and the data were collected without identifiers. I assured that no program information was duplicated by a check-off list.

**Results**

Of the 253 surveys that were sent, 219 responses (87%) were received from the program directors. The transfer rate was 76%; the withdrawal rate was 16%, and the dismissal rate was 8% (Figure 1). The year in which residents were most vulnerable for attrition was in the first year (49%), followed by the second year...
(34%), the third year (13%) and finally the fourth year (4%; Figure 2). Collectively, 83% attrition occurs in the first 2 years.

Of the transferring residents, 60% remained in obstetrics and gynecology (Figure 3); the remaining 40% entered various specialties. Fewer than 20% of residents transferred to CL specialties collectively. An equal percentage of transferring residents went into specialties that were considered to be NCL.

More residents (77.4%) in programs that had higher attrition rate (> 3 residents in 3 years) chose to remain in obstetrics and gynecology, compared with programs with low attrition (0-3 residents; 55.8%; rate ratio, 1.75; 95% CI, 1.32, 2.33).

Comment

Although resident attrition in obstetrics and gynecology has increased by 1% from 1992, these study data show that more residents are remaining in obstetrics and gynecology training programs. With the use of the data from ACGME ADS for the year 2003,4 the attrition rate currently is 4.5%, with 62.9% of residents transferring, 32.9% of residents withdrawing, and 4.2% of the residents being dismissed. The study in 1992 indicated that more transferring residents chose to continue training in a specialty other than obstetrics and gynecology.5 These survey results indicate a reversal of that trend, which may actually indicate a net of no real change in resident attrition from obstetrics and gynecology. When programs with higher attrition rates are compared with programs with low attrition rates, the data suggest that there is a 1.75-fold increase of residents who choose to remain in obstetrics and gynecology. It is interesting to speculate the reason for this observation, although that reason is not available from this studies information. Most of the residents who transferred did not leave their programs to join less rigorous or CL specialties.

In efforts to maintain brevity in the questionnaire information to stratify for postgraduate year, the number of resident total per program or gender was not obtained. Earlier reports have identified important gender differences in resident attrition.5

Why are residents choosing to leave their original programs, and how can these programs decrease their risk for losing residents to other obstetrics and gynecology residency programs? The American College of Obstetricians and Gynecologists (ACOG) Presidential Initiative Task Force on Residency Issues6 reported to the ACOG Executive Board in January 2004 and addressed these questions and presented innovative ideas for the improvement of the content of residency curriculum, the conduct of residency training, and the perception of post-residency lifestyle. At the request of the task force, questions regarding resident satisfaction were included in the 2004 CREOG examination survey. The executive committee at ACOG made the results known to me; however, the data has not been published yet. Interestingly, inadequate education and difficult work relationships lead the reasons that residents would leave their residency programs. Excessive workload and another career option were less important factors in attrition.

In conclusion, even though the residency program directors in obstetrics and gynecology put forth efforts in recruiting medical school candidates, they must also be attentive to the retention of residents early in their training. Collectively, we in education must identify factors that place residents at risk and modalities to decrease that risk.

References

Objective: The purpose of this study was to determine whether the Myers-Briggs type inventory extraversion is associated with clinical evaluation ratings that students earn during their (obstetrics/gynecology) junior medical student clerkship.

Study design: The Myers-Briggs type inventory was administered to medical students during their obstetrics/gynec clerkship. Bivariate correlations between clinical evaluations, National Board of Medical Examiners subject scores, and data from the Myers-Briggs type inventory extraversion scale were analyzed.

Results: Pearson product-moment correlation between clinical and National Board of Medical Examiners subject scores was not significant (r = .25; P = .05). The National Board of Medical Examiners did not show significant correlations with the Myers-Briggs type inventory extraversion data. The clinical evaluations showed a significant correlation (r = .35; P = .005) with Myers-Briggs type inventory extraversion.

Conclusion: Results show that Myers-Briggs type inventory extraversion is correlated positively with clinical evaluations. The National Board of Medical Examiners subject examination and clinical evaluations were not correlated significantly. Findings question whether clinical evaluation data should be included in the obstetrics/gynecology medical student evaluation process.

© 2005 Mosby, Inc. All rights reserved.

Medical student performance in obstetrics/gynecology clerkships is evaluated in different ways. One such mode of evaluation is clinical. Previous investigators have reported on the methods of clinical evaluations used in medical student clerkships in various specialties. These evaluations typically include an assessment of patient presentation skills, performance on ward rounds, and performance in outpatient clinics. Zahn et al surveyed various obstetrics/gynecology clerkships across the nation and found wide variation in clinical assessment methods and the ways in which these evaluations were incorporated into the final grades of students. Awad et al reported on clinical evaluations in a surgery clerkship and found poor correlations with more objective modes of evaluation. In 2001, Go reported that the introduction of more objective performance criteria produced more reliable results in the clinical evaluation in an emergency medicine clerkship. Ginsburg in 1985 commented on the high rate of variation of clinical evaluation results of students by teachers in an internal medicine clerkship.
At our study site, students are observed in various clinical settings. Faculty members and residents rate students on their ability to perform physical examinations, present patient information, prepare write-ups of findings, develop assessments and plans, and interact with patients and the rest of the clinical team. These observations contribute to the clinical evaluation score that each student receives after the completion of their obstetrics/gynecology clerkship. It is possible that personality characteristics of the medical student influence the clinical evaluation component. This important concern, although consistent with perceptions about clinical evaluations, has not been investigated empirically.

The Myers-Briggs Type Indicator (MBTI) is a personality inventory that has been used in behavioral research studies for many years. Past studies indicated that personality characteristics influence objective examination scores, clinical performance, and academic achievement. It is possible that these same characteristics affect the opinions of observers when they evaluate the clinical performance of medical students in obstetrics/gynecology. The particular type of personality features that might influence clinical evaluation scores is not known.

It is possible that students with similar clinical abilities are receiving different scores because of personality characteristics of the student, rather than clinical knowledge and performance. In some situations, such personality characteristics might enhance or perhaps have a deleterious influence on evaluations of clinical abilities.

The purpose of this study was to evaluate whether the extraversion component of the MBTI personality inventory correlates with clinical evaluation scores, and if clinical and National Board of Medical Examiners (NBME) subject examination scores were correlated.

A correlation between Myers-Briggs extraversion and clinical evaluations by evaluators would indicate that the current method of clinical evaluations is influenced by extra-test variables, such as personality characteristics of the medical student, and thus should be reconsidered as part of the evaluation process of obstetrics/gynecology clerkships.

### Material and methods

During the research period, 64 medical students matriculated through our obstetrics/gynecology clerkship. This included 3 clerkship periods, with each clerkship including from 14 to 20 students. The students were offered participation in the study during orientation. Once consent was obtained, study participants were administered the MBTI (form M) by a faculty member who was not otherwise involved in the clerkship evaluation.

During the rotation, students were evaluated clinically by the residents and faculty and, at the end of their clerkship, were administered the NBME subject examination. Faculty members had no knowledge of the results of the MBTI. For the subjective evaluation, the clerkship director integrated standardized evaluation feedback from residents and faculty members in the form of a scanned document. Categories included clinical performance, interaction with patients, individual initiative, work ethic, fund of medical knowledge, and interaction with residents/faculty. In each category, the students were given a score of 1 to 5 on a Likert-type scale, with 1 reflecting 0 and 5 reflecting superior performance. This information was then summed to derive a numeric score for each student. Clinical and NBME subject examination scores were then compared with the MBTI extraversion raw scores from the data of the MBTI.

The MBTI is a self-administered personality inventory that is based on the personality types as theorized by Carl Jung. The Clerkship Director was not in attendance during the administration of the MBTI, was given no information about the students during the data collection period, and was not given any information concerning student participation in the research study. The research study was approved by the Human Rights Advisory Committee at the University of Arkansas for Medical Sciences.

### Results

The MBTI was administered in accordance with recommended assessment instructions. Sixty-three of 64 medical students (98%) who were solicited for participation agreed to participate in the study. All participants successfully completed the MBTI and the obstetrics/gynecology clerkship rotation. NBME subject examination data and subjective evaluations were retrieved for all participants. Data were analyzed in SPSS software (version 12.0; SPSS, Inc, Chicago, IL). There were no missing data across all critical study variables for the 63 students who participated in the study. All statistical tests were 2-tailed, with a probability level of <.05 for significance.

There were 35 men (55%) and 28 women (44%) in the study. One student chose not to indicate gender as part of the study protocol. The average age of participants was 26 ± 2.8 (SD) years (range, 22-39 years) for the 63 participants. The \( t \)-test for independent means indicated no significant difference in age of men or women (\( t = 0.90; \) degrees of freedom, 61; \( P > .05)\).

NBME subject component scores for all groups ranged from 50 to 89 (mean, 70.25 ± 9.06). Clinical evaluation scores ranged from 70 to 100 (mean, 89.21 ± 5.80), which reflects a more restricted range of evaluation dispersion. MBTI extraversion raw scores ranged from 0 to 21 (mean, 11.10 ± 6.26). MBTI Introversion raw scores ranged from 0 to 21 (mean, 9.87 ± 6.19).
The Pearson product-moment correlations for the NBME subject examination and the clinical evaluations were not statistically significant (r = .25; P = .05), which suggests relative independence of these evaluations. The correlation between the NBME and MBTI extraversion raw score was also not significant (r = .22; P > .05), which suggests no statistically significant association between these assessments.

The correlation between the MBTI extraversion raw data and the clinical ratings was statistically significant (r = .35; P = .005) and reflective of moderate effect size. As support for construct validity of the MBTI, a significant and negative correlation (r = -.38; P = .002) was shown for MBTI introversion and clinical ratings. In addition, the MBTI extraversion and introversion were significantly and negatively correlated (r = -.97; P < .001). Figures 1 and 2 are scatter plots and regression lines of these correlations. As shown, both MBTI extraversion and introversion represent considerable dispersion.

**Comment**

Data indicate that, for medical students who rotate through our 6-week obstetrics/gynecology clerkship, the NBME subject scores do not have a significant statistical correlation with clinical evaluations and have no significant correlation with MBTI extraversion. This echoes past reports regarding clinical evaluations and a lack of meaningful correlations with more objective measures of clinical ability.\(^2,10\)

Clinical evaluations are significantly correlated with MBTI extraversion and introversion, which suggests that personality factors of the medical student may be associated with evaluation scores. This questions whether our clinical evaluation components of the obstetrics/gynecology clerkship should be included as part of the evaluation process. Others have also found poor correlations of clinical evaluations with medical student performance in other settings and questioned the role of clinical evaluations in the role of medical student evaluation.\(^2\) An alternative is to change the labeling of the “clinical evaluation” to “interpersonal skills” because this may reflect more accurately what is being measured. The Accreditation Council for Graduate Medical Education has recently recommended the inclusion of competencies in curricula that assess interviewing and interpersonal skills. Perhaps in our setting, clinical evaluations are the right tool being used in an inaccurate way.

It is also possible that clinical evaluations are influenced by personality characteristics of the evaluator. Further investigation of this possibility could involve an assessment of evaluators to consider the possibility of interactions between personality variables of the evaluator and student. Clerkships that use clinical evaluations should consider each of these possibilities in the evaluation of students.

Finally, the metrics of clinical evaluations need careful evaluation. In our setting, a Likert-type scale is being converted to a nominal scale, which introduces mathematical error because the 2 scales are incompatible.

With this consideration, it is important to keep the findings of this preliminary study in mind when clinical evaluations are incorporated into the formal clerkship grade. In addition, those persons who evaluate transcripts for residency applications are encouraged to consider how a clinical grade is reported and how such “grade” may be more influenced by personality rather than clinical skills.
It is important to note that these preliminary findings should be considered with some caution. It is also recommended that the findings of this study be replicated procedurally. Currently, we are in the process of expanding this trial over time with the use of our current methods to provide additional data regarding these preliminary findings. Repeating this investigation in other clerkships at The University of Arkansas for Medical Sciences and at other institutions with different and similar clinical evaluation methods would assist in a more robust evaluation of the educational usefulness of the subjective evaluation process. Other researchers might find it beneficial to assess similar personality characteristics of those persons who conduct the clinical evaluations. Furthermore, it is unclear whether findings of this study would be generalizable to other medical clerkship rotations. Further research is needed to address these concerns. As a final note, it is known that clinical medicine and a physician’s success as a practitioner and colleague are often influenced by a reflection of personality. The outgoing, extraverted physician is likely to interact and “connect” with others, including patients, in a different way than someone who is lower in extraversion.

References

Focused assessment of surgical performance: Difficulty with faculty compliance

Gabriella G. Gosman, MD,* Hyagriv N. Simhan, MD, MSCR, Richard S. Guido, MD, Ted T. M. Lee, MD, Suketu M. Mansuria, MD, Joseph S. Sanfilippo, MD, MBA

Department of Obstetrics, Gynecology and Reproductive Sciences, Minimally Invasive Gynecologic Surgery Program, University of Pittsburgh School of Medicine, Pittsburgh, PA

Received for publication February 21, 2005; revised June 1, 2005; accepted July 20, 2005

KEY WORDS
Education
Residency
Training program

Objective: This study evaluated faculty compliance in the use of the global surgical rating scale of the Objective Structured Assessment of Technical Skills to rate resident surgical performance after every endoscopic procedure.

Study design: For this prospective cohort study, 4 faculty members in the Minimally Invasive Gynecology Surgery Program were asked to rate resident surgical performance using the Objective Structured Assessment of Technical Skills instrument after every case. Faculty compliance was analyzed with respect to the influence of the resident or surgical case characteristics. Faculty and residents completed surveys about the value of the case-by-case ratings.

Results: Faculty members used the Objective Structured Assessment of Technical Skills instrument 36% of the time (range, 26%-60%). Faculty member compliance did not vary according to resident or surgical case characteristics. Faculty members did not think the forms had much impact on whether they gave feedback. Residents thought the opportunity to read their ratings was helpful.

Conclusion: Faculty member compliance with case-by-case surgical performance evaluation of the residents was low.

Residency training programs use the assessment of residents’ surgical performance to both formative and summative ends for the residents and the educational program. Resident surgical performance in our institution currently is evaluated with a global performance evaluation. Surgical skill is 1 of 9 categories that are rated by attendings on this form. Faculty members complete these forms at varying times after the end of each rotation. Such a single global rating of surgical performance may be very unreliable. To determine the feasibility of a more informative and intensive approach for the assessment of resident surgical performance, this study pilots the use of a validated surgical performance assessment instrument (the global rating scale of the Objective Structured Assessment of Technical Skills [OSATS]) for real-world use in an academic obstetrics and gynecology residency.

* Reprint requests: Gabriella G. Gosman, MD, 300 Halket St, Room 2314, Pittsburgh, PA 15213.
E-mail: ggosman@mail.magee.edu

0002-9378/S - see front matter © 2005 Mosby, Inc. All rights reserved.
The OSATS involves direct rater observation of residents performing technical or surgical tasks. Two scoring systems are used for these tasks, a task-specific checklist and an itemized global rating scale. Both scoring systems have been shown to be reliable and valid in formal multistation surgical testing environments that involve live surgery and bench model surgical tasks. Coleman and Mueller used the global rating scale to score obstetrics/gynecology resident surgical performance of laparoscopic salpingectomies to assess the effectiveness of a laparoscopic training intervention. We hypothesized that the global rating scale of the OSATS might be a means to improve significantly our ability to assess resident surgical performance on a case-by-case basis outside of the formal testing setting. The global scale was selected because it can be used for a variety of procedures. This study focuses on the feasibility of completing a global rating scale after every case that faculty members perform with a resident. Specifically, we evaluated faculty members’ ability to comply with this case-by-case rating protocol. For this pilot study, we limited faculty participation to the 4 faculty members of the Minimally Invasive Gynecology Surgery (MIS) Program in the Department of Obstetrics, Gynecology and Reproductive Sciences at the University of Pittsburgh School of Medicine, Magee-Womens Hospital. This group was selected because of the unique training needs of residents who are learning endoscopic surgery in an obstetrics/gynecology residency. We also evaluated faculty and resident attitudes toward this novel evaluation policy.

**Material and methods**

**Study design and population**

This is a prospective cohort study to evaluate the feasibility of surgical performance evaluation of residents after every case in which residents participated. The 4 faculty members of the MIS Program in the Department of Obstetrics, Gynecology and Reproductive Sciences underwent an orientation to the global rating scale portion of the OSATS, which is referred to as “rating form.” The Figure shows the adapted form. During the study period from February 9, 2004, through July 8, 2004, the 4 faculty members agreed to complete a rating form for resident surgical performance after every laparoscopic, hysteroscopic, or combined laparoscopic/
hysteroscopic surgical procedure that they performed with a resident or residents. Rating forms were printed on 5 × 7-inch cards and were given to faculty members. Additional information that was collected on the rating forms included resident level of participation, faculty initials, resident name, date of the procedure, and the procedure type. Faculty participants were oriented as to how to use the form to evaluate a resident. However, faculty members were not specifically told where or when to complete the form. Completed cards were collected by the primary investigator and kept in a locked cabinet until they were given to medical records personnel, who served as honest brokers. Procedures and resident participation were confirmed with de-identified operative notes. Procedure times were obtained with the use of de-identified operating room records.

Before the study period, residents were informed of the study purposes and procedure and were oriented to the rating form itself. At the end of the study period, rating forms and de-identified operative reports were compiled for each rated resident. Each resident had the opportunity to review the ratings and the operative notes of the cases in which they participated. Residents then completed a survey regarding their perceptions of the feedback value of the rating forms. At the end of the study period, faculty participants also completed a survey regarding the value of the rating forms in terms of resident surgical performance evaluation and the facilitation of feedback. The faculty participants met to discuss the strengths and weaknesses of the form. This protocol was reviewed by the Magee-Womens Hospital Institutional Review Board and given exempt status.

### Statistical analyses

Faculty percent compliance was defined as the total number of completed rating cards divided by the total number of surgical cases that were performed by faculty with residents during the study period. Compliance was calculated for the faculty group overall and also was stratified by individual faculty members. Differences in compliance by faculty member were analyzed using the \( \chi^2 \) test and logistic regression. Additionally, the influence of a number of other covariates on compliance was determined. Multivariable logistic regression was performed to identify factors that enhanced or detracted from compliance. These covariates included resident postgraduate year (PGY), the type of surgical procedure (laparoscopy, hysteroscopy, or both), long operative time, residents with a high number of cases performed with the 4 faculty members during the study period, and 2 resident participants in the case.

Faculty and resident survey data were handled separately. For faculty and resident survey items that were expressed in Likert format, scores were summarized with median and range. Survey questions were adapted from an obstetrics/gynecology resident feedback survey of Farrugia and Shapiro. Responses to open-ended questions were analyzed for content and thematic trends. All statistical analyses were performed with Stata software (version 8.0 for Windows; Stata Corp, College Station, TX).

### Results

The 4 faculty participants and 27 residents (PGY 1-4) performed 243 MIS cases. There were a total of 275 potential surgical performance rating encounters between faculty members and residents because 32 cases involved 2 residents operating together. Rating forms were completed for 98 cases (35.6%). Table I displays the number of cases that were performed, the number of

<table>
<thead>
<tr>
<th>Attending</th>
<th>Potential rating encounters (n)</th>
<th>Ratings completed (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>81</td>
<td>41 (50.6%)</td>
</tr>
<tr>
<td>2</td>
<td>138</td>
<td>36 (26.1%)</td>
</tr>
<tr>
<td>3</td>
<td>51</td>
<td>18 (35.3%)</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>3 (60%)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>275</td>
<td>98 (35.6%)</td>
</tr>
</tbody>
</table>

Table II: Attending completion of surgical performance rating form based on resident and case characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>n</th>
<th>Rated by attending</th>
<th>P value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGY</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>0</td>
<td>.07</td>
</tr>
<tr>
<td>2</td>
<td>42</td>
<td>18 (42.9%)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>156</td>
<td>61 (39.1%)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>73</td>
<td>19 (26.0%)</td>
<td></td>
</tr>
<tr>
<td>Procedure type</td>
<td></td>
<td></td>
<td>.37</td>
</tr>
<tr>
<td>Laparoscopy</td>
<td>166</td>
<td>59 (35.5%)</td>
<td></td>
</tr>
<tr>
<td>Hysteroscopy</td>
<td>44</td>
<td>16 (36.4%)</td>
<td></td>
</tr>
<tr>
<td>Laparoscopy and hysteroscopy</td>
<td>33</td>
<td>16 (48.5%)</td>
<td></td>
</tr>
<tr>
<td>Operation time greater than median for procedure type</td>
<td>122</td>
<td>44 (36.1%)</td>
<td>.66</td>
</tr>
<tr>
<td>Residents with &gt; 11 cases with MIS faculty during study</td>
<td>196</td>
<td>71 (36.2%)</td>
<td>.75</td>
</tr>
<tr>
<td>Cases with 2 residents operating together</td>
<td>32</td>
<td>10 (31.3%)</td>
<td>.44</td>
</tr>
</tbody>
</table>

* \( \chi^2 \) test.
cases that were rated, and compliance for each rater. Compliance was noted to be statistically different by rater \((P = .002, \chi^2\) test). Only rater 2 was noted to have statistically significantly decreased odds of compliance (odds ratio, 0.34; 95% CI, 0.19-0.61) when the rater was considered as an independent predictor of compliance in a logistic regression model. In Table II, we present characteristics of the surgical cases as stratified by compliance. Long operative time was defined by an operative time that was greater than the median for a given procedure type. In theory, the number of cases that attendings perform with a given resident might influence the likelihood that an attending would evaluate that resident’s performance. To ascertain this phenomenon, resident MIS surgical volume was dichotomized as high volume (greater than or equal to the median number of cases performed per resident with MIS attendings) or low volume (fewer than the median number of cases performed per resident with these attendings). There was no statistically significant univariate association between compliance and the level of resident participant, procedure type, long operative time, high volume rater exposure, and the performance of the case by 2 residents. However, the univariate association of compliance with resident year approached statistical significance. We assessed this relationship, which was adjusted for long operative time and high volume rater exposure, using multiple logistic regression. We included these variables as possible confounders in the model because of an a priori assumption of a possible and plausible relationship of long operative time with compliance and because resident year and high volume rater exposure were highly associated \((P < .001, \chi^2\) test). In this multiple variable model, we found that PGY status was not significantly associated with compliance.

Seventeen of 24 rated residents (70.8%) completed a poststudy survey. Table III describes the responses of the residents to the performance assessment instrument. On supplemental multiple-choice questioning, 76.4% of the residents preferred using the rating cards after every case, with a summary at the end of the rotation; 17.7% of the residents preferred a single summary evaluation at the end of the rotation. One resident (5.8%) preferred

Table III  Resident responses* to selected questions on the end of study survey regarding faculty feedback about their surgical performance

<table>
<thead>
<tr>
<th>Category</th>
<th>Item</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall feedback on surgical performance</td>
<td>1. Faculty feedback on my surgical performance is important.</td>
<td>5 (4-5)</td>
</tr>
<tr>
<td></td>
<td>2. Faculty feedback on my surgical performance is accurate.</td>
<td>4 (1-5)</td>
</tr>
<tr>
<td></td>
<td>3. I use faculty feedback on my surgical performance to improve.</td>
<td>4 (2-5)</td>
</tr>
<tr>
<td></td>
<td>4. I ask for faculty feedback on my surgical performance if it is not offered.</td>
<td>3 (1-5)</td>
</tr>
<tr>
<td></td>
<td>5. I generally receive faculty feedback in the operating room.</td>
<td>3 (2-5)</td>
</tr>
<tr>
<td>Feedback on surgical performance during study</td>
<td>6. I got immediate feedback from the faculty on my surgical performance.</td>
<td>4 (2-5)</td>
</tr>
<tr>
<td></td>
<td>7. I was more likely to get immediate feedback during this time period.</td>
<td>3 (1-5)</td>
</tr>
<tr>
<td></td>
<td>8. The summary of faculty feedback from the rating cards is helpful to me.</td>
<td>4 (2-5)</td>
</tr>
</tbody>
</table>

* Rating scale: 0 = not able to evaluate; 1 = hardly at all; 2 = to a small degree; 3 = to a moderate degree; 4 = to a considerable degree; 5 = to a very high degree.

Table IV  Faculty participant responses* to selected questions about the end of the study survey regarding their feedback to residents about surgical performance

<table>
<thead>
<tr>
<th>Category</th>
<th>Item</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall feedback on surgical performance</td>
<td>1. I give feedback to residents on their surgical performance.</td>
<td>3 (2-5)</td>
</tr>
<tr>
<td></td>
<td>2. This feedback is well received.</td>
<td>4 (3-5)</td>
</tr>
<tr>
<td></td>
<td>3. Residents ask for feedback about their surgical performance.</td>
<td>2 (1-3)</td>
</tr>
<tr>
<td></td>
<td>4. Residents use my feedback to improve their surgical performance.</td>
<td>3 (2-4)</td>
</tr>
<tr>
<td>Feedback on surgical performance during study</td>
<td>5. I gave immediate feedback to residents on their surgical performance on cases that I used the rating forms.</td>
<td>2 (1-4)</td>
</tr>
<tr>
<td></td>
<td>6. I was more likely to give immediate feedback during the period that I used the rating forms.</td>
<td>1 (1-4)</td>
</tr>
</tbody>
</table>

* Rating scale: 0 = not able to evaluate; 1 = hardly at all; 2 = to a small degree; 3 = to a moderate degree; 4 = to a considerable degree; 5 = to a very high degree.

Data are given as median (range).
feedback during surgical cases exclusively, and 6 residents (35.3%) desired verbal feedback during and/or immediately after surgical cases in addition to the rating card and/or end-of-rotation feedback. One resident noted that her intraoperative feedback was not consistent with what was recorded on the rating cards.

All 4 faculty members filled out a poststudy survey. Table IV contains these faculty member responses. Faculty members estimated that completion of the rating card took 3.75 ± 1.5 minutes. One faculty participant indicated that a distinct rating system was needed for residents who participated as assistants, rather than surgeons, on rated cases. The faculty group session revealed that all of the participants completed the rating cards when they were back in their offices, rather than in the presence of the resident immediately after the surgical procedure.

Comment

We found that real-world faculty compliance with the global rating scale of the OSATS as a performance assessment tool was poor. Our single and multiple variable analyses failed to identify specific barriers to compliance with the OSATS rating forms, related to identifiable traits of the resident or of the surgical case itself. Faculty participants noted that a perceived barrier to compliance was that they held the onus of responsibility to initiate the evaluation process, rather than the resident. Some faculty participants felt that if residents were responsible for presenting faculty with the rating forms at the conclusion of each case, then the proportion of cases that were rated might improve. One faculty member thought that this particular instrument was not well suited for evaluating residents who participate as surgical assistants (rather than primary surgeon); therefore, this individual was less likely to complete the rating forms.

Technical skill assessment in obstetrics and gynecology resident education is essential for several reasons. Residency programs strive to graduate trainees who are capable of independently performing all of the core procedures in the specialty.7 Assessment of residents’ surgical performance is an important way to evaluate the adequacy of the educational program. Assessment of each individual resident’s surgical skill must be a key component of a program’s promotion and graduation policy. Furthermore, residents are adult learners for whom feedback is a critical element of the educational process.8 Surgical performance assessment obviously serves this purpose as well.

In terms of quality of feedback in our study, the faculty group forum found that the forms did not facilitate immediate feedback because faculty members completed the forms when they returned to their offices rather than at the conclusion of the surgical case in the presence of the resident. This is reflected in the fact that faculty members felt that the rating cards did not make them more likely to give immediate feedback (Table IV). Our survey demonstrated that residents strongly valued faculty feedback on their surgical performance and felt that they would use this feedback to improve. Most residents who were surveyed perceived the rating cards and summary that were derived from these assessments to be a valuable addition to currently used techniques for feedback. More than one third of residents who were surveyed desired verbal feedback immediately on completion of the surgical case. This immediacy of feedback was not facilitated by the rating form policy as implemented in this study.

Our study is unique in that it evaluates faculty compliance with the OSATS global surgical rating scale in the real world of resident performance assessment. Studies by Fung et al9,10 using a similar but less widely used instrument help to interpret our current findings and guide further study in this area. This group used a 3-item interactive voice response instrument to rate obstetrics/gynecology resident performance of laparoscopic procedures. The instrument used a 5-point Likert scale to rate residents on (1) knowledge and handling of instruments, (2) ability to plan and perform operative moves, and (3) knowledge of anatomy. A fourth question asked the rater to describe critical incidents that occurred during the case. Faculty members used the system to rate the resident participant. Residents used the system to rate themselves. Compliance for this study was reported as percent of rated cases in which both faculty and resident used the system. They did not calculate rated versus unrated cases. For feedback purposes, faculty members and residents could listen to each other’s recorded comments. Similar to our findings, residents preferred the new system to the traditional postrotation single-item evaluation of surgical skill. A follow-up validation study of this instrument showed that it required a minimum of 12 ratings of performance during laparoscopic procedures to reach a reliability of 0.80.10 The most practical application of these surgical performance ratings in obstetrics/gynecology residencies would be for assessment during a broad range of surgical procedures. Such a context might increase the number of ratings that are required to attain reliability.

The use of the OSATS global rating scale or interactive voice response instrument facilitates a focused assessment of resident surgical performance in a similar manner to the clinical evaluation exercise (CEX) and mini-CEX. These techniques use direct faculty observation of trainee-patient clinical encounters for formative and summative assessment of trainee performance. The American Board of Internal Medicine strongly recommends the use of mini-CEX in the evaluation of
residents because this technique is efficient and effective. Consistent with the surgical performance assessment instrument, Norcini et al\textsuperscript{11} found that 12 to 14 mini-CEX encounters were needed to reach a reliability of 0.80 for internal medicine residents. Patients in this study had a range of clinical settings with a wide variety of clinical problems. It remains to be determined whether the OSATS global rating scale will function reliably with approximately 12 ratings when applied to the full range of gynecologic endoscopic procedures or further to the full range of gynecologic surgical procedures.

Unlike the mini-CEX, the global rating scale of the OSATS has been validated only in formal testing environments rather than in routine clinical care. Our study represents a necessary step towards the goal of validating this widespread surgical evaluation tool in the daily surgical practice of an obstetrics/gynecology residency training program. Given the compliance rate in our study, very few residents had \( \geq 12 \) rated cases. If compliance cannot be improved substantially, it may require 1 year of collecting these forms in our program (our study took place over 5 months) to achieve this number of rated endoscopy cases for residents at each level. The reliability and validity of the instrument that was used in the context of laparoscopy, hysteroscopy, and combined procedures could then be assessed.

In the future, we will implement faculty participant suggestions to shift initiative for card completion to the residents. They will give a card to the attending surgeon immediately after the conclusion of the endoscopic case. We will explicitly request that faculty members complete the rating at that time. By modifying the protocol and making the timeframe for completion of the form more explicit, we hope to achieve 3 goals: (1) to improve faculty compliance with completion of the form, (2) to encourage more timely feedback to residents by summarizing their performance during a rated case, and (3) to test the reliability and validity of this instrument used in this context.

References
Self-assessment of resident surgical skills: Is it feasible?

Lynn S. Mandel, PhD,* Barbara A. Goff, MD, Gretchen M. Lentz, MD

Department of Obstetrics and Gynecology, University of Washington, Seattle, WA

Received for publication February 21, 2005; revised July 6, 2005; accepted July 20, 2005

Objective: The purpose of this study was to examine obstetrics and gynecology residents’ self-assessment of proficiency on a variety of surgical bench procedures and to compare their ratings with those of trained faculty observers who used instruments that have been shown to be reliable and valid.

Study design: As part of a 6-station Objective Structured Assessment of Technical Skills, 74 residents from 5 institutions estimated their overall open and laparoscopic skill level before the testing. After completing each station, residents evaluated their overall and global skills performance.

Results: Residents rated their proficiency higher on open skills than on laparoscopic skills. Task-specific, overall, and global assessments were correlated significantly with the faculty ratings ($P < .001$). Residents tended to rate themselves lower than did faculty on almost all measures; even those residents with poor skills indicated that they were aware of their deficiencies. Overall and global self-assessments increased with each resident level, which indicated good construct validity.

Conclusion: Residents can rate their overall open and laparoscopic skills, task-specific performance, and global skills with good reliability and validity. Although they tended to score themselves lower than did faculty observers, the correlations are high (ie, residents who give themselves a higher score tended to receive a higher score from faculty, and vice versa). One of the concerns about self-assessment is that residents with poor skills might not be aware of their deficiencies. We did not find that to be the case. Therefore, when residents work on self-directed exercises, task-specific and global checklists can be used for both learning and self-assessment.

Supported in part by a grant from the United States Surgical Corporation and by the Ortho McNeil Educational Grant from the Association of Professors of Obstetrics and Gynecology and a grant from the National Board of Medical Examiners (NBME) Medical Education Research Fund Grant. The project does not necessarily reflect NBME policy, and NBME support provides no official endorsement.

* Reprint requests: Lynn Mandel, PhD, Department of Obstetrics and Gynecology, University of Washington, 1959 NE Pacific, Room BB 667A, Seattle, WA 98195-6460.
E-mail: lmandel@u.washington.edu

With the current emphasis on self-directed and life-long learning, the ability to assess one’s own skills has assumed increasing importance. In a process that is analogous to medical diagnosis and treatment, learners typically must assess their knowledge or skills, determine their learning needs, formulate learning goals, engage in a learning activity, and then reassess proficiency. This ability to assess one’s strengths and weaknesses is essential to the process of life-long learning. Additionally, the development of high-fidelity laparoscopic and surgical simulators and virtual reality...
machines has added impetus to the value of being able to self-assess, especially in simulation-based training.

In a meta-analysis of 44 studies in higher education (including medicine), Falchikov and Boud\(^6\) found that, on average, the correlation between self-assessment and external measures was low. Gordon\(^5\) reviewed 18 studies in the health professions that dealt with knowledge and clinical skills but not technical or surgical skills and found that learners were not able to assess their proficiency accurately. Ward et al\(^6\) updated the analysis by combining these studies with 20 newer studies and found that methods issues limited the applicability of the results. Although >80% of these studies compared ratings of self-assessment with those of an external evaluator, the reliability of these “gold standard” raters was suspect, therefore calling into question both the reliability and validity of their findings.

More recent studies have focused on surgical skills and have shown that trainees are able to assess their own skills, especially after viewing a videotape of their performance.\(^7\) Surgical residents were asked to rate their skill level after performing a laparoscopic procedure in a pig, after viewing a videotape of their performance, and after a review of 4 “benchmark” performances that illustrated a range of proficiencies. The correlations between faculty and resident ratings increased significantly after viewing their own videotapes but not after reviewing the benchmark tapes. It is not surprising that the correlation would be higher after seeing their own performance because faculty were scoring them in real time during the test, whereas the residents had to attempt to recall at the end what they did during the entire procedure. With the video to prompt them, they could review and reflect on their performance at each step.

Repetition of simulated surgical tasks has also shown that medical student trainees can improve the accuracy of their self-assessment with practice.\(^3\) As the repetitions increased, the number of actual and estimated errors decreased. As the authors point out, the fact that the participants felt that they were improving can be a powerful motivational force during learning.

A study at Madigan Army Medical Center\(^8\) found high reliability and validity when obstetrics and gynecology residents were asked to assess their skills in midline episiotomy repair that was performed on a life-like model. Although residents rated their own proficiency lower than did faculty observers, the scores on the component skills were similar.

Direct observation with specific criteria is considered the most reliable and valid method of assessing technical skill during operative procedures.\(^9\) Pioneering work in surgical skills assessment by Reznick\(^10\) and Martin et al\(^11\) in Toronto has shown the Objective Structured Assessment of Technical Skills (OSATS) to have high reliability and validity. Because our previously published studies have shown the same high reliability and validity,\(^12\)-\(^14\) comparison of resident self-assessment scores with the OSATS results should yield reliable results with high concurrent validity and avoid the pitfalls of comparison with an unreliable gold standard. The purpose of this study was to examine resident assessments of their own proficiency on a variety of surgical bench procedures and to compare their ratings with the ratings of trained faculty observers who used instruments that have been shown to be reliable and valid.

### Material and methods

A total of 92 residents at 5 institutions were asked to estimate their overall open and laparoscopic skill levels on a scale of 1 (poor) to 5 (excellent) before starting a 6-station OSATS that included both open and laparoscopic skills that are performed on life-like bench models (Table I). The residents were assured that performance on the OSATS had no bearing on their progress in the residency program. In fact, the individual resident results were not shared with the residency program directors.

Each examination had 3 open and 3 laparoscopic procedures, with a balance of difficult and relatively easy tasks.\(^14\) Immediately after each station, the residents rated their overall performance on that skill (scale, 1-5) and completed the global skills checklist,\(^11\) which covers the same 7 areas, regardless of task (maximum, 35 points for each station). These global skills are applicable to almost any surgical procedure (eg, respect for tissue, knowledge of the procedure, instrument handling, and time and motion). The overall ratings and the global skills checklist had been validated previously when used by faculty members to rate resident performance.\(^15\) Asking residents to complete these forms involved approximately 5 minutes of explanation before the OSATS, and only a few minutes between stations to complete the evaluations. Physicians from the University of Washington who were experienced in administering the OSATS acted as the faculty raters at all testing sites.

Resident and faculty ratings were compared with the use of paired \(t\)-tests, which gave both the correlations between the 2 scores and the statistical significance of the differences. Comparisons by resident level or testing site were carried out with the use of one-way analysis of variance. Additional analyses used Pearson bivariate correlations. Data were analyzed with SPSS for Windows (version 12.0; SPSS, Inc, Chicago, IL).

### Results

Complete sets of faculty and self-assessment data were obtained for 74 residents. No resident refused to participate; however, some data were missing because of
logistic problems (eg, residents on call having to leave before the completion of testing). Self-assessments were compared with the faculty overall and global ratings for each individual skill and a combination of the 3 laparoscopic tasks, 3 open tasks, and the total overall and global scores that were a composite of the 6 tasks that each resident had undertaken.

Overall assessments of open and laparoscopic skills on a scale of 1 (poor) to 5 (excellent) before the start of testing revealed a mean of 3.12 for open and 2.65 for laparoscopic skills ($P < .001$). The 2 measures were highly correlated ($r = .754; P < .001$), which indicated that residents who rated themselves high on open skills also rated themselves high on laparoscopic skills and vice versa.

After completing their overall and global assessments, residents rated their overall performance higher on the 3 open skills than on the 3 laparoscopic skills ($10.2 vs 8.4; P < .001$). Identical trends were found on the global assessments ($74.0 vs 63.4; P < .001$). Ratings for the open and laparoscopic skills were highly correlated ($P < .001$). Overall self-assessment and overall global scores also were correlated significantly with the total overall and total global faculty ratings and the residents’ estimate of their open and laparoscopic skills before they started the OSATS ($P < .001$). Results for the overall assessments are shown in Table II.

Residents tended to rate themselves lower than did faculty members on both the individual tasks and composite ratings and global skills ($P < .001$). For example, with a maximum score of 5 on each task, the mean overall ratings for residents and faculty were abdominal closure (3.3 vs 3.9; $P < .001$), Burch (2.1 vs 2.8; $P < .001$), bowel repair (2.9 vs 3.0; $P = .702$), cystectomy (3.0 vs 3.4; $P = .014$), and laparoscopic salpingostomy (2.7 vs 3.2; $P = .007$).

The mean global scores (maximum, 35) for residents and faculty were abdominal closure (23.4 vs 28.3; $P < .001$), Burch (19.2 vs 22.6; $P < .001$), bowel repair (22.6 vs 24.3; $P = .193$), and cystectomy (20.0 vs 23.3; $P = .001$). Even though faculty ratings of overall performance were higher for laparoscopic salpingostomy, the residents rated themselves slightly higher on the global checklist for this skill (22.1 vs 20.7; $P = .154$). Results for selected individual tasks are shown in Table III. There was also a high degree of correlation between resident and faculty ratings on specific tasks such as abdominal closure ($r = .60$), Burch ($r = .77$), and laparoscopic salpingostomy ($r = .65$; all $P < .001$). The same trend was seen with the global skills checklists: abdominal closure ($r = .59$), Burch ($r = .80$), and laparoscopic salpingostomy ($r = .65$). Regardless of the

### Table II: Means and Correlations for Overall and Global Ratings

<table>
<thead>
<tr>
<th>Composite</th>
<th>Self-Assessment</th>
<th>Faculty Rating</th>
<th>Correlation</th>
<th>$P$ value difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open skills</td>
<td>Overall</td>
<td>8.35</td>
<td>10.22</td>
<td>0.743*</td>
</tr>
<tr>
<td></td>
<td>Global</td>
<td>63.64</td>
<td>74.41</td>
<td>0.753*</td>
</tr>
<tr>
<td>Laparoscopy skills</td>
<td>Overall</td>
<td>7.50</td>
<td>8.57</td>
<td>0.665*</td>
</tr>
<tr>
<td></td>
<td>Global</td>
<td>58.56</td>
<td>64.47</td>
<td>0.679*</td>
</tr>
<tr>
<td>Total overall rating score (maximum, 30)</td>
<td>15.99</td>
<td>18.51</td>
<td>0.759*</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Total global score (maximum, 210)</td>
<td>125.13</td>
<td>136.92</td>
<td>0.664</td>
<td>.001</td>
</tr>
</tbody>
</table>

* $P \leq .001$.
component skills on the OSATS, both faculty member assessments and self-assessments were lower for laparoscopic tasks than for open tasks. Faculty and resident ratings were highly correlated, which means that those residents who rated themselves highest were rated the highest by the faculty. This is illustrated in Figures 1 and 2, which show the total overall and total global ratings from both groups.

Only 22% of the residents rated their overall performance higher than did the faculty; 6% of the residents had the exact same rating, and 72% of the residents rated their performance lower. For the global scores, 30% of the residents rated their overall performance higher; 1% of the residents rated their overall performance the same, and 69% of the residents rated their overall performance lower. Figure 3 shows the distribution for the global scores; the negative numbers indicate that the residents rated themselves higher than did faculty members, and positive numbers show higher faculty member ratings.

As has been reported previously, the OSATS tasks have shown high construct validity (ie, scores improved as resident level increased, with senior residents scoring higher than junior residents). For almost all individual skills, global self-assessments increased with each resident level, which indicates good construct validity (most probability values <.001). The same was true for the residents’ overall 1 to 5 rating on each task. The overall assessments of open and laparoscopic skills before and after testing and the total global score after testing also showed significant differences by level ($P < .001$). As training increased, demonstrated proficiency increased (ie, postgraduate year 4 residents scoring higher than postgraduate year 3 residents who scored higher than postgraduate year 2 residents who scored higher than postgraduate year 1 residents).
One of the major concerns about self-assessment is that residents with poor skills might not be aware of their deficiencies. In comparing overall faculty ratings versus self-assessment for the lowest-scoring residents (≥2 SDs below the mean), we found a significant correlation between the 2 assessments \((P < .05)\) and very few instances in which the faculty rated the resident low but the resident rated himself/herself high. There was no significant difference in the ability of residents to assess their own performance, regardless of whether they were in a program that had a formal surgical skills curriculum. There were also no consistent differences in the correlations between resident and faculty ratings based on the difficulty of the task.

**Comment**

Residents can rate their overall open and laparoscopic skills, overall task-specific assessments, and global skills with good reliability and validity. Although they tended to score themselves lower than did faculty observers, the correlations were high (ie, residents who gave themselves a higher score tended to receive a higher score from faculty members and vice versa). One of the concerns about self-assessment is that residents with poor skills might not be aware of their deficiencies. However, we did not find that to be the case. The process involved only an extra 5 minutes at the beginning of the OSATS to give the residents the forms and to explain the logistics. Completion of the assessments after each task required only a few minutes while the residents were waiting to proceed to the next station. The only costs were duplicating the checklists and a few minutes of faculty time.

Students and residents now can choose from a variety of life-like models, many of which allow them to practice on their own. However, as MacDonald et al\(^3\) pointed out, this type of training provides little feedback and relies heavily on self-assessment. Learning psychomotor skills is enhanced when learners are given immediate, specific feedback. In the context of surgical skills learning, this feedback can come from faculty members, a surgical simulator, and/or the learner’s own ability to match his/her performance with knowledge of the specific procedure and general surgical principles. Therefore, when residents work on self-directed exercises, task-specific and global checklists can be used for both learning and self-assessment.

One only has to review Medline citations over the past few years to see the growing trend to provide laboratory skills training that uses the many bench models and virtual reality simulators that currently are available. Therefore, the ability to assess one’s own proficiency can complement faculty assessment in these simulations, help match the learner’s self-identified needs with the activity, and help reduce the faculty time that is required for supervision and providing feedback.

Ward et al\(^7\) reported that, when residents viewed a videotape of their performance, the accuracy of their self-assessment was significantly closer to that of a trained faculty observer than was their initial assessment immediately after completing the skill. MacDonald et al\(^3\) found that repetition improved the accuracy of residents’ self-assessment. Because the OSATS was a testing situation, it was not feasible to evaluate whether the ability to assess one’s own skills improved with repetition. However, repeating a skill after receiving feedback and/or reviewing a taped performance are options worth investigating.

Another potential research study would involve having both the resident and faculty evaluator complete task-specific and global checklists. Subsequently asking the learner to describe how he/she has performed opens up a dialogue and makes it easier for the faculty evaluator to provide specific feedback. By reviewing the checklists together, faculty members can also point out possibly erroneous self-assessments, which could lead to an improved ability to assess one’s own skills.

In summary, a reliable and valid self-assessment tool can be used to document ongoing learning on specific tasks and provide feedback to the learner, the faculty, and the training program itself. The entire process took only a few minutes before the OSATS and between each testing station; the cost was almost nothing, and the task was easy for residents to complete.

**References**

Objective: This study was undertaken to assess job satisfaction and quality of life aspects among residents in obstetrics/gynecology before and after the implementation of duty-hour requirements.

Study design: We administered a survey to residents before and after duty-hour restrictions, addressing satisfaction with residency training, quality of life, and predictions/impressions of the effect of reduced work hours.

Results: Satisfaction with overall residency training as a discrete survey item did not change; however, the composite score from all responses to specific items increased. Several specific clinical and academic items garnered higher satisfaction scores in 2004. Residents reported less-than-anticipated increases in healthiness of their lifestyle and a decrease in interest in teaching.

Conclusion: This is a prospective assessment of the effect of duty-hour requirements, improvements in residents’ perception of their time and ability to study and pursue research and in clinical areas are encouraging. The perception that there is less interest in teaching is of concern.

The Accreditation Council for Graduate Medical Education instituted duty-hour requirements for all accredited residency training programs as of July 1, 2003. These requirements were put into effect in response to many factors that included historic events in New York State, studies such as the 1999 Institute of Medicine report on the number of medical errors and related patient deaths each year, and proposed legislation that would have made duty-hour restrictions a matter of federal law.1-3 Recent studies have suggested that sleep-deprived residents are at increased risk of motor vehicle accidents and that decreasing the length of call shifts decreases the incidence of serious medical errors among interns.4,5 The overall intent of the duty-hour requirements was to balance the needs of patient safety, resident well-being, and academic and clinical education.6

At the same time duty-hour requirements were implemented for all medical specialties, obstetrics and gynecology training programs faced the problem of declining interest in the specialty on the part of senior medical students in the United States. In the 2004 residency match, 65% of available training positions in obstetrics and gynecology were filled by US medical school graduates, which is a steady decrease from 75% in 2000.7 The rigors of residency training and the perceived lack of a good “quality of life” for obstetrician-gynecologists have been cited as key reasons for this declining interest.8 Although the institution of duty-hour requirements has been found to correlate positively with residents’ perceived “quality of life,”9,10 this phenomenon has not been studied extensively to identify those...
components of quality of life that are improved. Similarly, the relevance of physician job satisfaction to such outcomes as disability claims, job turnover, and altered prescribing practices has been established among practicing physicians but largely is unstudied among residents in training. Finally, the effect of duty-hour requirements on such factors as medical student education by residents is as yet unstudied.

In a previous study, we reported on components of resident job satisfaction, quality of life, and predictions as to the effect of reduced work hours before the actual implementation of duty-hour requirements at our institution. In this report, we again assess this cohort of obstetrics and gynecology residents to evaluate whether 1 year of duty-hour requirements had a significant impact on resident job satisfaction and quality of life.

**Material and methods**

**Survey description**

A structured questionnaire that contained 59 items was administered to all residents in a university-based obstetrics and gynecology residency program, immediately before (June 2003) and 1 year after (June 2004) the duty-hour requirements went into effect. This survey consisted of 4 segments: Five demographic items, level of current satisfaction with aspects of the residency training experience, comprising 38 items, 5 “current quality of life” indicators, and 11 questions as to the direct effect of the duty-hour requirements on residency training and quality of life. For the latter segment, questions were worded as predictions for change before the duty-hour requirements (2003 survey) and as assessments of change afterwards (2004 survey). Questions were answered on a 5-point Likert scale, unless otherwise stated. Details on the development of this survey have been published previously by the authors. The survey was designated as “exempt” by the Colorado Multiple Institutional Review Board.

**Survey administration**

This survey was administered to all obstetrics and gynecology residents at the University of Colorado in June 2003 and again in June 2004. Individual survey responses were anonymous and identified by a code number. Surveys were returned by 33 of 35 residents (94%) in 2003 and by 32 of 35 residents (91%) in 2004. Twenty-three residents completed both surveys. This included 9 residents who were at the postgraduate year–1 level in 2003 (graduating class of 2006), 7 residents from the class of 2005, and 7 residents from the class of 2004.

**Statistical analysis**

Survey responses were entered into a spreadsheet and independently hand-checked for accuracy by 2 of the authors. Descriptive statistics, t-tests, and linear regression analysis were calculated with the use of SPSS for Windows (version 12.0; SPSS, Inc, Chicago, IL). The paired t-test was used to evaluate change in mean responses for identical survey items to control for nonindependence of samples.

**Results**

**Population characteristics**

The mean age of the residents was not significantly different between years. Thirty-one of 33 residents in 2003 were female; 29 of 32 residents in 2004 were female. Seventeen residents in 2003 were married, 2 of whom had children, compared with 20 married residents in 2004, 6 of whom had children.

**Global satisfaction with residency**

The response of residents to the question, “Please rate your overall satisfaction with your obstetrics and gynecology residency program,” was analyzed controlling both for year of training or among different groups of trainees in comparable postgraduate training years and for the effect of promotion or the results from the same group of residents over time. Although one comparison, residency satisfaction among postgraduate year 3 residents, was significantly decreased among those postgraduate year 3 residents who completed the survey in 2004 than among those who completed the survey in 2003; in general, there was no significant change in overall satisfaction with residency between 2003 and 2004. The average satisfaction rating for all residents who completed the survey in 2003 and 2004 was also not different.

**Satisfaction with specific aspects of residency training**

Responses to 37 questions about current levels of satisfaction with specific aspects of training were analyzed for change for those residents who completed both surveys. These questions covered clinical training, academic environment, interpersonal relationships, and lifestyle issues. A full list of items that were queried has been published previously. Of these 37 items, none received significantly lower scores in 2004, and 9 received significantly higher satisfaction ratings (Table 1). Of note, despite an overall 30% decrease in rotating in-house call taken by residents who were working traditional day shifts, there was no significant increase in satisfaction with either the setup of night call or with the amount of night call taken.

In addition to analysis of responses to individual questions, a composite score that comprised the average of each individual’s response to all 38 questions in part II of our survey was generated. These composite scores
were then analyzed for change over time and for correlation either with year of training or with previous composite score. The mean composite score for all 23 residents increased from 3.60 in 2003 to 3.84 in 2004 ($P < .005$). Composite scores for each resident in 2004 were not predicted by postgraduate year level or by previous composite in 2003.

**Quality-of-life indicators**

Responses to quality-of-life indicators were analyzed for the subgroup of 23 residents who completed both the 2003 and 2004 surveys. The paired $t$-test was used because of the relatively small numbers of respondents and the likely significant effect on the variance of responses by the nonindependence of the samples. Weekly mean hours spent sleeping, days of exercise, minutes of exercise per session, and hours spent in educational reading were highly predicted by the response the previous year, did not correlate with advancing year of postgraduate training, and did not exhibit significant change overall. However, the reported hours spent reading narrowly missed statistical significance (mean increase of 0.82 hours or 28% change; $P = .058$).

**Perceived effect of the work week reduction on aspects of training and lifestyle**

The third part of the survey asked residents to predict changes in training and lifestyle measures because of duty-hour requirements in 2003 and to then assess changes in the same 11 measures in 2004. A summary of responses is given in Table II. Of the 11 measures, 2 items received significantly lower scores than predicted after the institution of the limited work week: the assessment of the effect of reduced work hours on healthiness of lifestyle and the effect on interest in teaching. The remainder of the responses did not differ significantly between the 2003 and 2004 surveys.

**Comment**

Physician job satisfaction and quality of life are important factors that affect clinical performance and recruitment to and attrition from any specialty training program. Obstetrics and gynecology programs, in particular, are faced with declining US medical student interest. Many educators are also concerned about the ability to train competent physicians and surgeons with limited time and resources. The imposition of restricted work hours has the potential to affect, positively or negatively, the training and overall satisfaction of obstetrics-gynecology residents.

In this study, we attempted to assess what impact, if any, the duty-hour requirements had on residents’ perceptions of their training and quality of life. We did this in a prospective fashion by asking the same questions both before and after 1 year of duty hours. Each segment of the survey was designed to evaluate satisfaction from a slightly different perspective: Residents were asked first simply to rate current satisfaction with both overall residency experience and specific components of that experience, then to quantify, before and after, certain indicators such as sleep and study time, and finally to estimate the direct effect of the duty hours themselves on key aspects of training and quality of life.

<table>
<thead>
<tr>
<th>Table I</th>
<th>Significant changes in satisfaction with specific elements of residency training, 2003-2004*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question</td>
<td>2003</td>
</tr>
<tr>
<td>Ability to pursue educational reading</td>
<td>2.35</td>
</tr>
<tr>
<td>Ability to research a given medical topic</td>
<td>3.39</td>
</tr>
<tr>
<td>Personal participation in research</td>
<td>3.04</td>
</tr>
<tr>
<td>Gynecologic surgical experience</td>
<td>3.13</td>
</tr>
<tr>
<td>Obstetric experience</td>
<td>4.35</td>
</tr>
<tr>
<td>Exposure to subspecialties</td>
<td>3.59</td>
</tr>
<tr>
<td>Attending supervision</td>
<td>3.73</td>
</tr>
<tr>
<td>Quality of hospital nursing staff</td>
<td>3.43</td>
</tr>
<tr>
<td>Amount of leisure time</td>
<td>2.43</td>
</tr>
</tbody>
</table>

* Stem wording: “Please rate your satisfaction with the following aspects of your program”: Items were scored on a 5-point scale, 1 = “very dissatisfied,” 5 = “very satisfied.”

<table>
<thead>
<tr>
<th>Table II</th>
<th>Effect of duty-hours restrictions on training and quality of life (n = 23)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>2003</td>
</tr>
<tr>
<td>Will residency improve?</td>
<td>3.96</td>
</tr>
<tr>
<td>Happier in residency?</td>
<td>3.87</td>
</tr>
<tr>
<td>More free time?</td>
<td>4.17</td>
</tr>
<tr>
<td>Happier in personal life?</td>
<td>4.04</td>
</tr>
<tr>
<td>Additional time to study?</td>
<td>3.74</td>
</tr>
<tr>
<td>Healthier lifestyle?</td>
<td>4.22</td>
</tr>
<tr>
<td>Improved patient care?</td>
<td>3.83</td>
</tr>
<tr>
<td>More difficult continuity of care?</td>
<td>3.74</td>
</tr>
<tr>
<td>Will be well-trained?</td>
<td>3.83</td>
</tr>
<tr>
<td>More time to teach?</td>
<td>3.39</td>
</tr>
<tr>
<td>More interest in teaching?</td>
<td>3.65</td>
</tr>
</tbody>
</table>

* Wording of these items changed from the 2003 to the 2004 surveys. In 2003 the wording was “After the institution of the 80-hour work week, I will...” (Example, “I will have more interest in teaching”; 1 = disagree, 5 = agree). In 2004 the wording was “Since the institution of the 80-hour work week, please rate the following:...” (Example: “Interest that I have in teaching”; 1 = worse/less, 5 = more/better).
We made several changes in the structure of the University of Colorado residency in response to the duty-hour requirements. Although our average work week was <80 hours before July 2003, some rotations, particularly in the first 2 years of training, were significantly out of compliance. Frequency of in-house night call was reduced, both by the addition of another night float resident in postgraduate year 1 and by the allocation of 2 previous in-house positions in the postgraduate 3 and 4 years to at-home call. The average work week for all residents, based on resident self-reporting, decreased from 73 to 66 hours.

The single outcome variable of “overall satisfaction with residency” did not change after the institution of duty-hour requirements. However, a composite average of responses to all other more specific elements did increase in a significant fashion. Nine elements in particular increased significantly. Of these 9, 3 elements related to academic issues (reading, reviewing the literature, and doing research) and 3 elements related to clinical training (obstetric, gynecologic, and subspecialty exposure). In our program, the amount of clinical experience did not increase; in fact, it most likely decreased. For example, many clinical services do not have a resident 1 day per week because of post-call work restrictions. Many gynecologic surgical procedures no longer have >1 resident participating, where previously 2 residents (one primary surgeon and one assistant) were able to take part in surgical procedures.

Our conclusion is that, despite static or declining volume of clinical material, residents feel better able to take advantage of their time with patients and with their supervising attending physicians. In addition, residents are very positive about their increased ability to partake of educational components of residency education (such as reading and research). This is reflected not only in increased ratings of satisfaction with these components but also with actual reports of increased time spent reading, which narrowly missed statistical significance.

One area of concern is certainly the decreased estimation by residents of their interest in teaching after the institution of duty-hour requirements. Again, our specialty is faced with the challenge of improving student recruitment into obstetrics and gynecology, and residents play a crucial role in teaching and mentoring students. Our residents report anecdotally that the need to cover clinical responsibilities and still meet the duty-hour requirements has made them less able to teach, and this perception seems to be borne out by our study. Further investigation of this issue to optimize the medical student experience in obstetrics and gynecology is certainly an important future goal.

This study has several important limitations. Although this residency program is larger than some, the absolute numbers of residents who were studied is small, which limits our ability to detect significant changes. However, expanding the survey to multiple residency programs would have introduced confounding factors such as different structures of residency programs and the differing effects of what are necessarily unique responses between different programs to the duty-hour challenge. This study measured only residents’ perceptions about the effect of the duty-hour requirements and does not measure absolute change in any of the areas that were studied. Nonetheless, this prospective cohort study raises several important areas for future investigation as obstetrics and gynecology educators aim for continuous improvement in trainee satisfaction, clinical and didactic education and skills, and medical student recruitment.

References

The influence of an audience response system on knowledge retention: An application to resident education

Archana Pradhan, MD, MPH,a,* Dina Sparano, BS,a Cande V. Ananth, PhD, MPHb

Department of Obstetrics, Gynecology, and Reproductive Sciences, Divisions of General Obstetrics and Gynecologya
Epidemiology and Biostatistics, b UMDNJ-Robert Wood Johnson Medical School, New Brunswick, NJ

Received for publication February 21, 2005; revised May 31, 2005; accepted July 19, 2005

Objective: The purpose of the study was to compare delivery methods of lecture material regarding contraceptive options by either traditional or interactive lecture style with the use of an audience response system with obstetrics and gynecology residents.

Study design: A prospective, randomized controlled trial that included 17 obstetrics and gynecology residents was conducted. Group differences and comparison of pre/posttest scores to evaluate efficacy of lecture styles were performed with the Student t test. Each participant completed an evaluation to assess usefulness of the audience response system.

Results: Residents who received audience response system interactive lectures showed a 21% improvement between pretest and posttest scores; residents who received the standard lecture demonstrated a 2% improvement (P = .018). The evaluation survey showed that 82% of residents thought that the audience response system was a helpful learning aid.

Conclusion: The results of this randomized controlled trial demonstrate the effectiveness of audience response system for knowledge retention, which suggests that it may be an efficient teaching tool for residency education.

© 2005 Mosby, Inc. All rights reserved.

Academic medical centers that are affiliated with medical schools are charged with the task of providing a strong educational base for future physicians. This is no small feat because the material to teach is vast and the time to teach is limited. Every medical specialty faces the challenge of developing a complete curriculum that promotes the professional growth of the medical student and the resident.

In the United States, most residency programs use a core curriculum of lectures to impart residents with a basic fund of knowledge. Education regarding contraception options is one women’s health topic that is important to teach effectively. Any resident who has chosen to specialize in obstetrics and gynecology must be able to educate their female patients regarding contraceptive options. Currently, there is neither a standardized curriculum for a core lecture on contraceptive choices nor an established method of delivering a
core lecture on contraceptive choices. Thus, there is both a need for a standardized curriculum regarding this issue and a need to evaluate whether the manner in which this body of knowledge imparted to residents is successful.

Alternative methods of delivering lectures are gaining ground among those in the medical community who are interested in improving the efficiency of programs because retention rates of traditional didactic lectures often prove less than desirable. The audience response system (ARS) uses wireless keypads for audience participation during lectures. Participants are able to respond to test questions and receive immediate feedback while the content is still the current topic of discussion. The instructor is able to see the responses immediately and can make changes to the instruction as needed to ensure that the audience fully comprehends the content.

A study by Schackow et al. was designed to measure the effectiveness of an ARS to help lecture-attendees retain information. The study at the Chicago-based family practice residency program demonstrated a significant difference in posttest scores between residents who received the lecture with the ARS versus the “traditional” format lecture. With the active participation required by an ARS and the straightforward presentation of key learning points that are typical of ARS lecture delivery, the lecture material is given much more staying power than before, which proves the values of including such technology in the medical education setting. The ARS maintains students’ anonymity, thus encouraging more active participation and more truthful responses than might otherwise be possible. Data collection is also more accurate because inputs are collected electronically and can be managed automatically with the accompanying software. Human error in transcribing written answers, losing hard copies of answer forms, and other such causes of lost data are eliminated.

Although studies show that learners prefer lectures which use the ARS, there is little data on whether the ARS can enhance knowledge retention in medicine. The purpose of this study was to compare delivery methods of lecture material to obstetrics and gynecology residents regarding contraceptive options by way of either traditional didactic lecture style or interactive lecture style with the use of an ARS. Specifically, the study was designed to evaluate the effectiveness of the teaching method on the long-term retention of the lecture material and to determine whether the method of delivery affected the residents’ satisfaction of the lecture by questionnaire.

### Material and methods

A prospective, randomized trial that was designed to compare the effect of 2 different teaching styles on knowledge retention was conducted from July through August of 2004. Participants included 17 obstetrics and gynecology residents at the UMDNJ-Robert Wood Johnson Medical School, NJ, who voluntarily consented to participate in this study. Residents were assigned randomly to attend either the contraception technology lecture or the interactive case-discussion class. Both lectures were delivered by the same instructor.

The 8 residents in group 1 received the interactive lecture; the 9 residents in group 2 were taught the same material in a standard didactic lecture format. A 15-minute pretest to establish resident baseline knowledge of contraceptive technology was administered to all residents before the presentation of either lecture. The ARS was used for group 1. Group 2 received the same lecture material in traditional lecture format. Approximately 6 weeks after the delivery of both lectures, all residents completed the same posttest evaluation to assess their knowledge of the material that was covered in the lectures. Also, each participant completed an evaluation form to assess his/her opinion regarding the usefulness of the ARS.

Each pre- and posttest consisted of 15 questions; 1 to 2 questions pertaining to each of the following areas were asked: epidemiology of contraception, basic science of contraception, oral contraception, transdermal contraception, contraception by intravaginal ring, injectable contraception, intrauterine devices, barrier contraception, emergency contraception, and natural family planning.

### Statistical analysis

The Student t test was used to compare pre- and posttest scores to evaluate efficacy of the intervention. No formal sample size analysis was performed because it was decided a priori to include all current obstetrics and gynecology residents on rotation. A probability value of <.05 was considered to denote statistical significance, and all statistical tests were 2-tailed.

### Results

The 8 residents who received the ARS interactive lecture scored a mean ± SD of 78 ± 1.41 on the pretest and a mean of 95 ± 1.60) on the post-test, which shows a 21% improvement between the 2 tests. The traditional didactic lecture group of 9 residents scored a mean of 80 ± 2.80 on the pretest and a mean of 82 ± 2.32 on the posttest, which demonstrates a 2% improvement in scores. These data show that residents who received the interactive lecture had significantly longer retention of the lecture material compared with the residents who received the traditional lecture (P = .018; Table).

The results of the evaluation survey revealed that only one half of the students in the traditional group felt they had a thorough understanding of contraceptive
Cwiak et al\(^4\) compared 2 methods of lecture delivery to changing needs of today's adult learners. A study by their methods of lecture delivery to accommodate the superior when educating the adult learner.’’

methods of teaching are considered by many to be et al,\(^2\) ‘‘Resident doctors are adult learners and interactive across disciplines and cultures. According to Homme a preferred mode of education, with popularity ranging the traditional lecture format as a means of effectively Previous studies consistently have illustrated failure in
technology after the lecture. The evaluation survey also showed that all 17 residents found the ARS easy to use and that 14 of the residents thought that the ARS was a helpful learning aid in the lecture (Table).

**Comment**

Previous studies consistently have illustrated failure in the traditional lecture format as a means of effectively teaching material for sustained application. Relevant case-based learning, on the other hand, has been cited as a preferred mode of education, with popularity ranging across disciplines and cultures. According to Homme et al,\(^2\) ‘‘Resident doctors are adult learners and interactive methods of teaching are considered by many to be superior when educating the adult learner.’’

Faculty members across the country are changing their methods of lecture delivery to accommodate the changing needs of today’s adult learners. A study by Cwiak et al\(^4\) compared 2 methods of lecture delivery to third-year medical students. Students were given either a didactic lecture or an interactive lecture that incorporated the tenets of problem-based learning. Findings showed an equivalent increase in posttest scores whether students received the didactic lecture or the interactive lecture. Satisfaction scores were also consistently above average and equivalent for the 2 lecture styles. In a similar study on diabetes mellitus and hypertension, although the medical students preferred the small group discussions, tests scores that compared small group discussions with traditional lectures were not significantly different between the 2 teaching methods.\(^5\)

The significance of our study is that it showed increased knowledge retention in the group that received the interactive ARS lecture compared with the group that received the traditional lecture format. With this technology, the instructor would stop periodically to ask the students a question regarding the material that had just been delivered. The students responded to these questions using individual electronic keypads, and the instructor was able to determine instantly the percentage of the class that answered correctly with the use of the corresponding computer program. The immediate feedback allowed the instructor to measure the effectiveness with which the material was being delivered and thus rephrase and repeat concepts when necessary. The ARS ‘‘combines many of the key components important for adult learners: interactivity; commitment to an answer from the learner; timely feedback, and the highlighting of key concepts with ‘take home points’ by the faculty expert.’’\(^2\) Perhaps, therein lies the explanation as to the reason that knowledge retention was affected in our study; our interactive lecture included the ARS as compared with previous studies that evaluated knowledge retention after interactive lecture alone. Our study also showed that residents find the ARS easy to use and consider it helpful to their learning. These data support the findings of other studies that compared interactive teaching with didactic teaching.\(^4,5\)

These results demonstrate the effectiveness of the ARS on knowledge retention and further suggest that the ARS may be a valuable and resourceful teaching tool for residency education. However, it is possible that the novelty of the experience may have lead to the increased attention to the lecture, a Hawthorne effect of sorts. It is possible that, if all lectures were given with the ARS, one would not see the increased knowledge retention. A future study, which will compare pre- and posttest scores in which the ARS is used for several different lectures given on same day, may clarify this issue.

The ARS is an innovative and creative educational tool that may help medical faculty become more efficient and effective educators. The ARS may not be widely available for use in all lecture settings (ie, an institution may not have enough keypads for each audience member to receive his or her own.) In this situation, the ARS could be used initially when the teacher is developing the lecture. The ARS could alert the lecturer to key points and concepts that require additional attention for retention by students.

With an ever-expanding body of material to learn and integrate into daily practice, all physicians are often challenged to find, on their own, a means by which to efficiently acquire new skills. Because of its effectiveness in enhancing knowledge retention and its well-received usefulness, we conclude that the ARS is a valuable education tool and that it should be considered for implementation into settings that span the educational spectrum.

**References**

1. Schackow TE, Chavez M, Loya L, Friedman M. Audience response system: effect on learning in family medicine residents. Fam Med 2004;36:496-504
Obstetrics and gynecology residents as teachers of medical students: Predictors of excellence

Joseph A. (Tony) Ogburn, MD,a,* Eve L. Espey, MD, MPH,a Maxine H. Dorin, MD,a Chen Ming, MS,b William F. Rayburn, MDa

Departments of Obstetrics and Gynecology a and Mathematics and Statistics, b University of New Mexico, Albuquerque, NM

Received for publication February 21, 2005; revised May 19, 2005; accepted July 19, 2005

KEY WORDS
Medical student
Resident
Teaching
Excellence

Objective: The purpose of this study was to assess variables that might predict which intern candidates will become excellent teachers of medical students.

Study design: This retrospective cohort study compared demographic characteristics, previous work experience, United States Medical Licensing Examinations scores, honors on core clerkships, membership in Alpha Omega Alpha, and match list ranking of 43 residents to identify predictors of excellent teaching evaluations during residency.

Results: Fifteen residents (35%) were identified as excellent teachers. They were more likely to have had previous work experience, to be older, or to be male. They were not more likely to have higher United States Medical Licensing Examinations test scores, more honors grades, Alpha Omega Alpha membership, or a higher rank list position.

Conclusion: Work experience, age, and male gender are associated with increased likelihood of being identified as an excellent teacher by medical students. Programs in which residents have a significant role as teachers of students may consider these factors in the residency selection process.

© 2005 Mosby, Inc. All rights reserved.

The significant role that is played by resident physicians in the education of medical students has been recognized for many years.1,2 The Liaison Committee for Medical Education has acknowledged the importance of resident teaching by including a resident training requirement as part of medical school accreditation standards.3 In its institutional requirements, the Accreditation Council on Graduate Medical Education (ACGME) specifies that residents participate fully in the educational and scholarly activities of the residency program, which include a major responsibility for teaching and supervision of medical students.4 The ACGME competency of practice-based learning and improvement requires that residents facilitate the learning of medical students.5 Almost 25% of resident activities are related to supervising, instructing, and evaluating medical students.
and junior residents. Medical students estimate that one third of their knowledge derives from resident teaching. On the most recent Association of American Medical Colleges graduate exit questionnaire, which is a national survey of all graduating medical students about their medical school experiences, students estimated that more than 50% of teaching was performed by residents and fellows on their internal medicine (58%), general surgery (55%), and obstetrics and gynecology (50%) rotations.

Despite the importance of resident teaching and the focus on successful teaching by national committees such as the Liaison Committee for Medical Education and ACGME, the ability to teach is typically not a criterion that is used to select residents. Instead, program directors focus on criteria such as United States Medical Licensing Examinations (USMLE) scores, grade point average, class rank, membership in Alpha Omega Alpha (AOA), and letters of recommendation. Because a major part of resident success depends on being an effective educator, we sought to identify applicant characteristics that would predict the ability to teach. The objective of this study was to determine whether we could identify intern applicants who, after becoming residents, would more likely be rated as excellent teachers by medical students and what specific traits correlated with teaching excellence.

**Material and methods**

Subjects included all 43 obstetrics and gynecology residents between 1997 and 2004 who completed at least 2 years of residency at the University of New Mexico. Before 1997, students uniformly did not evaluate resident teaching; therefore, our sample was limited to residents after that time. No structured training of residents to become more effective teachers was conducted during this time period. The University of New Mexico Human Research Review Committee reviewed and approved the study.

Demographic characteristics, measures of academic performance, work history, and rank list position were recorded for all subjects. All information came from resident personnel files and was analyzed in the aggregate only. Work experience was defined as a period of at least 1 year during which the subject was not enrolled in school and was employed full-time. Type of work experience was not considered. Demographic characteristics included age at onset of residency and gender. Measures of academic performance included USMLE scores, membership in AOA, and honors grades in third-year clinical clerkships.

Given our average medical school class of 74 graduating students, approximately 10 to 13 students rotate in each clerkship group. At the conclusion of each 8-week clerkship, the medical students anonymously evaluate all residents. The evaluation form directs the student to rate each resident on a scale of 1 (below average) to 4 (excellent) for the following 4 items: (1) communicates clearly and effectively, (2) is an enthusiastic and stimulating teacher, (3) models professional behavior, and (4) serves as a role model. Residents who achieve an annual composite score of ≥3.5 are considered excellent teachers. This score was chosen because it reflects a generally accepted standard of excellence. We have used it successfully for 7 years to select faculty and residents for teaching excellence awards. Data were entered and analyzed with SPSS software (version 10.1; SPSS Inc, Chicago, IL). Logistic regression, Spearman rank order correlation, and chi-squared test were used to determine differences between excellent resident teachers and teachers who did not meet the criteria for excellence. We reported 95% CIs and considered a probability value of <.05 to be statistically different.

**Results**

We received approximately 200 applications for our residency program annually during the study period and interviewed approximately 60 candidates from across the United States. Few foreign-trained or osteopathic students were interviewed. The 6 interns who were chosen each year were typically non-Hispanic white or Hispanic and ranged in age from 25 to 37 years. Approximately 75% of each intern class was female.

Information was available for all 43 residents. Fifteen of the residents (35%) met the criteria for being designated excellent teachers. Comparisons of residents who were excellent teachers with those residents who were not excellent teachers in the areas of previous work experience, gender, honors in obstetrics clerkship, honors in most clerkships, and membership in AOA are summarized in the Table. Applicants with previous work experience were more likely than the others to be excellent teachers (odds ratio, 8.4; 95% CI, 1.9-37.6). Men were more likely than women to be identified as excellent teachers (odds ratio, 5.6; 95% CI, 1.1-27.0). Older residents were more likely to be excellent teachers (P = .04). There was no association between gender and previous work experience (P = .72) or gender and age (P = .11). Previous work experience was found to be an independent predictor of excellence after controlling for the effect of age (P = .01).

Higher USMLE step 1 score (P = .35), USMLE step 2 score (P = .21), or higher rank list position (P = .94) were not predictive of being rated an excellent teacher.

Only 1 of the 15 excellent teachers had previous work experience in an actual teaching position. That individual served as a teacher of English as a second language. The other excellent teachers had varied types of employment, but most of the positions involved significant direct contact with people.
Residents may enter their training programs as excellent teachers or they may develop these skills during their training. A number of studies have assessed the impact of a training program’s ability to improve the teaching skills of residents with conflicting results. A study of obstetrics and gynecology residents found that a 1-time 4.5-hour teaching program improved residents’ self-assessment of teaching skills but did not affect student evaluations of residents. A 6-hour teaching program in an internal medicine department resulted in a significant increase in resident teaching scores as reported by medical students. Overall, a significant impact of training programs on the ability to teach has not been documented consistently.

Although trying to identify potential excellent teachers had not been part of our selection process, medical student education is emphasized in our residency program. During residency interviews, applicants are informed that teaching will be a major part of their responsibilities. Our intern orientation includes a 2-hour teaching session that provides an overview of the medical student curriculum, goals and objectives for the third-year medical student clerkship, teaching tips, and an overview of the evaluation of medical students. Our educational program includes 2 to 3 teaching workshops that are focused on student education and that are attended by all residents each year. Residents evaluate students at the mid point of each clerkship with a dedicated 1-hour review session that uses the Reporter-Interpreter-Manager-Educator framework. These sessions provide not only an evaluation of student performance but also develop resident skills in several aspects of teaching, which includes evaluation and feedback. Finally, we recognize all residents who were identified as excellent teachers with an award at our annual resident banquet.

Our study has several potential limitations. We used a 4-item questionnaire that can be completed quickly and easily for students to rank all residents and faculty members. More detailed questionnaires have been described. Although our questionnaire is not validated, it does include important domains that are covered in more detailed questionnaires that include professionalism, commitment to teaching, and communication skills. We also did not investigate some variables that may be associated with excellence in teaching that include previous research experience or subjective personality traits. Because of the small numbers of subjects and the wide variety of work experiences, we were not able to determine whether any certain type of work experience was more predictive of being an excellent teacher. Additionally, our sample was limited to 43 subjects in obstetrics and gynecology. A larger, multicentered study is being considered to confirm these findings.

Excellent resident teachers may improve the performance of learners. Subjectively, we believe that excellent

| Table Comparisons of characteristics of residents regarding excellence in teaching |
|---------------------------------|---------------------------------|-------------------|
| Characteristic                  | Proportion of excellent teachers (%) | Odds ratio for being excellent teachers (95% CI) | P value |
| Work experience                 |                                  |                                 |         |
| Yes                             | 57.1                             | 8.4 (1.9, 37.6) | .004    |
| No                              | 13.6                             |                                 |         |
| Gender                          |                                  |                                 |         |
| Male                            | 66.7                             | 5.6 (1.1, 27.0) | .046    |
| Female                          | 26.5                             |                                 |         |
| Honors in obstetrics            |                                  |                                 |         |
| Yes                             | 34.5                             | 0.9 (0.2, 3.6)  | .937    |
| No                              | 35.7                             |                                 |         |
| Honors in majority clerkships   |                                  | 0.8 (0.2, 2.7)  | .686    |
| Yes                             | 31.6                             |                                 |         |
| No                              | 37.5                             |                                 |         |
| Membership in AOA               |                                  |                                 |         |
| Yes                             | 50.0                             | 2.1 (0.4, 11.9) | .402    |
| No                              | 32.4                             |                                 |         |

Comment

Our study indicates that residents who have previous work experience and who are older are more likely to be rated as excellent teachers by medical students. We were unable to locate other investigations that examined any association between work experience or age and the rating of teaching. It appears that work experience is a predictor of teaching excellence that is independent of the effect of age. Although increasing age may be responsible for some component of increased excellence in teaching, we believe it is likely that the life experience that is gained by working contributes to increased maturity and tolerance. These traits may improve an individual’s effectiveness as a teacher.

We were interested to find that male gender was associated with a higher likelihood of being identified as an excellent teacher. Because of the small number of men in the sample, the actual significance of this finding is unclear. Male applicants who are attracted to obstetrics and gynecology may possess traits, however, that make them more effective teachers. A larger study would be necessary to confirm this finding.

Although academic performance has been evaluated for its ability to predict success in residency, it has not been evaluated for its role in the prediction of teaching excellence. Although certain measures of academic achievement are associated with successful passage of licensing examinations and specialty boards, we found no association between any academic measure and teaching excellence.
resident teachers may improve medical student perception of obstetrics and gynecology as a specialty and as a career choice. Because fewer medical students currently are choosing our medical specialty as a career, it is critical that residents be viewed favorably and with respect.

The selection of residents who are likely to become excellent teachers should be a goal of residency program directors. On the basis of the findings of this study, we have incorporated the consideration of previous work experience into our intern candidate selection process. Program directors should continue to recruit male students into the specialty.

Acknowledgments

We thank Cristina Beraun and Cindy Garcia for their assistance in the collection of data.

References

3. Liaison Committee for Medical Education. Functions and structure of a medical school: Accreditation and the Liaison Committee on Medical Education, Standards for the Accreditation of Medical Education Programs Leading to the MD degree. Washington, DC, and Chicago, IL: Association of American Medical Colleges and the American Medical Association; 2000.
A multicenter study to determine motivating factors for residents pursuing obstetrics and gynecology

May Hsieh Blanchard, MD, Amy M. Autry, MD, Haywood L. Brown, MD, John R. Musich, MD, Leah Kaufman, MD, Dylan R. Wells, MD, Robert D. Stager, MD, Jennifer L. Swanson, MD, Kirsten J. Lund, MD, Donald W. Wiper III, MD, Jennifer L. Bailit, MD, MPH

Case Western Reserve University-MetroHealth Medical Center/Cleveland Clinic Foundation, Cleveland, OH; University of California, San Francisco, CA; Duke University, Durham, NC; William Beaumont Hospital, Royal Oak, MI; Long Island Jewish Medical Center, New Hyde Park, NY; University of Tennessee, Memphis, TN; Medical College of Georgia, Augusta, GA; Western Pennsylvania Hospital, Pittsburgh, PA; University of Colorado Health Sciences Center, Denver, CO; Maine Medical Center, Portland, ME; Case Western Reserve University-MetroHealth Medical Center, Cleveland, OH

Received for publication February 21, 2005; revised July 1, 2005; accepted July 19, 2005

KEY WORDS
Obstetrics and gynecology residency
Student interest
Specialty choice
Motivators

Objective: This study was undertaken to determine why residents choose obstetrics and gynecology.

Study design: Applicants to obstetrics and gynecology residency programs were surveyed; a 5-point scale (5 = most important) was used to rate various aspects of the specialty. Univariate statistics were performed. Bivariate analysis comparing results that were based on gender and timing of decisions was completed with Student t test, \( \chi^2 \), and Kruskal-Wallis tests.

Results: A total of 153 applicants (42% response rate) from 10 programs participated; 85.3% of respondents were female. Surgical opportunities, variety of clinical experience, and fast-paced/high-acuity experiences attract applicants to obstetrics and gynecology. When considering programs, resident camaraderie, gynecologic experience, and commitment to education were most important. Over 70% of residents decided to pursue obstetrics and gynecology during or after their third-year clerkship.

Conclusion: Surgical opportunities and clinical variety appeal to applicants. The majority choose obstetrics and gynecology during or after their core clerkship. In addition, program dynamics are important when choosing a residency.

© 2005 Mosby, Inc. All rights reserved.

Interest in obstetrics and gynecology, as evidenced by the percentage of US medical school graduates entering the first year at a residency program (PGY-1), peaked in 1993 at 7.9%. Since that time, there has been a continued decline in the number of US graduates pursuing training in obstetrics and gynecology. In 2004, 5.5% (743/13,572) of US seniors participating in the National Resident Matching Program (NRMP) match successfully applied to obstetrics and gynecology PGY-1.
positions. However, only 65.1% of the 1,142 obstetrics and gynecology PGY-1 positions offered in the 2004 match were filled by graduating US medical students.\(^2\) These statistics reflect an ongoing decline in interest among graduating US medical students for the field of obstetrics and gynecology. Proposed reasons for this decline have included the liability crisis, decreased reimbursement, and unpredictability of lifestyle. There has been national attention drawn to this situation. Initiatives among obstetrics and gynecology educator (CREOG [The Council on Resident Education in Obstetrics and Gynecology]; APGO [The Association of Professors of Gynecology and Obstetrics]) and membership (ACOG [The American College of Obstetricians and Gynecologists]) organizations have issued directives toward focused recruitment into the specialty, promoting the positive aspects of the field.\(^3\)

Despite the above-mentioned trends, there continue to be substantial numbers of medical students who pursue training in obstetrics and gynecology. These students can provide insight regarding opportunities to optimize recruitment efforts.

In this study, we sampled applicants to obstetrics and gynecology residency programs from across the country. Our objective was to identify those students (ie, those who applied to residency programs in obstetrics and gynecology via the 2004 NRMP match) who were NOT deterred by the negative aspects identified with the field of obstetrics and gynecology, and to determine why they still choose training in this specialty. Our secondary aim was to evaluate for regional, as well as program size and affiliation predilections among these applicants. Ultimately, we hoped to elucidate which aspects of the specialty are appealing to current medical students, what helps applicants differentiate among the various residency programs, and how these findings are affected by applicant gender or timing of decision to enter the field.

**Material and methods**

In this multicenter survey, we evaluated motivating factors for choice to pursue residency training in obstetrics and gynecology, as well as differentiating factors (as perceived by applicants) among residency programs. Ten obstetrics and gynecology residency programs participated in this survey. Institutional Review Board approval was obtained at each site.

Surveys were sent to medical students applying for a first-year residency position in each of the programs, based on the 2004 NRMP rank list for each program. Duplicate applicants among the 10 programs’ rank lists were identified. To ensure that each applicant was surveyed only once, those applicants were then assigned to 1 of the programs that had the applicant on its rank list, but duplicates were preferentially distributed to the programs that had lower volumes of surveys to be generated (in attempts at equalizing the number of surveys dispersed by each study site). An applicant was to receive the survey only from a program where he/she had applied/interviewed. Each program sent 2 forms (e-mail, US mail) of the survey to each of the remaining applicants on its list. Surveys were mailed to applicants after the 2004 NRMP match results were made available. Each survey was labeled with a unique code indicating both the program generating the survey and the recipient; this code was used for response tracking purposes only, and was discarded after the completed survey was received. The cover letter accompanying the survey was generated and signed by the individual study site director, and included instructions indicating that participation was voluntary, that the survey was part of a multicenter research study, that the identifier code was to be used for tracking response to request for participation, and that only 1 form of the survey should be returned. The tracking code was also used to guard against duplicate responses.

Respondents were asked about their choice of obstetrics and gynecology as a specialty. The survey listed various aspects of a career in obstetrics and gynecology (surgical/obstetric/research/primary care opportunities, patient type, variety of clinical experiences in day-to-day practice, lifestyle factors) and asked respondents to rank each using the 5-point scale (1 = low importance, 5 = high importance) as to which aspects were deemed to be appealing. Respondents were queried regarding the timing of their decision to pursue obstetrics and gynecology, activities/experiences during medical school that contributed to this choice, and other fields that were seriously considered other than obstetrics and gynecology.

The survey also queried respondents as to considerations in choosing an obstetrics and gynecology residency program. Respondents were given a list of factors pertaining to residency programs (location; type of program, ie, university-based vs community-based; educational experiences during residency/afforded by the program; faculty size and dynamics; historic post-residency placement success; size of program; “fringe” benefits), and asked to use a 5-point scale to rank the importance of each factor in deciding which obstetrics and gynecology residency program to place on their NRMP rank list. The respondent was then presented with a list of recruitment tools that are used by residency programs (eg, Web site, preinterview dinner/social, interview with faculty, interview with residents, follow-up visits) and asked to evaluate their usefulness in helping the candidate to decide whether and how to rank a particular obstetrics and gynecology program (5-point importance scale).

Survey participants were given the option to provide demographic information. In addition, applicants were asked to provide information as to where on their rank list they matched, their satisfaction with the outcome of
To protect participant anonymity, surveys sent by US mail included a stamped return envelope addressed to the primary study site (Cleveland, OH); surveys sent by e-mail included the e-mail address of the primary study site director.

Demographic characteristics of the survey population were described using ranges, means, and percentages, where appropriate. Questions that used a 5-point scale were reported with means and SDs. All other questions were described using percentages. Responses were then analyzed by gender using Student’s t test and χ² test, where appropriate. Lastly, questions were evaluated by timing of candidate decision to pursue training in obstetrics and gynecology (before the core clerkship, during the core clerkship, or after the core clerkship) with the use of the Kruskal-Wallis test.

Results

Surveys were sent to 367 applicants and 153 (42%) responded. The average respondent age was 27.7 years (range: 24-48 years); 90% were 31 years or younger. Of the 153 respondents, 129 (84.3%) were female, and 21 (13.7%) were male (3 respondents did not identify gender). Self-reported race/ethnic categories included white (75.2%), black (10.5%), and Asian or Pacific Islander (8.5%). Six percent were of other racial/ethnic groups, did not wish to give racial/ethnicity information, or did not respond to this question.

Study site residency programs varied in geographic location, size (number of residents, number of hospitals), affiliation (university- vs community-based), and patient population (Table I). Of the respondents, 30.7% had no plans for subspecialty fellowship training in obstetrics and gynecology (1.3% nonrespondents). Those who desired postresidency training reported interest in reproductive endocrinology (20.3%) most frequently (gynecologic oncology [11.8%], maternal-fetal medicine [7.8%], urogynecology [5.2%], other [22.9%]). Practice plans (postresidency/fellowship) were as follows: academic/university based (37.2%), community (20.2%), undecided (39.9%), other (2%), and nonrespondent (0.7%).

Participants were asked to rate the most appealing aspects of the specialty of obstetrics and gynecology. Respondents reported surgical opportunities (mean score 4.7), variety of clinical experience in day-to-day practice (office/surgery/labor unit) (4.6), and fast-paced/high-acuity experiences (4.3) to be of highest importance (Table II). The opportunity to be a women’s health advocate or to have influence on public health was more important to women than men (4.3 vs 3.1, and 4.0 vs 2.9, respectively; \( P < .001 \) for each). Caring specifically for female patients was an aspect of obstetrics and gynecology that women (3.7) found more important than men.

<table>
<thead>
<tr>
<th>Study site</th>
<th>Location</th>
<th>No. residents per y</th>
<th>Affiliation</th>
<th>No. hospitals covered</th>
<th>No. beds/hospital</th>
<th>Insurance*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case Western Reserve University-MetroHealth Medical Center/Cleveland Clinic Foundation</td>
<td>Cleveland, OH</td>
<td>7</td>
<td>University</td>
<td>2</td>
<td>750</td>
<td>35/55/10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1000</td>
<td>100/0/0</td>
</tr>
<tr>
<td>University of California, San Francisco</td>
<td>San Francisco, CA</td>
<td>8</td>
<td>University</td>
<td>2</td>
<td>588</td>
<td>47/51/2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>686</td>
<td>14/56/30</td>
</tr>
<tr>
<td>Duke University</td>
<td>Durham, NC</td>
<td>7</td>
<td>University</td>
<td>4</td>
<td>1019</td>
<td>50/46/4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>369</td>
<td>24/42/35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>426</td>
<td>270</td>
</tr>
<tr>
<td>William Beaumont Hospital</td>
<td>Royal Oak, MI</td>
<td>6</td>
<td>Community</td>
<td>1</td>
<td>997</td>
<td>55/45/0</td>
</tr>
<tr>
<td>Long Island Jewish Medical Center</td>
<td>New Hyde Park, NY</td>
<td>5</td>
<td>University</td>
<td>1</td>
<td>829</td>
<td>60/30/10</td>
</tr>
<tr>
<td>University of Tennessee</td>
<td>Memphis, TN</td>
<td>10</td>
<td>University</td>
<td>5</td>
<td>500</td>
<td>10/90/10</td>
</tr>
<tr>
<td>Medical College of Georgia</td>
<td>Augusta, GA</td>
<td>4</td>
<td>University</td>
<td>1</td>
<td>450</td>
<td>40/45/15</td>
</tr>
<tr>
<td>Western Pennsylvania Hospital</td>
<td>Pittsburgh, PA</td>
<td>3</td>
<td>Community</td>
<td>1</td>
<td>524</td>
<td>36/60/4</td>
</tr>
<tr>
<td>University of Colorado</td>
<td>Denver, CO</td>
<td>9</td>
<td>University</td>
<td>3</td>
<td>450</td>
<td>50/50/0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>398</td>
<td>0/100/0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>420</td>
<td>100/0/0</td>
</tr>
<tr>
<td>Maine Medical Center</td>
<td>Portland, ME</td>
<td>3</td>
<td>Community</td>
<td>1</td>
<td>598</td>
<td>40/56/4</td>
</tr>
</tbody>
</table>

* % payer mix (commercial/medicare-medicaid/self-pay).
In the match outcomes, 66.5% of applicants matched at the residency program ranked first on their rank order list, and 92.1% matched in one of their top three choices. Applicants were typically happy with their match outcome (4.7). Applicants ranked an average of 8.9 programs. Of the respondents, 76.5% listed at least 10 programs on their rank list.

When asked about positive influences toward choice of a career in obstetrics and gynecology, the third-year clerkship (4.4), fourth-year–related electives (4.5), and mentors in the field of obstetrics and gynecology (3.7) were of high importance; the dean of the medical school was noted to be of low importance (1.4). Within the third-year obstetrics and gynecology core clerkship, aspects deemed to have contributed to the decision to pursue obstetrics and gynecology included hands-on opportunities (4.5), operative experience (4.5), and delivering infants (4.2) (Table V).

Most respondents (83.0%) seriously considered one or more other residency field, citing general surgery (31.4%), internal medicine (30.7%), and pediatrics (22.0%) most frequently. However, only 2 respondents reported actually ranking a specialty other than obstetrics and gynecology. Of note, 34.0% of students made their decision to become an obstetrician-gynecologist during their obstetrics and gynecology core clerkship. Nonetheless, many respondents (38.6%) did not finalize their choice to pursue the field of obstetrics and gynecology until later in their third year (after completing their core clerkship) or in the fourth year of medical school. In evaluation of respondent groups based on timing of their decision to become an obstetrician-gynecologist (before the obstetrics and gynecology core clerkship, during the core clerkship, after completion of

### Table II

<table>
<thead>
<tr>
<th>Factors</th>
<th>Mean ± SD</th>
<th>Male</th>
<th>Female</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgical opportunities</td>
<td>4.7 ± 0.6</td>
<td>4.9 ± 0.1</td>
<td>4.6 ± 0.1</td>
<td>.11</td>
</tr>
<tr>
<td>Obstetric opportunities</td>
<td>4.2 ± 1.1</td>
<td>3.9 ± 0.3</td>
<td>4.2 ± 0.1</td>
<td>.27</td>
</tr>
<tr>
<td>Primary care opportunities</td>
<td>3.4 ± 1.3</td>
<td>2.8 ± 1.2</td>
<td>3.5 ± 1.3</td>
<td>.02</td>
</tr>
<tr>
<td>Research opportunities</td>
<td>2.6 ± 1.2</td>
<td>2.4 ± 0.3</td>
<td>2.7 ± 0.1</td>
<td>.42</td>
</tr>
<tr>
<td>Academic opportunities</td>
<td>3.1 ± 1.1</td>
<td>2.7 ± 3.2</td>
<td>3.2 ± 1.1</td>
<td>.04</td>
</tr>
<tr>
<td>Postresidency fellowship opportunities</td>
<td>3.4 ± 1.3</td>
<td>3.9 ± 0.3</td>
<td>3.4 ± 0.1</td>
<td>.07</td>
</tr>
<tr>
<td>Healthy patients</td>
<td>3.6 ± 1.1</td>
<td>3.5 ± 0.2</td>
<td>3.7 ± 0.1</td>
<td>.002</td>
</tr>
<tr>
<td>Variety of patient problems</td>
<td>4.0 ± 1.0</td>
<td>3.8 ± 0.3</td>
<td>4.1 ± 0.1</td>
<td>.30</td>
</tr>
<tr>
<td>Female patients</td>
<td>4.1 ± 1.2</td>
<td>2.9 ± 1.3</td>
<td>3.7 ± 1.1</td>
<td>.002</td>
</tr>
<tr>
<td>Variety of clinical experiences in day-to-day practice</td>
<td>4.6 ± 0.7</td>
<td>4.6 ± 0.2</td>
<td>4.6 ± 0.1</td>
<td>.84</td>
</tr>
<tr>
<td>Fast-paced/high-acuity experiences (obstetrics/operating room)</td>
<td>4.3 ± 0.8</td>
<td>4.0 ± 1.1</td>
<td>4.4 ± 0.8</td>
<td>.04</td>
</tr>
<tr>
<td>Opportunity to be a women’s health advocate</td>
<td>4.1 ± 1.1</td>
<td>3.1 ± 1.3</td>
<td>4.3 ± 1.0</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Opportunity to have influence on public health</td>
<td>3.9 ± 1.2</td>
<td>2.9 ± 1.4</td>
<td>4.0 ± 1.1</td>
<td>.0001</td>
</tr>
<tr>
<td>Potential income</td>
<td>2.6 ± 1.0</td>
<td>2.5 ± 0.2</td>
<td>2.6 ± 0.1</td>
<td>.68</td>
</tr>
<tr>
<td>Lifestyle</td>
<td>2.0 ± 1.0</td>
<td>1.8 ± 0.2</td>
<td>2.1 ± 0.1</td>
<td>.26</td>
</tr>
<tr>
<td>Prestige</td>
<td>2.2 ± 1.0</td>
<td>2.0 ± 0.2</td>
<td>2.2 ± 0.1</td>
<td>.57</td>
</tr>
</tbody>
</table>

OR, Operating room; L&D, labor and delivery.

(2.9); $P = .002$. Respondents of both genders reported lifestyle (2.0), prestige (2.2), and potential income (2.6) to be of lowest importance when considering the appeal of a career in obstetrics and gynecology.

In evaluation of obstetrics and gynecology residency programs, the most important factors to applicants when considering whether to rank a particular program were: resident camaraderie (4.6), gynecologic experience afforded by the program (4.6), faculty commitment to resident education (4.5), faculty “approachability” (4.4), and responsiveness of the program/faculty/program director to resident concerns (4.4) (Table III). Resident gender mix was significantly more important to male respondents (3.5 vs 2.6; $P = .002$). Female respondents placed higher importance on the primary care experience afforded by the program (3.0 vs 2.3; $P = .01$), having family/friends in the area of a given program (3.7 vs 2.9; $P = .02$), and the program’s leave (maternity, the Family and Medical Leave Act [FMLA]) policy (2.4 vs 1.5; $P = .001$) as compared with the male respondents.

In choosing obstetrics and gynecology residency programs to rank, moonlighting opportunities (1.5), salary (2.1), and paid conference opportunities (2.1) were reported to be of lowest importance for applicants.

Regarding recruitment methods used by residency programs, applicants reported favorable impressions toward those programs that allowed for opportunities to interview with residents (4.3), opportunities to interview with faculty other than the program director and chairperson (4.1), and one-on-one faculty interviews (4.1), as opposed to “group” interviews (1 faculty member interviewing multiple applicants [1.9] or multiple faculty members interviewing 1 applicant [2.2]) (Table IV).
Table III  Factors differentiating among residency programs in obstetrics and gynecology. Which of the following were your strongest considerations in choosing obstetric/gynecologic residency programs to rank? (Please rate each of the following factors individually on a scale of 1-5; 1 = low importance, 5 = high importance.)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Mean ± SD</th>
<th>Male</th>
<th>Female</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>City/geographic location of program</td>
<td>4.3 ± .9</td>
<td>4.2 ± .3</td>
<td>4.4 ± .1</td>
<td>.37</td>
</tr>
<tr>
<td>Family/friends in the area</td>
<td>3.6 ± 1.4</td>
<td>2.9 ± 1.5</td>
<td>3.7 ± 1.4</td>
<td>.02</td>
</tr>
<tr>
<td>University-based vs community-based program</td>
<td>4.1 ± 1.1</td>
<td>3.9 ± 0.3</td>
<td>4.1 ± 0.1</td>
<td>.40</td>
</tr>
<tr>
<td>Reputation of affiliated university (if applicable)</td>
<td>3.5 ± 1.1</td>
<td>3.5 ± 0.3</td>
<td>3.5 ± 0.1</td>
<td>.88</td>
</tr>
<tr>
<td>Reputation of hospital(s) covered by the program</td>
<td>3.6 ± 1.1</td>
<td>3.7 ± 0.3</td>
<td>3.7 ± 0.1</td>
<td>.75</td>
</tr>
<tr>
<td>Number of hospitals covered by the residency</td>
<td>2.8 ± 1.2</td>
<td>2.7 ± 0.3</td>
<td>2.7 ± 0.1</td>
<td>.96</td>
</tr>
<tr>
<td>Obstetric experience afforded by program</td>
<td>4.1 ± 1.0</td>
<td>3.8 ± 0.3</td>
<td>4.2 ± 0.1</td>
<td>.10</td>
</tr>
<tr>
<td>Gynecologic experience afforded by program</td>
<td>4.6 ± 0.7</td>
<td>4.8 ± 0.1</td>
<td>4.5 ± 0.1</td>
<td>.22</td>
</tr>
<tr>
<td>Primary care experience afforded by program</td>
<td>2.9 ± 1.2</td>
<td>2.3 ± 1.1</td>
<td>3.0 ± 1.1</td>
<td>.01</td>
</tr>
<tr>
<td>Research opportunities during residency</td>
<td>3.0 ± 1.1</td>
<td>2.9 ± 0.3</td>
<td>3.0 ± 0.1</td>
<td>.51</td>
</tr>
<tr>
<td>Exposure to subspecialties during residency</td>
<td>4.0 ± 0.9</td>
<td>4.0 ± 0.1</td>
<td>4.0 ± 0.1</td>
<td>.90</td>
</tr>
<tr>
<td>Reputation of program</td>
<td>4.3 ± 0.8</td>
<td>4.1 ± 0.3</td>
<td>4.3 ± 0.1</td>
<td>.24</td>
</tr>
<tr>
<td>Reputation of program director</td>
<td>3.1 ± 1.1</td>
<td>2.6 ± 0.2</td>
<td>3.1 ± 0.1</td>
<td>.07</td>
</tr>
<tr>
<td>Reputation of departmental chairperson</td>
<td>3.0 ± 1.1</td>
<td>2.7 ± 0.2</td>
<td>3.1 ± 0.1</td>
<td>.18</td>
</tr>
<tr>
<td>Stability of faculty, director (ie, low &quot;turn-over&quot;)</td>
<td>3.7 ± 1.1</td>
<td>3.7 ± 0.3</td>
<td>3.7 ± 0.1</td>
<td>.96</td>
</tr>
<tr>
<td>Faculty size</td>
<td>3.1 ± 1.0</td>
<td>2.8 ± 0.3</td>
<td>3.1 ± 0.1</td>
<td>.17</td>
</tr>
<tr>
<td>Faculty gender mix</td>
<td>2.5 ± 1.1</td>
<td>2.4 ± 0.3</td>
<td>2.6 ± 0.1</td>
<td>.54</td>
</tr>
<tr>
<td>Faculty specialty mix</td>
<td>3.4 ± 1.1</td>
<td>3.5 ± 0.2</td>
<td>3.3 ± 0.1</td>
<td>.64</td>
</tr>
<tr>
<td>Faculty commitment to resident education</td>
<td>4.5 ± 0.7</td>
<td>4.4 ± 0.2</td>
<td>4.5 ± 0.1</td>
<td>.29</td>
</tr>
<tr>
<td>Faculty “approachability”</td>
<td>4.4 ± 0.8</td>
<td>4.4 ± 0.2</td>
<td>4.4 ± 0.1</td>
<td>.90</td>
</tr>
<tr>
<td>Call schedule</td>
<td>3.1 ± 1.1</td>
<td>2.8 ± 0.3</td>
<td>3.2 ± 0.1</td>
<td>.11</td>
</tr>
<tr>
<td>Didactics, protected time</td>
<td>3.7 ± 0.9</td>
<td>3.4 ± 0.2</td>
<td>3.7 ± 0.1</td>
<td>.06</td>
</tr>
<tr>
<td>Boards (written/oral) pass rate</td>
<td>3.2 ± 1.2</td>
<td>3.3 ± 0.2</td>
<td>3.2 ± 0.1</td>
<td>.79</td>
</tr>
<tr>
<td>Postresidency job/practice placement</td>
<td>3.6 ± 1.1</td>
<td>3.5 ± 0.3</td>
<td>3.6 ± 0.1</td>
<td>.78</td>
</tr>
<tr>
<td>Postresidency fellowship competitiveness</td>
<td>3.8 ± 1.1</td>
<td>4.2 ± 0.2</td>
<td>3.8 ± 0.1</td>
<td>.14</td>
</tr>
<tr>
<td>Couples match considerations</td>
<td>1.6 ± 1.4</td>
<td>1.8 ± 0.4</td>
<td>1.6 ± 0.1</td>
<td>.60</td>
</tr>
<tr>
<td>Nonmedical job/school opportunities for partner</td>
<td>2.2 ± 1.6</td>
<td>2.0 ± 0.3</td>
<td>2.3 ± 0.2</td>
<td>.47</td>
</tr>
<tr>
<td>Program size/number of residents</td>
<td>3.8 ± 1.0</td>
<td>3.9 ± 0.3</td>
<td>3.8 ± 0.1</td>
<td>.86</td>
</tr>
<tr>
<td>Resident gender mix</td>
<td>2.7 ± 1.2</td>
<td>3.5 ± 1.1</td>
<td>2.6 ± 1.2</td>
<td>.002</td>
</tr>
<tr>
<td>Resident camaraderie</td>
<td>4.6 ± 0.8</td>
<td>4.6 ± 0.2</td>
<td>4.6 ± 0.1</td>
<td>.92</td>
</tr>
<tr>
<td>Responsiveness of program/faculty/director to resident concerns</td>
<td>4.4 ± 0.8</td>
<td>4.3 ± 0.2</td>
<td>4.4 ± 0.1</td>
<td>.36</td>
</tr>
<tr>
<td>Salary</td>
<td>2.1 ± 1.1</td>
<td>2.5 ± 0.3</td>
<td>2.0 ± 0.1</td>
<td>.08</td>
</tr>
<tr>
<td>Paid conference opportunities</td>
<td>2.1 ± 1.0</td>
<td>2.5 ± 0.3</td>
<td>2.0 ± 0.1</td>
<td>.09</td>
</tr>
<tr>
<td>Perks (meals, parking, facilities)</td>
<td>2.5 ± 1.1</td>
<td>2.8 ± 0.2</td>
<td>2.5 ± 0.1</td>
<td>.26</td>
</tr>
<tr>
<td>Availability/quality of ancillary support staff (laboratories, phlebotomy, transport)</td>
<td>2.8 ± 1.2</td>
<td>2.9 ± 0.3</td>
<td>2.8 ± 0.1</td>
<td>.93</td>
</tr>
<tr>
<td>Leave (maternity, FMLA) policy</td>
<td>2.3 ± 1.2</td>
<td>1.5 ± 0.8</td>
<td>2.4 ± 1.2</td>
<td>.001</td>
</tr>
<tr>
<td>Moonlighting opportunities</td>
<td>1.5 ± 0.8</td>
<td>1.5 ± 0.1</td>
<td>1.5 ± 0.1</td>
<td>.84</td>
</tr>
</tbody>
</table>

the clerkship), there was a significant difference among those 3 groups in the importance of obstetric opportunities (4.6, 4.1, 3.9, respectively; P = .01) when choosing a career in obstetrics and gynecology.

Comment

In our study of applicants to 10 obstetrics and gynecology residency programs, we selected a population already interested in pursuing the field to elucidate what aspects of the field are attractive and presently effective. This investigation is unique in that this multicenter study involved residency programs of varying sizes from across the country, both academic- and community-based, and sampled those who had already expressed the commitment (ie, interest) to pursuing training in obstetrics and gynecology, thus enabling us to query for motivators from among a larger population than previously attempted. We did not try to define negative influences on decision to pursue obstetrics and gynecology; those are evident in the popular and specialty press. Although there has been considerable discussion about the negative influence of the liability crisis on recruitment, it is clear that those students who are interested in obstetrics and gynecology still pursue the specialty, despite adequate foreknowledge and forewarning of the implications of the liability crisis on future practice. Thus, the question is: what in obstetrics and gynecology still appeals to these students, notwithstanding the obvious detractors?
When considering which aspects of the field of obstetrics and gynecology are most desirable to entering residents, it remains the surgical and procedural aspects of the field. Furthermore, appeal of the variety of clinical experience in day-to-day practice, and the fast-paced/high-acuity experiences (obstetric, surgical) predominate across both genders. Interestingly, the fast-paced/high-acuity experiences were significantly more important to female respondents as compared with the male respondents, although both ranked it highly.

Respondents were interested in the gynecologic experience afforded by a given residency program, which speaks to the attraction of the procedural aspects of the field, as noted above. However, they also indicated the importance of the “atmosphere” of the residency, namely, resident camaraderie, faculty “approachability,” and the responsiveness of the program/faculty/program director to resident concerns. In addition, when in the process of interviewing, they highly valued the ability to meet with residents and nonadministrative faculty, presumably to get a better “feel” for what the residency program is really like. The significance of this weight placed on the less-tangible, less-quantitative aspects of residency may be its reflection of the changing values and perspectives among today’s future doctors when considering what it means to be a resident and/or physician in medicine (in general) and/or obstetrics and gynecology (in specific), as it pertains to views toward work, work ethic, fiduciary responsibility, and a career in medicine.

The quest for effective recruitment into our specialty has been discussed since at least the early 1970s. In 1991, Metheny et al surveyed senior students at 11 medical schools to determine predictors of specialty choice among the respondents who ranked obstetrics and gynecology as one of their top 4 specialty choices. They found that students attracted to the specialty were more likely female, and gave higher ratings of importance to obstetrics and gynecology procedures and risks, the opportunity to work with mostly healthy patients, patient contact, the opportunity to perform surgery, their beliefs about reproductive issues, and a perceived need for more physicians in the specialty. In our study, the survey recipients were actual applicants to obstetrics and gynecology residency programs, thus allowing for a larger sample size of those who have already been identified to have interest in the field. There were a higher number of respondents who were female, reflecting the national status of applicants to residency in obstetrics and gynecology. However, although our respondents also found procedures and surgical opportunities to be most important, patient health status and advocacy were of lesser importance, particularly for the male respondents.

In 2003, Fogarty et al investigated the same question, but asked it of the previous ten year’s graduates from one Midwest medical school, and found that factors significant in influencing choice of obstetrics and gynecology as a specialty were gender, second-year rotation, obstetrics and gynecology staff, continuity of patient care, primary care opportunities, surgical opportunities, female patients, healthy population, lifestyle opportunities, and financial opportunities. Although our study found surgical opportunities and faculty/staff to be important factors for choice of residency, primary care, lifestyle, and income were of little import to applicants to obstetrics and gynecology residency. That same year, Schnuth et al surveyed current medical students, years

<table>
<thead>
<tr>
<th>Table IV</th>
<th>Residency program recruitment techniques. Different programs use different recruitment methods. If you encountered any of these techniques, please rate whether they influenced your decision to rank a given program (scale of 1-5; 1 = influenced me to NOT RANK/rank lower a program, 5 = influenced me to RANK higher a program)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recruitment technique</td>
<td>Mean ± SD</td>
</tr>
<tr>
<td>Residency information on own website</td>
<td>3.6 ± 0.9</td>
</tr>
<tr>
<td>Communications before interview</td>
<td>3.4 ± 1.0</td>
</tr>
<tr>
<td>Opportunity to interview with residents</td>
<td>4.3 ± 0.9</td>
</tr>
<tr>
<td>Opportunity to interview with faculty members (other than director and chairperson)</td>
<td>4.1 ± 0.9</td>
</tr>
<tr>
<td>Preinterview social (night previous)</td>
<td>4.0 ± 1.1</td>
</tr>
<tr>
<td>Other opportunity for informal interactions with residents</td>
<td>3.8 ± 1.1</td>
</tr>
<tr>
<td>Postinterview communications</td>
<td>3.9 ± 1.1</td>
</tr>
<tr>
<td>Follow-up (second-look) visit</td>
<td>3.2 ± 1.6</td>
</tr>
<tr>
<td>Solo interview day (you were the only applicant interviewing that day)</td>
<td>1.9 ± 1.2</td>
</tr>
<tr>
<td>Group interview day (there were other applicants interviewing that day)</td>
<td>3.5 ± 1.0</td>
</tr>
<tr>
<td>One-on-one faculty interviews</td>
<td>4.1 ± 0.9</td>
</tr>
<tr>
<td>Multiple faculty members interviewing 1 applicant (at once)</td>
<td>2.2 ± 1.1</td>
</tr>
<tr>
<td>One faculty member interviewing multiple applicants (at once)</td>
<td>1.9 ± 1.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table V</th>
<th>Positive influences toward choosing obstetrics and gynecology. Which of the following positively influenced your decision to choose a career in obstetrics/gynecology? (Please rate each of the following factors individually on a scale of 1-5; 1 = low importance, 5 = high importance.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ± SD</td>
<td></td>
</tr>
<tr>
<td>Third-year obstetrics/gynecology clerkship</td>
<td>4.4 ± 0.9</td>
</tr>
<tr>
<td>Interaction with residents</td>
<td>3.7 ± 1.2</td>
</tr>
<tr>
<td>Interaction with faculty</td>
<td>3.9 ± 1.1</td>
</tr>
<tr>
<td>Delivering infants</td>
<td>4.2 ± 1.1</td>
</tr>
<tr>
<td>OR experience</td>
<td>4.5 ± 0.8</td>
</tr>
<tr>
<td>Outpatient clinic experience</td>
<td>3.7 ± 1.1</td>
</tr>
<tr>
<td>Inpatient experience</td>
<td>3.8 ± 0.9</td>
</tr>
<tr>
<td>Didactics</td>
<td>3.3 ± 1.1</td>
</tr>
<tr>
<td>Hands-on opportunities</td>
<td>4.5 ± 0.8</td>
</tr>
</tbody>
</table>
1 through 4, in another Midwest medical school to determine factors that influence selection of the specialty of obstetrics and gynecology, but with the additional aim of evaluating gender differences. They found that female students felt more strongly than male students that interest in women’s health, ability to improve the health of women, emphasis on patient education and prevention, and intellectual content of obstetrics and gynecology were positive influences. In contrast, male students felt more strongly than female students that opportunities to perform surgery, gender, and working under pressure were influences. The proclivities expressed by the female students were supported by our study. However, it was the female respondents in our study who placed higher value on fast-paced/high-acuity experience over the male respondents.

In a 2004 study, Vaidya et al evaluated the relationship between specialty choice and medical student temperament and character, using the Cloninger Temperament and Character Inventory (TCI; a 240-question, self-administered, true-false questionnaire). Cloninger et al identified 4 temperament dimensions—novelty seeking (NS), harm avoidance (HA), reward dependence (RD), and persistence (P). The survey was administered to third- and fourth-year medical students of a Midwest medical school. Of those who chose obstetrics and gynecology as their specialty, they were noted to score high in NS and RD. (NS is the tendency to explore, to be curious about novel stimuli, and to actively avoid punishment and frustrating non-reward. High novelty seekers are curious about new situations, and the riskier the situation, the greater the “buzz” they experience. RD is the tendency to develop behaviors that lead to positive reinforcement and to maintain rewarded behaviors. Persons with high RD want to please others and social institutions that were previously rewarding, and tend to conform to the reward system.) These findings support our study findings, in that the surgical and procedural aspects appealing to applicants to obstetrics and gynecology are within the category of “novel situations” with potential for participation in risky situations. Furthermore, the importance of the residency camaraderie and faculty support speaks to the desire for rewarding social interactions.

Clearly, there are practical and professional crises facing practicing obstetrician-gynecologists. These need continued attention toward resolution, both at an individual and at an organized level. Nevertheless, in striving to promote the specialty of obstetrics and gynecology to potential residents, it is key to capitalize on its strengths (procedural/surgical aspects, variety and novelty in day-to-day practice). This study demonstrates that these strengths are also what entering residents find most appealing about our specialty. Behavioral research suggests that there are personality correlates in those that are attracted to the field of obstetrics and gynecology. Furthermore, stratification by gender did not demonstrate any differences in the perceived importance of these aspects of the specialty. Given that more than 70% of our respondents did not choose to become an obstetrician-gynecologist until their third-year core obstetrics and gynecology clerkship or later, there are ample opportunities to demonstrate and showcase the appealing aspects of the field.

References
Candid candidate comments: The relationship between residency program selection factors and match list placements from ranked applicants

Marilyn J. Raymond, PhD, PT,a,* Robert J. Sokol, MD,a Louis A. Vontver, MD, MEd,b Kenneth A. Ginsburg, MDa

Obstetrics and Gynecology Department, Wayne State University School of Medicine,a Detroit, MI; Obstetrics and Gynecology Department, University of Washington,b Seattle, WA

Received for publication February 17, 2005; revised May 26, 2005; accepted July 18, 2005

Objective: The purpose of this study was to explore the relationship between residency program selection factors and placements we received on candidates’ match lists.

Study design: An annual (2001-2004) telephone survey of applicants (n = 140) ranked for acceptance to the department’s residency program addressed 16 program selection factors important in choosing our site. Candidates’ comments about these program selection factors were recorded. Logistic regression was conducted with program selection factors as independent variables and program receipt of candidates’ top quartile match list rankings as the dependent variable.

Results: The overall response rate was 68% (n = 95). Applicants’ positive ratings on 3 program selection factors were related to our receipt of top quartile match list rankings: 1) our capacity to meet personal career goals (odds ratio [OR] 4.59); 2) positive faculty-resident relationships (OR 1.97); and 3) suitable location (OR 1.96).

Conclusion: The single most important factor contributing to receipt of candidates’ top quartile match list placements was positive perceptions of our program’s abilities to meet their personal career goal needs.

Obstetrics and Gynecology National Resident Match Program (NRMP) results continue to show tight competition among programs to secure best applicants. These trends hit home when the Wayne State University Obstetrics and Gynecology Residency Program in Detroit, Michigan experienced a 73% vacancy rate (8/11) in the 2001 match! This shocked us because we had a strong tradition of full matches for decades before this incident. Though empty positions were filled quickly, program administrators placed a high priority upon aggressive identification and change of factors contributing to this outcome. Part of our plan involved exploring the relationship between factors considered important to selecting our residency site by candidates.

No financial support was received for this study.

* Reprint requests: Marilyn J. Raymond, PhD, Hutzel Women’s Hospital, Obstetrics and Gynecology Medical Education Office, 3990 John R, 7 Brush North Room 7606, Mail Stop 165, Detroit, MI 48201.
E-mail: mraymond@med.wayne.edu

0002-9378/$ - see front matter © 2005 Mosby, Inc. All rights reserved.
we ranked, and placements we received on match lists from these candidates.

An accepted, standardized tool to assess residency program selection factors is not available. We reviewed 16 previous studies spanning 2 decades on candidates’ perceptions of factors important to their program selection. They revealed different target audiences, measurement tools, analytical methods, and numbers/kinds of program selection factors. Examples of program selection factors commonly used included resident satisfaction, program quality/reputation (eg, subspecialty opportunities, diverse training, research), interpersonal relationships, lifestyle issues, work environment (eg, salary, benefits, schedules), candidates’ intuitions, and interviews.3-18 Candidates rate location most frequently as an important program selection factor.3,5,6,9,10-15 Debate continues on aspects of program quality/reputation, interpersonal relationships, resident satisfaction, the work environment, or lifestyle as the single most important program selection factor.3-18

Studies on program selection factors typically use perspectives from graduating medical students, resident reflection, or specialties other than ob/gyn.3-16 We found only 1 study addressing program selection factors from viewpoints of ranked candidates.9 Very little is known about ob/gyn candidates because previous investigations included our specialty as a small part of aggregate results in national studies.9,12 None of these studies linked program selection factors to how candidates ranked a particular site on their match lists. To this end, our purpose was to fill this gap by evaluating the relationship between program selection factors used to choose our ob/gyn residency and placements we received on candidates’ NRMP match lists from applicants we ranked.

Material and methods

After review of the literature, we developed a survey instrument on 16 factors considered important to selecting our residency program (Table 1). Items were adapted from the Council for Resident Education in Obstetrics and Gynecology recommendations19 to applicants on this topic. This document was chosen as the basis for our items because it 1) reflected common items found in previous studies, 2) was created and endorsed by ob/gyn educational scholars from a prestigious professional organization, and 3) contained accessible, current information.

To measure participant responses to the importance of program selection factors, a 5-point Likert scale asked candidates to judge the quality of an item examining its utility as a selection factor for choosing our program (anchors: 1 = “strongly disagree,” 2 = “disagree,” 3 = “neutral—neither agree or disagree,” 4 = “agree,” and 5 = “strongly agree”). Some survey items were worded positively, and others negatively. Negatively worded items necessitated subsequent recoding of candidates’ responses in reverse order to correspond to ratings given on positively worded ones. The higher rating an item received, the greater it was considered a positive program selection factor for our site.

We asked participants to tell us the number of programs to which they had applied. They were queried about where they placed us on their final match lists (anchors: 1 = “did not rank our program,” 2 = “first quartile representing the lowest 25th percent or bottom quartile match list placement,” 3 = “second quartile representing between 26 to 50th percent position in match list placement,” 4 = “third quartile representing 51 to 75th percent position in match list placement,” and 5 = “fourth quartile representing 76th -100% position in match list placement or top quartile”). This grading system was implemented because applicants applied to different numbers of programs. The candidates’ recall of the number of total programs they ranked and where we fell in this schema led to their placement of our site into a particular quartile. These quartiles could be compared without the mismatch of rankings that would have occurred if we had used a strict rank order rating scale.

We trained our Medical Education Office personnel to conduct annual telephone surveys from 2001 through 2004. These surveys targeted 140 applicants for first-year residency positions who we ranked in the NRMP selection process. Interviews took place after the annual match results were official in mid-March. Each session lasted 20 minutes on average. Exempt institutional review board status was granted because of categorization of this study as an educational survey procedure with subject confidentiality and no reasonable risks to subjects upon disclosure of responses.

We used SPSS 12.0 (Chicago, IL) to conduct statistical analyses. We performed nonparametric testing because of the ordinal measurements used, and the lack of normality for some variables coupled with the small sample size. Univariate analyses included descriptive statistics: mean, mode, and median. Cross-tabulations between match results and candidates’ final match list rankings were performed. Bivariate calculations utilized the Spearman rank-order test (rho) for correlation coefficients. Multivariate analysis with backward, stepwise (conditional) logistic regression was conducted.

Because of the small sample size (n = 95), the number of independent variables in the logistic regression model needed to be 9 or less. We decreased the number of program selection factors (n = 16) used as independent variables to meet this criterion by choosing only those variables that had correlations significant at $P < .15$ between the program selection factors and the dependent variable. The purpose of this logistic regression was to determine which program selection factors
worked together in relationship to our program receiving candidates’ top quartile rankings. The dependent variable measuring how candidates ranked our program was recoded into 2 points: 0 = “candidates did not rank us in the top quartile” from original scores ‘1 to 4,’ and 1 = “candidates ranked us in the top quartile” from original scores of ‘5’ only. The independent variables consisted of 9 program selection factors that correlated significantly with the dependent variable of how candidates ranked our program ($P < .15$).

The logistic regression analysis indicated that the odds-ratio (OR) for “program meets personal career goals needs” was more than twice as high as the OR for either of the other 2 significant factors. We therefore decided to explore this factor further. The purpose of this additional logistic regression was to determine which program selection factors worked together in relationship to determine the “meaning” of program meets candidates’ personal career goal needs. We performed a regression similar to the procedure described previously. The dependent variable measuring program meets candidates’ personal career goals needs was recoded into 2 points: 0 = “candidates did not feel we met their personal career goals” from original scores ‘1 to 4,’ and 1 = “candidates did feel that we met their personal career goals needs” from original scores of ‘5’ only. The independent variables consisted of 7 program selection factors that correlated significantly with the dependent variable of program meets personal career goal needs ($P < .15$). The 7 independent variables were resident impressions at interview, gynecologic surgical opportunities, faculty-resident relationships, resident-resident relationships, program leadership, specialty opportunities, and curriculum.

To evaluate reliability of our findings, we performed leave-one-out cross-validation through discriminant analysis to assess generalization error. This technique sorts each case into a group with respect to classification functions calculated over the entire data set except the particular case that is being classified. The purpose of this methodology is to decrease overly positive bias that occurs when you classify the same cases that are used in the discriminant function coefficients. When similar correct classification occurs between the original cases and the cross-validated cases, the data can be generalized. In addition to reliability, face validity was established by evaluation of survey items from 5 ob/gyn medical education faculty members.

We recorded candidates’ comments about negative and positive residency program selection factors for qualitative analyses. These remarks added breadth and depth to the quantitative findings.

<table>
<thead>
<tr>
<th>Table I</th>
<th>Sixteen program selection factor survey items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor</td>
<td>Survey item</td>
</tr>
<tr>
<td>Subspecialty opportunities</td>
<td>The subspecialty education opportunities in the department (eg, reproductive endocrinology, maternal-fetal medicine, reproductive genetics, gynecologic-oncology) appeal to me.</td>
</tr>
<tr>
<td>OB surgical opportunities*</td>
<td>There appears to be inadequate surgical training in obstetric procedures.</td>
</tr>
<tr>
<td>Faculty impressions at interview</td>
<td>Faculty left me with a positive impression about the program.</td>
</tr>
<tr>
<td>Program meets personal career goals*</td>
<td>The program does not match my future career goals with respect to the kind of practice I hope to conduct.</td>
</tr>
<tr>
<td>Curriculum</td>
<td>The program appears committed to education, as evidenced by the implementation of a structured 4-year curriculum.</td>
</tr>
<tr>
<td>Lectures and conferences</td>
<td>The program appears committed to education, as evidenced by the number of formal teaching conferences.</td>
</tr>
<tr>
<td>Salary and benefits</td>
<td>I feel that this program’s salary and benefits packages are competitive with the program I selected.</td>
</tr>
<tr>
<td>Living location</td>
<td>The Southeast Michigan location is a desirable living place for my personal and family needs.</td>
</tr>
<tr>
<td>Program leadership</td>
<td>The program displays stable leadership, accreditation, and staffing patterns.</td>
</tr>
<tr>
<td>Faculty-resident relationships*</td>
<td>The quality of faculty and resident relationships appears poor.</td>
</tr>
<tr>
<td>GYN surgical opportunities*</td>
<td>There appears to be inadequate surgical training in gynecologic procedures.</td>
</tr>
<tr>
<td>Resident impressions at interview</td>
<td>Residents left me with a positive impression about the program.</td>
</tr>
<tr>
<td>Interview structure/content*</td>
<td>The interview process needs improvement.</td>
</tr>
<tr>
<td>Work schedule</td>
<td>This program’s work schedule is reasonable in comparison to the program I selected.</td>
</tr>
<tr>
<td>Medical Center location</td>
<td>The downtown Detroit Medical Center location is a desirable employment place.</td>
</tr>
<tr>
<td>Resident-resident relationships</td>
<td>The quality of the upper level/lower level resident relationships appears high.</td>
</tr>
</tbody>
</table>

* These items were worded negatively. Negatively worded items necessitated subsequent recoding of candidates’ responses in reverse order to correspond to ratings given on positively worded items. Thus, the higher rating an item received, the greater it was considered a positive program selection factor for our site.
Results

The overall response rate was 68% (n = 95). Applicants had interviewed at an average of 8.3 programs with a range from 1 to 20 programs. We received 33 (35%) top quartile rankings from these candidates.

Table II provides descriptive statistics on the scores of the 16 program selection factors. Because of skewed distributions on several factors, we used median values to report central tendencies. These scores ranged from ratings of neutral ‘3’ to highly positive ‘5.’

Table III displays correlation findings significant at P < .15 between the program selection factors and how candidates ranked our program significant at P < .15.

Table IV describes logistic regression with the factor as the dependent variable and with 7 independent variables that correlated with the dependent variable at P < .15. We found that this regression model’s predicted correct percentage improved from 52.2% to 73.9% using 3 program selection factors: 1) program leadership OR 2.23, CI 1.31-3.82, P = .003; 2) faculty-resident relationships OR 2.14, CI 1.06-4.32, P = .033; and 3) subspecialty opportunities OR 2.14, CI 1.02-2.31, P = .042. The strength of association for this regression model using Nagelkerke $R^2$ equaled 0.370.

Candidates’ comments regarding the 3 program selection factors related to receiving a top quartile ranking provided a better understanding to us of their negative and positive perceptions. Negative concerns about meeting personal career goals included those who wanted a community program after reviewing what we offered. Other candidates felt faculty-resident relationships were malignant (eg, not friendly, disengaged, unsupportive). One candidate summed up our living location by stating, “Detroit is old, gray, cold and the streets are full of potholes!” Others were concerned with safety issues. Having no family or friends in the area was considered problematic. In contrast, examples of positive comments included many candidates stating that the program met their career goal needs by the diverse and complex patient population, the outstanding faculty, and opportunities in subspecialty training, research, and...
the pursuit of fellowships. Some applicants felt faculty-resident relationships were good after conversing with them during the interviews, discussing faculty relationships with current or past residents, or from their personal clinical experiences affiliating with our program. Our living location was considered positively by individuals who grew up in the area, or had relatives and friends nearby. Some desired the area because of employment opportunities for spouses, excellent public schools, and competitive living costs compared with many other large cities.

Our reliability and validity results supported the utility of the measures selected. Leave-one-out cross-validation analysis indicated that findings were likely to be generalizable. The technique showed similar results comparing original grouped cases with cross-validated grouped cases as follows. Correct classification of the original group was 72.8% in comparison to 70.7% for the cross-validated group. Five faculty reviewers unanimously agreed that the survey items represented residency program selection factors to confirm face validity.

Comment

Of the 16 program selection factors that we assessed, 1 stood out as most important in relationship to receipt of candidates’ top quartile rankings. That program selection factor was how well the program met our candidates’ personal career goal needs. To explore this concept further we looked at our candidates’ remarks. These remarks revealed that a university setting, a program that cared for a diverse and complex patient population, offered research and subspecialty training, and gave them skills to help pursue fellowships often explained how their career goals were met by the program. Another logistic regression revealed that “outstanding faculty” members through strong program leadership and supportive faculty-resident relationships were key characteristics that helped define this factor. In addition, this logistic regression indicated that availability of subspecialty training was another critical component to meeting our candidates’ personal career goal needs. This analysis also revealed to us that faculty-resident relationships were even more important than the initial findings showed. Faculty-resident relationships actually were intertwined with meeting candidate’s personal career goal needs, as well as standing alone in relationship to our receipt of top quartile rankings from candidates.

The high importance candidates placed on our program meeting their personal career goal needs indicated that candidates considered their professional aspirations above personal ones when selecting and ranking residency programs. Nevertheless, the learning environment and lifestyle issues also played into decisions to find the right site. A remarkably strong association with receiving a top quartile ranking (ie, Nagelkerke $R^2 = 0.529$) was obtained when a high rating on the factor “program meets candidates’ personal career goal needs” combined with good “faculty-resident relationships” and a suitable “living location.” In order to receive top quartile rankings from candidates, they needed to feel that our site provided major educational resources in an enhanced learning environment, and in a location that supports their personal lifestyles.

When we looked at central tendency scores for each program selection factor (Table II), the results showed us that we could work to improve candidates’ ratings on all program selection factors. Two factors on which we focused most were highlighting the benefits of living in Southeastern Michigan, and working internally to improve faculty-resident relationships.

Because in the real world we could not wait 4 years to respond to our recruitment needs, we made changes according to preliminary findings from annual surveys. One limitation of this study is that these actions may have colored the interpretation of the overall results. However, our work in using these surveys has met with success indicating the study’s reliability and validity. After 3 years of post-match activities to handle vacant positions, we filled all NRMP openings in 2004 and 2005. Our interview numbers of candidates who meet or exceed our admittance standards have increased from 45 in past years to 70 in 2005.

Other limitations of our study included small numbers of participants. This necessitated the modification of the logistic regression procedures previously described. In addition, we did not have the demographic information available to assess and compare nonresponders with responders in order to determine whether response bias occurred. Finally, changes in the applicant pool, the resident pool, and the curriculum were not incorporated into the model.

An important difference in our study from others was that no other study used our survey item entitled “program meets candidates’ personal career goal needs.” Our qualitative and quantitative findings helped us to interpret the meaning of this survey item to reflect other investigations’ descriptions of program quality/reputation (eg, outstanding faculty, opportunities in subspecialty training, diverse and complex patient population, research, and pursuit of fellowships).

Thus, we join those authors who have found program quality/reputation as a key aspect of resident recruitment. We add that a program’s quality/reputation includes meeting candidates’ professional career goal needs through maintaining outstanding faculty and providing subspecialty opportunities.

While several investigators have found interpersonal relationships were more important than program quality/reputation as the number 1 ranked program
selection factor, our study found interpersonal relationships to be significant, but of lesser importance than program quality/reputation. Despite controversy over ranked importance, the identification of interpersonal relationships as a fundamental program selection factor has been established by other studies. Further, our results confirm the findings of many others that location remains an important program selection factor. The identification of similar important program selection factors despite differences in study designs supports the validity of these results.

The information gained from these surveys is similar to market research. It has proven a rich and valuable resource in identifying program selection factors our applicants consider when ranking us on their match lists. Other programs can adapt this process to assess their recruitment needs or use our findings. These surveys have provided us with timely feedback from the source that matters to us most—the NRMP candidates we wanted to join our program. We believe their input helped us focus on many of the issues that led to our previous poor match results. The most valuable lesson learned is that if we meet our candidates’ needs, we meet our program’s needs for recruiting best candidates.

Acknowledgments

Special thanks go to several people who helped with this investigation. Alaina Bruce conducted the surveys from 2002 to 2004. Michael Kruger assisted with the data analysis and interpretations. Gloria Kuhn, MD, provided numerous writing tips and editing suggestions. John M. Malone, Jr, MD, and Theodore B. Jones, MD, supplied support to keep me on target and energized me to complete this project. The APGO Solvay Scholars program provided structure and helpful information. Finally, appreciation is due to each of the 95 candidates who took the time to participate in this study and share their thoughts.

References

Educational games in an obstetrics and gynecology core curriculum

Sharon O’Leary, MD,a,b,* Lisa Diepenhorst, MD,a Ruth Churley-Strom, MS, RN,a Diane Magrane, MDc

Department of Obstetrics and Gynecology,a St Joseph Mercy Hospital, Ypsilanti, MI; The University of Michigan School of Medicineb; Faculty Development and Leadership Programs,c Association of American Medical Colleges, Ann Arbor, MI

Received for publication February 18, 2005; revised July 18, 2005; accepted July 18, 2005

KEY WORDS
Game theory in medical education
Interactive learning of ectopic pregnancy

Objective: The purpose of this study was to compare learning satisfaction and effectiveness using traditional lecture or educational game in teaching medical students about ectopic pregnancy.

Study design: Third-year medical students were randomized to instruction about ectopic pregnancy through either standard lecture or educational Jeopardy style game. Students in each group completed a pretest, posttest, and satisfaction survey. Experts in ectopic pregnancy validated the pretest and posttest. The satisfaction survey was taken from published validated tests. Paired samples t test was used to compare pretest and posttest scores. Independent samples t tests were used to compare test scores and satisfaction responses between groups. Chi-square tested dichotomous satisfaction responses.

Results: All 104 students in both groups showed significant improvement in learning about ectopic pregnancy (P < .001) on pre- and posttest comparison, with scores being almost identical. Students in the group randomized to game format rated it higher in stimulating faculty/student interaction, helping retain information, and overall enjoyment than students participating in the lecture method (P < .001). In addition, students in the game group responded positively that the format was interactive, stimulated their interest, and kept them engaged in class content (P < .001).

Conclusion: The innovative educational game format was as effective as standard lecture in educating students about ectopic pregnancy, while being more enjoyable and stimulating. Based on these conclusions, we hope to motivate other teachers in obstetrics and gynecology to use innovative teaching methods to provide a more enjoyable, stimulating, and active means of effective medical education.

Creative teaching techniques engage students, stimulate participation, and sustain their interest. Educational games meet these criteria and can serve as an additional interactive model for enhancing the academic environment. Games are effective, stimulating, and encourage active participation. A variety of educational games have been used in medical education, with approaches including a card game to teach immunology, a “Survivor” game to review pulmonary physiology, and a “Scrupules” game to teach managed care ethics.
specific game format or the topic is not as important as
the novel approach to teaching that encourages active
participation, and offers a respite from the world of
“hum-drum” didactic lectures, which is the usual do-
main of the medical students.

While educational games have been found to be
effective and enjoyable in medical education, none have
been studied while teaching obstetrics and gynecology
topics to medical students. Currently, the standard
method of instruction of third-year medical students is
the didactic lecture. The goal of this study was to evaluate
student learning and student satisfaction with an inter-
active game compared with a standard lecture format in
teaching medical students about ectopic pregnancy.

Material and methods

This study was deemed exempt from Institutional Re-
view Board review because it involved research con-
ducted in established or commonly accepted educational
settings dealing with normal educational practices, such
as research on instructional strategies and evaluation of
the effectiveness of the strategies.

The University of Michigan academic scheduling
system assigns 32 to 37 students to an 8-week clerkship
in obstetrics and gynecology. The first 5 days are
devoted to orientation activities and classes. On day 1,
the investigators used a random numbers table to assign
students to 1 of 2 methods for instruction on ectopic
pregnancy. The principal investigator taught both ses-
sions, the standard lecture with open questions and
answer period, and the game format. The class sessions
immediately followed each other. Thus, for the opening
week of each ob-gyn clerkship, 16 to 18 students
attended a lecture/discussion on ectopic pregnancy and
16 to 18 students participated in a game organized to
instruct about ectopic pregnancy. The principal investi-
gator prepared both lecture and game using Association
of Professors of Gynecology and Obstetrics (APGO)
learning objectives, with careful cross-referencing of
teaching points to ensure consistency in content pre-
sented. All students were encouraged to ask questions at
the end of each session.

Students completed a 20-item multiple-choice pretest
to assess their general knowledge about ectopic preg-
nancy. The test questions were developed by the prin-
cipal investigator and included 2 questions extracted
from the APGO test bank. Face and content validity
were judged by teaching faculty from the University of
Michigan and St Joseph Mercy Health System based on
the 5 APGO teaching objectives for ectopic pregnancy.

After students completed the pretest, those assigned
to the lecture format received a standard 45-minute
lecture with supporting slides by the principal investi-
gator. Students assigned to the game format were
divided into 4 teams. The principal investigator pro-
vided the game rules, moderated the “Jeopardy” style
game, and elaborated on correct answers. The game
categories of 1) epidemiology and differential diagnosis,
2) risk factors, 3) signs/symptoms, 4) diagnosis, and 5)
treatment matched the APGO teaching objectives.
Cards with increasing dollar values were displayed on
the game board below these categories. Teams were
awarded points based on the degree of difficulty and
complexity of the question they selected. The team with
the most points received a high-energy edible reward.

Immediately following each class, students completed
the posttest and a satisfaction survey consisting of 4
questions using a 5-point Likert scale and 3 true-false
questions. Students were additionally queried about their
interest in obstetrics and gynecology as a specialty and
the number of hours spent preparing for the class. The
satisfaction survey was based on validated satisfaction
surveys,5-7 and judged to have face and content validity
by an educational researcher and a faculty mentor
involved in the APGO/Solvay Scholars program.

A power analysis was done to determine a sample size
required to achieve a minimum power of 0.80. Calcula-
tions using nQuery Advisor Release 4.0: Study Planning
Software 1995-2000 (Statistical Solutions, Ltd, Cork,
Ireland) identified that a sample size of 20 in each group
would have 92% power to detect a difference in means
of 3 assuming that the common standard deviation is
2.700 using a 2 group t test with a 0.050 2-sided
significance level.8

Pre- and posttest scores, as well as survey responses
rating increase in general knowledge, stimulation of
faculty/student interaction, retention of information,
and enjoyment of the format were compared between
lecture and game methods by independent samples t test.
Paired samples t test was used to compare student
pretest and posttest percent correct scores. A chi-square
statistic was used to compare true or false responses
with questions about the interactive nature of the
presentation, the appropriateness of the method used
for teaching medical facts about ectopic pregnancy, and
stimulation of student interest and engagement in the
lecture and game groups.

Results

During the study period, a total of 104 students were
assigned to the ob/gyn clerkship. All students who
arrived for the class agreed to take part in the study.
Fifty-two students participated in the lecture and 52
students in the game classes.

Mean pretest scores were 11.33 (56.6%) for students
in the lecture group and 12.52 (62.6%) for students in
the game format group. Although there was a significant
difference in pretest scores between groups (P = .005),
students in both groups demonstrated improvement in
learning with similar mean posttest scores of 16.98
To what extent has the educational format increased your general knowledge regarding the subject matter of Ectopic pregnancy?

To what extent did the educational format stimulate faculty/student interaction?

To what extent did the educational format help you retain Ectopic pregnancy topic information?

Overall, I enjoyed the educational Game format for this class?

Figure 1 Satisfaction survey questions comparing game versus lecture using a 5-point Likert scale. The dark bars refer to the Lecture group, and the light bars refer to the Game group.

Figure 2 Percentage of students answering yes to the following questions. The dark bars refer to the Lecture group and the light bars refer to the Game group.

Comment

No subject has educational value until it is adjusted to the learner. Students who are actively participating in problem-based curricula report more meaningful learning and higher motivation and interest compared with students in more traditional learning environments, where students relied on rote learning and completion of the course was the goal. Key concepts of adult learning include active participation by learners and application of knowledge, yet didactic lectures remain the standard method of teaching medical students. Games serve as a model for innovation in medical education wherein students are not passive listeners, but active participants in their own learning.

Our study demonstrated that there was no significant difference in learning on posttest scores between traditional didactic lecture and use of an interactive game. Most notable, however, is that students taught by game were significantly more likely to find the class increased faculty-student interaction, and was more enjoyable,
stimulating, and interactive. Based on our findings we hope to motivate other teachers of obstetrics and gynecology to develop innovative strategies that provide enjoyable effective teaching methods.

This study focused on preexisting knowledge, learning, and short-term retention through use of pre- and posttests. Future studies to test long-term recall of content taught using an educational game format would add evidence to the value of educational games.

References

Medical student identification of domestic violence as measured on an objective, standardized clinical examination

Susan E. Hoffstetter, PhD, WHNP, Robert J. Blaskiewicz, MD, Gail E. Furman, PhD, Jennifer A. McCabe, BS

Department of Obstetrics, Gynecology, and Women’s Health, Saint Louis University School of Medicine, St Louis, MO

Received for publication February 18, 2005; revised May 26, 2005; accepted July 18, 2005

**KEY WORDS**
Standardized patients
Domestic violence
Medical student education

**Objective:** This study examined the incidence and predictors of domestic violence screening by third-year medical students at an end of clerkship Objective Standardized Clinical Examination.

**Study design:** Two hundred and seventy-five third-year medical students completed an 8-station end of clerkship Objective Standardized Clinical Examination as part of this retrospective observational study, one with nonspecific abdominal pain and possible domestic violence. Checklists on history, physical, communication, and interpersonal skills were collected. Domestic violence screening was analyzed by logistic regression and analysis of variance.

**Results:** The incidence of domestic violence screening by history alone was 34% before the physical. Interpersonal scores on the overall exam and domestic violence station, but not gender or rotation sequence, predicted domestic violence questioning.

**Conclusion:** In this standardized patient study there was a low rate of domestic violence screening by history.

© 2005 Mosby, Inc. All rights reserved.

Domestic violence is a public health problem, a major cause of morbidity and mortality in the United States, and a serious social problem impeding women’s participation and contributions to society. Domestic violence is used to establish power and control over another person through actual or threatened physical/sexual violence or psychologic/emotional abuse and its effects cross ethnicity, race, age, national origin, sexual orientation, and religious and socioeconomic lines.

Approximately 1 to 4 million women a year in the United States are abused by an intimate partner. These women suffer immediate and long-term physical and mental health consequences, accruing an estimated cost of 4.1 billion dollars for medical and mental health care. Women experiencing violence have 50% higher medical care costs compared with women not currently experiencing violence.

The need for violence education was first emphasized in the 1985 Surgeon General’s Workshop on Violence and Public Health. In 1995, the Association of American Medical Colleges challenged medical schools to create the domain of a family violence curriculum comparable to traditional subjects in whom students would gain the
ability to screen, assess, and support victims of violence, as well as develop community support networks, resources, and information on health care professional involvement.6

Medical schools have adopted varying formats from lectures and workshops to training sessions to address the topic. There is little information on the incidence of screening for domestic violence by medical students in the clinical setting. Objective Standardized Clinical Examination (OSCE) evaluations are now frequently used to measure clinical skills, and successful completion of a standardized patient examination is required for licensure. The purpose of this study was to determine the incidence and predictors of domestic violence screening by third-year medical students at an end of clerkship OSCE.

Material and methods

The study was a retrospective, observational study approved by the Institutional Review Board of Saint Louis University conducted for the academic years of 2002 and 2003. All third-year medical students (n = 276) completed an 8-station OSCE at the end of the obstetric and gynecology clerkship. One student was excluded from the analysis because of insufficient data, leaving 275 students in the study. The formal curriculum at Saint Louis University during the time of this study included a preclinical presentation on domestic violence and a didactic session during the ob/gyn clerkship.

One of the OSCE stations presented a patient complaining of nonspecific abdominal pain with an associated abdominal bruise. Domestic violence was a consideration in the differential diagnosis of this patient. The standardized patient was trained to offer information about her abusive relationship if the student asked specifically about domestic violence. A simulated bruise was placed on the patient’s abdomen that was clearly visible once the abdomen was properly exposed during the physical examination. If the student asked about the bruise the standardized patient was instructed to convey the information about her domestic violence. In the absence of a specific question about domestic violence or a question concerning the bruise the student did not receive that information.

Data were collected on a subset of 236 consecutive students by timing of domestic violence inquiry (before the physical examination, during/after the exam, or not at all). The gender of the student was recorded. The time of the students obstetric and gynecology rotation in the academic year was noted. Rotations were grouped into quarters of the academic year for analysis.

Data collected on all students for the domestic violence station included a patient-completed history checklist about the character of her abdominal pain and associated symptoms, in addition to domestic violence questions. The student was also assessed on their communication response to the patient’s domestic violence situation, including the student’s ability to offer resources and reassurance, and identify patient safety. The students’ interpersonal skills were assessed by the use of a previously validated tool, Patient Perception Questionnaire (PPQ).

Data were collected from the 8 stations across the domains of history taking, physical examination skills, communication, diagnosis, treatment, and PPQ. Overall OSCE scores were calculated by both including and excluding the domestic violence station. The incidence of domestic violence screening was calculated by including the number of students asking about domestic violence at any time during the patient encounter. Statistical analysis was completed using multivariate analysis of variance (MANOVA), univariate analysis (ANOVA), Bonferroni-adjusted pairwise mean comparisons, and simultaneous-entry, multiple linear regression.

Results

Of the 275 students, 243 (88.4%) inquired about domestic violence with at least 1 question at some point in the standardized patient encounter. In the subset of 236 students for whom the timing of this inquiry was known, 80 students (33.9%) asked about domestic violence before the physical exam, 133 students (56.4%) asked about domestic violence during or after the physical exam, and 23 students (9.7%) never asked. The percentage of students in each of these 3 categories did not vary significantly by student gender, $\chi^2(2) = 1.9, P = .38$. Similarly, rotation by academic quarter was not significantly associated with the timing of the domestic violence questions, $\chi^2(6) = 8.6, P = .20$.

Multivariate analysis of variance (MANOVA) was used to determine whether timing of domestic violence questions (before exam vs during/after exam vs never) was associated with other indicators of clinical performance. The MANOVA included the OSCE total scores (omitting the domestic violence station) for history taking, physical examination, communication, PPQ, and diagnosis/treatment. The PPQ score for the domestic violence station was also included as a separate variable. The multivariate effect was significant across the 6 dependent variables, Wilk’s $\lambda = 0.83, P < .001$. Follow-up univariate ANOVAs indicated that of the 6 dependent variables, only the PPQ score for the domestic violence station was significantly different as a function of domestic violence question timing, $F(2,233) = 18.7, P < .001$. Descriptive data are shown in Table I. Bonferroni-adjusted pairwise mean comparisons indicated that students who never asked about domestic violence had significantly lower PPQ scores on the domestic violence station than students who asked about domestic violence before or during/after the exam ($P < .05$).
Simultaneous-entry, multiple linear regression was used to predict, from the other indicators of clinical performance, the number of domestic violence questions asked by the student (out of 3 questions). The mean number of domestic violence questions asked was 2.4 (SD 1.0; range 0-3). Neither student gender nor rotation was significantly correlated with number of domestic violence questions asked: gender, \( r = 0.03, P = .53 \); rotation, \( r = -0.03, P = .54 \). These variables were not included in the regression analysis. The number of domestic violence questions asked was predicted from the OSCE total scores for history taking, physical examination, communication, PPQ, and diagnosis/treatment (again, omitting the domestic violence station). The PPQ score for the domestic violence station was again included as a separate predictor. The overall regression equation with six predictors was significant, \( R = 0.22, F(5,269) = 3.1, P < .05 \). Results are displayed in Table III.

Overall performance on the domestic violence station (total score combining history taking, physical exam, communication, and PPQ) was also examined. Neither student gender nor rotation was significantly correlated with overall performance on the domestic violence station: gender, \( r = 0.11, P = .16 \); rotation, \( r = 0.003, P = .96 \). Using multiple linear regression, domestic violence station performance was predicted from the OSCE total scores for history taking, physical examination, communication, PPQ, and diagnosis/treatment (again, omitting the domestic violence station). The overall regression equation with 5 predictors was significant, \( R = 0.22, F(5,269) = 3.1, P < .05 \). Among the 5 predictors, only the PPQ score (across all other stations besides the domestic violence station) significantly predicted overall domestic violence station performance, \( \beta \) (standardized regression coefficient) = 0.18, \( P < .01 \). Results are displayed in Table III.

### Table I  Clinical performance as a function of domestic violence question timing

<table>
<thead>
<tr>
<th>DV question timing</th>
<th>Before exam</th>
<th>During/after exam</th>
<th>Never</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>History taking*</td>
<td>63.3 (10.0)</td>
<td>63.1 (7.9)</td>
<td>64.6 (8.8)</td>
</tr>
<tr>
<td>Physical exam*</td>
<td>68.3 (7.6)</td>
<td>67.2 (7.4)</td>
<td>68.1 (6.7)</td>
</tr>
<tr>
<td>Communication*</td>
<td>67.2 (13.0)</td>
<td>69.1 (14.0)</td>
<td>68.8 (17.6)</td>
</tr>
<tr>
<td>PPQ*</td>
<td>75.9 (8.1)</td>
<td>76.1 (8.2)</td>
<td>74.7 (5.7)</td>
</tr>
<tr>
<td>Diagnosis/treatment*</td>
<td>86.0 (14.2)</td>
<td>83.3 (15.2)</td>
<td>80.9 (14.1)</td>
</tr>
<tr>
<td>PPQ DV station(^{1})</td>
<td>81.4 (13.3)</td>
<td>76.8 (14.9)</td>
<td>60.5 (16.1)</td>
</tr>
</tbody>
</table>

* OSCE scores; values do not include data from the domestic violence station.

\(^{1}\) Bonferroni-adjusted pairwise comparisons: never < before exam, during/after exam (\( P < .05 \)).

### Table II  Prediction of number of domestic violence questions asked from clinical performance indicators (predictor variables)

<table>
<thead>
<tr>
<th>Predictor variables</th>
<th>( \beta )</th>
<th>( t )</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>History taking*</td>
<td>-0.06</td>
<td>-0.90</td>
<td>.37</td>
</tr>
<tr>
<td>Physical exam*</td>
<td>0.06</td>
<td>0.91</td>
<td>.36</td>
</tr>
<tr>
<td>Communication*</td>
<td>0.06</td>
<td>0.96</td>
<td>.34</td>
</tr>
<tr>
<td>PPQ*</td>
<td>0.003</td>
<td>0.04</td>
<td>.96</td>
</tr>
<tr>
<td>Diagnosis/treatment*</td>
<td>0.03</td>
<td>0.55</td>
<td>.58</td>
</tr>
<tr>
<td>PPQ DV station</td>
<td>0.39</td>
<td>6.6</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

* OSCE scores; values do not include data from the domestic violence station.

### Table III  Prediction of overall performance on domestic violence station from clinical performance indicators (predictor variables)

<table>
<thead>
<tr>
<th>Predictor variables</th>
<th>( \beta )</th>
<th>( t )</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>History taking*</td>
<td>-0.11</td>
<td>-1.5</td>
<td>.13</td>
</tr>
<tr>
<td>Physical exam*</td>
<td>0.11</td>
<td>1.5</td>
<td>.13</td>
</tr>
<tr>
<td>Communication*</td>
<td>0.09</td>
<td>1.4</td>
<td>.15</td>
</tr>
<tr>
<td>PPQ</td>
<td>0.18</td>
<td>2.6</td>
<td>.009</td>
</tr>
<tr>
<td>Diagnosis/treatment*</td>
<td>-0.002</td>
<td>-0.03</td>
<td>.97</td>
</tr>
</tbody>
</table>

* OSCE scores; values do not include data from the domestic violence station.

**Comment**

This study found a low incidence of domestic violence questioning before the physical exam in a standardized patient simulating an episode of nonspecific abdominal pain. Medical students were better at domestic violence screening after identification of an abnormal physical finding. The timing of the domestic violence questioning was not influenced by gender or time in the academic year. Students who were rated higher on the PPQ score on the domestic violence station were more likely to inquire about domestic violence. This would suggest that students who have better interpersonal skills are better at screening and identifying patients at risk for an abusive situation. Alternatively, the standardized patient may have perceived a more empathetic student if they inquired about domestic violence.

The depth of domestic violence inquiry was not influenced by gender or amount of previous clinical experience. Possible explanations could relate to personality traits, violence exposure, personal experiences,
and value attributed to the social history by role models. In our study the overall OSCE PPQ score was predictive of the student’s performance on the domestic violence station. This performance was unrelated to other aspects of history taking and physical examination and suggests a correlation between student’s ability to relate to the patient and their score. There may be a factor inherent in the student’s physician patient relationship that is a predictor of domestic violence screening.

Standardized patients are widely used in medical education. Our study used this tool to measure the student’s behavior regarding a significant social problem in America. The OSCE exam can be used to measure the effect of curricular change addressing domestic violence.

The manifestations of partner violence touch all medical specialties, reinforcing the need for all physicians to be trained in screening, intervening, and advocating support for violence victims. Medical students need a consistent approach in developing the clinical skills to enable screening for domestic violence. Students need to incorporate domestic violence screening in routine assessments of all patients.

Acknowledgments

We thank John T. Chibnall, PhD, Saint Louis University School of Medicine, Associate Professor of Psychiatry, for his assistance with statistical analysis.

References

A new curriculum for hysteroscopy training as demonstrated by an objective structured assessment of technical skills (OSATS)

Amy L. VanBlaricom, MD,* Barbara A. Goff, MD, Michael Chinn, MD, Melodie M. Icasiano, MD, Peter Nielsen, MD, Lynn Mandel, PhD

Departments of Obstetrics and Gynecology, University of Washington, Madigan Army Medical Center, Seattle, WA

Received for publication February 18, 2005; revised May 16, 2005; accepted July 18, 2005

Objective: The purpose of this study was to develop an effective curriculum to teach hysteroscopy, as well as to develop an objective assessment instrument (OSATS).

Study design: All 24 residents in our training program attended a 3-hour hysteroscopy curriculum in each of 3 years. Each year after the curriculum, an OSATS was administered consisting of an assembly and operative task. Each task was evaluated with a task-specific checklist and a previously validated global assessment form. Twenty-four residents from an outside institution served as controls. All residents were examined by blinded and unblinded examiners.

Results: The reliability coefficient was .7857 for the checklist and .9763 for the global scale. Interrater reliability for the checklist was .7478 and ranged from .4222 to .9329 for the global instruments. Evaluation of construct validity revealed that the junior residents at both locations did significantly worse on the checklist and global scale than the senior residents for all 3 years of the study (P < .001). Those residents provided the curriculum did significantly better on both the checklist and the global scale than the resident controls with a mean of 45% versus 63% for the checklist and 18.3 versus 24.9 for the global score. (P = .001 checklist, P = .007 global scale).

Conclusion: This curriculum is an effective way to impart knowledge and skill in the assembly and use of the operative hysteroscope. The checklist developed for this OSATS has excellent reliability and construct validity.

Surgical training is an important part of any residency program. Historically, surgical training occurred in the operating room where residents first observe an attending perform procedures, and then take on increasing roles in the surgical cases under direct supervision.

This type of training results in variable acquisition of surgical skills, often caused by inherent differences in attending technique, as well as variation in how much each resident is allowed to do in a given procedure. These differences are sometimes based on perceived “difficulties” in the case, including patient factors such as age, weight, or associated medical conditions, which carry with them a need to complete the procedure in a timely manner. Other differences in resident experience...
are based on attending personal preference as well as their own comfort level with the procedure they are performing.

In the last 10 years, advances in minimally invasive surgical techniques have been astounding. Endoscopic surgeries are rapidly becoming more complex and technically difficult and in some cases replace historically tried and true procedures, which require laparotomy. In many cases, faculty are finding themselves with inadequate experience with new technologies and, therefore, residents are presented with fewer opportunities to perform these new techniques on their own even with direct supervision. Additionally, with the advent of work hour restrictions there is now an increasing concern about decreased surgical experience in resident training.1-4 Because a significant depth and breadth of such experience is required for surgical privileges and is even being considered for standardization in National Board Exams, it is incumbent upon us as academic surgeons to expose residents to the experience they need to demonstrate adequacy once they leave our program. The responsibility for training competent surgeons lies heavily upon us in academics. There are excellent data to support the effectiveness of formal, objective teaching of surgical skills outside of the operating room.5-15 Studies in the general surgery literature indicate that this may be a more efficient way of teaching surgical skills.12,16-18

Despite its recognized importance in the age of minimally invasive surgery, little work has been done investigating teaching of hysteroscopic skills.19-24 For the most part, investigations in this area have been in development of realistic models and simulators, with minimal work in curriculum development and appropriate assessment of skills.25,26 Our goal was to develop an effective method to teach and evaluate knowledge and technical skill in hysteroscopy in obstetrics and gynecology residents.

Material and methods

A 3-hour curriculum was developed, consisting of a 1-hour didactic session covering surgical indications,
contraindications, anticipated complications, and basics of equipment and media options, followed by a 2-hour hands-on lab where residents were given an opportunity to assemble the diagnostic and operative hysteroscopes and review troubleshooting points in their use, as well as the use of other endoscopic equipment such as camera and light source (Figures 1 and 2). Each resident was then given an opportunity to practice assembly and resection with the operative hysteroscope on an appropriate inanimate model with one-on-one supervision from a faculty expert. The didactics were given each year in 3 sessions of 8 residents per session.

All 24 residents in our training program were exposed to the hysteroscopy curriculum in 2001, 2002, and 2004. Each year, approximately 4 to 6 months after the curriculum, a posttest was administered consisting of an OSATS (Objective Structured Assessment of Technical Skills) where residents would demonstrate assembly of an operative hysteroscope and hysteroscopic resection of pathology from a realistic inanimate model in front of an examiner (Figures 3 and 4). Specifically, the residents were asked to resect a large polyp from an inanimate uterine model. Reusable models were purchased from Simulations (formerly Limbs and Things, Bristol, England) and cost approximately $20 per resident. Twenty-four residents from an outside institution who were not provided with the curriculum were tested similarly. These residents served as the untrained cohort.

In any residency program there is concern about bias in evaluation as the examiners have often worked closely with the residents they are evaluating. Thus, an important component to the project was testing residents at
Madigan Army Medical Center, where examiners from the UW were blinded to level of training. In addition, we had faculty members from Madigan evaluate our residents in a completely blinded fashion. Faculty at each institution evaluated their own residents as well. Blinded and unblinded assessments were then compared. Previous evaluation of blinded examiners demonstrated a high interrater reliability.\textsuperscript{27,28} In doing so, we determined the true objectivity of the OSATS.

For this project, a 2-station OSATS was designed to assess diagnostic and surgical hysteroscopy. A task-specific checklist was developed for each station inclusive of steps of the procedure deemed necessary and appropriate to safe and effective completion of the task. Each of the tasks in the OSATS was timed and evaluated with a pass/fail judgment, the task-specific checklist evaluation, and a previously validated global assessment form\textsuperscript{16-18} (Figures 5 and 6). Residents could fail a task based on time (longer than 10 minutes to complete the task) or technique. The time limit was chosen at 10-minutes per task as it was used in previous endoscopic OSATS. Previous research in surgical education has demonstrated that reliability and validity of an assessment tool are important.\textsuperscript{16-18}

By adding outside blinded examiners for the OSATS, we attempted to demonstrate a high interrater reliability. In addition, by looking at all residents in the program we attempted to show a difference across years.
That is, a PGY4 (postgraduate year) would be more knowledgeable and technically adept than a PGY1 (construct validity).

The reliability of each of the instruments was assessed by Chronbach’s alpha. Interrater reliability was calculated using intraclass correlation coefficients.

Construct validity for each instrument was evaluated by comparing scores among the residency levels. This was accomplished with a one-way analysis of variance with resident year as the independent variable. This study was approved for human subjects’ exemption given that it involved assessment of educational tools.

Results

At our institution, residents begin their exposure to hysteroscopy in their R1 year. The amount of the procedure they are allowed to perform is determined by the individual attending based on the level of difficulty and some personal comfort. As part of the preliminary research for this project, a survey was performed of all 24 residents in our department. This revealed that residents felt that in most cases they were lacking in hysteroscopy experience, including discomfort with operative indications and contraindications, assembly of the complex operative hysteroscopy equipment, operative technique, and management of complications. Residents reported a range of between 21 and 45 hysteroscopic procedures by the completion of the chief year. In many cases they were only the secondary surgeon, with the attending performing much of the technically difficult portion of the case. Initial testing of the residents at all levels of training proved that hysteroscopy skills were inadequate, with only 1 of 6 second-year residents and 3 of 6 chief residents able to assemble a resectoscope in an appropriate time frame. Only 1 chief resident was able to correctly name the type and strength of media used at our institution for operative hysteroscopy, and none could describe why this particular media was chosen. The results of this survey clearly underscore the need for additional training in hysteroscopy.

The OSATS instruments developed to assess hysteroscopic skills were evaluated for reliability and validity. The reliability coefficient was .7857 for the checklist and .9763 for the global scale. Interrater reliability for the task specific checklist was .7478 and ranged from .4222 to .9329 for the global instruments.

Evaluation of construct validity revealed that the senior residents (PGY3 and 4) at both locations did significantly better on the checklist and global scale than
1. Assembly
   - Able to assemble correctly and efficiently: No, Yes (3)
   - Overall Technique:
     - Poor: 1, 2, 3, 4, 5 Excellent
   - Time__________________
   - Comments_________________________________________

2. Resection Task
   - Places obturator first: No, Yes (3)
   - Opens inflow and outflow during resection: No, Yes (3)
   - Pulls loop toward scope rather than away: No, Yes (3)
   - Uses cut instead of coagulation: No, Yes (3)
   - Removes bx specimen w/o removing obturator: No, Yes (3)
   - Removes scope after turning off inflow: No, Yes (3)

3. Overall Technique
   - Poor: 1, 2, 3, 4, 5 Excellent

4. Resection time ________________________________
   - Comments________________________________________

Figure 5  Hysteroscopy Testing.
the junior residents (PGY1 and 2) for all 3 years of the study. The mean scores for the checklist were 43% (PGY1), 53% (PGY2), 68% (PGY3), and 70% (PGY4) ($P < .001$).

### Table I

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Respect for tissue</strong></td>
<td>Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments.</td>
<td>Careful handling of tissue, but occasionally caused inadvertent damage.</td>
<td>Consistently handled tissues appropriately with minimal damage.</td>
<td></td>
</tr>
<tr>
<td><strong>Time and motion</strong></td>
<td>Many unnecessary moves.</td>
<td>Efficient time/motion, but some unnecessary moves.</td>
<td>Economy of movement and maximum efficiency.</td>
<td></td>
</tr>
<tr>
<td><strong>Instrument handling</strong></td>
<td>Repeatedly makes tentative or awkward moves with instruments.</td>
<td>Competent use of instruments although occasionally appeared stiff or awkward.</td>
<td>Fluid moves with instruments and no awkwardness.</td>
<td></td>
</tr>
<tr>
<td><strong>Knowledge of instruments</strong></td>
<td>Frequently asked for the wrong instrument or used an inappropriate instrument.</td>
<td>Knew the names of most instruments and used appropriate instrument for the task.</td>
<td>Obviously familiar with the instruments required and their names.</td>
<td></td>
</tr>
<tr>
<td><strong>Use of assistants</strong></td>
<td>Consistently placed assistants poorly or failed to use assistants.</td>
<td>Good use of assistants most of the time.</td>
<td>Strategically used assistant to the best advantage at all times.</td>
<td></td>
</tr>
<tr>
<td><strong>Flow of operation and forward planning</strong></td>
<td>Frequently stopped operating or needed to discuss next move.</td>
<td>Demonstrated ability for forward planning with steady progression of operative procedure.</td>
<td>Obviously planned course of operation with effortless flow from one move to the other.</td>
<td></td>
</tr>
<tr>
<td><strong>Knowledge of specific procedure</strong></td>
<td>Deficient knowledge. Needed specific instruction at most operative steps.</td>
<td>Knew all important aspects of the operation.</td>
<td>Demonstrated familiarity with all aspects of the operation.</td>
<td></td>
</tr>
</tbody>
</table>

Overall, on this task, should this candidate: Pass__________ Fail__________

Comments:


**Figure 6** Hysteroscopy.
The senior residents (PGY3 and PGY4) did significantly better on both the task-specific checklist and global scale than the junior residents (PGY1 and PGY2) ($P < .001$) (Table II).

The next step was to evaluate the effectiveness of the curriculum. Those residents provided the curriculum did significantly better on both the task-specific checklist and the global scale than those residents not provided the curriculum, with a mean score of 45% versus 63% for the checklist and 18.28 versus 24.88 for the global score ($P = .001$ checklist, $P = .007$ global scale) (Table III). This difference was also found when evaluating residents by PGY year (Table IV). There were a total of 5 trained residents who participated in all 3 years of the study. When their individual scores were reviewed, there was a significant improvement in global score for each resident for each year they received the curriculum ($P = .012$). The improvement in the task-specific checklist was also a pronounced trend that approached significance ($P = .059$) (both in Table V).

The curriculum cost involved price of the models and dedicated faculty time. The reusable hysteroscopy pelvic trainer was purchased from Limbs and Things for $407. The replaceable advanced resectable uteri were purchased for $61 each and were used for training and OSATS of 3 residents per uterus. The hysteroscopic instruments, camera, and video towers were all borrowed from the operating room of the hospital where the program was conducted.

The largest expense of the curriculum was dedicated faculty time. One dedicated faculty member provided the 3-hour curriculum on 3 occasions (for 8 residents each occasion) for each of the 3 years of the study (a total of 9 hours per year). At each OSATS session there were 2 faculty evaluators (1 blinded and 1 unblinded). Time for OSATS was 30 minutes per resident for a total of 96 faculty hours (32 hours per year).

Residents who received the curriculum rated it very highly. They were asked to rate the course in terms of

<table>
<thead>
<tr>
<th>Table I</th>
<th>Construct validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGY year</td>
<td>n</td>
</tr>
<tr>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>4</td>
<td>25</td>
</tr>
</tbody>
</table>

Baseline results on checklist and global scale by PGY year for all residents in study.

<table>
<thead>
<tr>
<th>Table II</th>
<th>Construct validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>Checklist</td>
</tr>
<tr>
<td>Junior residents (PGY1 and 2)</td>
<td>36</td>
</tr>
<tr>
<td>Senior residents (PGY3 and 4)</td>
<td>36</td>
</tr>
</tbody>
</table>

Residents provided the curriculum by PGY year.

<table>
<thead>
<tr>
<th>Table III</th>
<th>Curriculum effectiveness trained versus untrained residents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global scale</td>
<td>Checklist</td>
</tr>
<tr>
<td>Trained residents</td>
<td>72</td>
</tr>
<tr>
<td>Untrained residents</td>
<td>18</td>
</tr>
</tbody>
</table>

$P = .007$ $P = .001$

<table>
<thead>
<tr>
<th>Table IV</th>
<th>Trained versus untrained results broken down by postgraduate year</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGY1</td>
<td>Checklist</td>
</tr>
<tr>
<td>Trained</td>
<td>51%</td>
</tr>
<tr>
<td>Untrained</td>
<td>22%</td>
</tr>
</tbody>
</table>

$P = .046$ $P = .002$

| PGY2 | Checklist | Global |
| Trained | 53% | 22.38 |
| Untrained | 51% | 20.00 |

$P = .842$ $P = .620$

| Junior PGY (1 and 2) | Checklist | Global |
| Trained | 52% | 20.83 |
| Untrained | 36.5% | 13.89 |

$P = .030$ $P = .023$

| PGY3 | Checklist | Global |
| Trained | 73% | 27.23 |
| Untrained | 52% | 19.00 |

$P = .018$ $P = .198$

| PGY4 | Checklist | Global |
| Trained | 75% | 30.42 |
| Untrained | 55% | 24.50 |

$P = .053$ $P = .096$

| Senior PGY (3 and 4) | Checklist | Global |
| Trained | 74% | 28.91 |
| Untrained | 54% | 22.67 |

$P = .002$ $P = .055$

<table>
<thead>
<tr>
<th>Table V</th>
<th>Individual improvement in residents participating in all 3 years of the curriculum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resident</td>
<td>Year of curriculum</td>
</tr>
<tr>
<td>1</td>
<td>2001</td>
</tr>
<tr>
<td>1</td>
<td>2002</td>
</tr>
<tr>
<td>1</td>
<td>2004</td>
</tr>
<tr>
<td>2</td>
<td>2001</td>
</tr>
<tr>
<td>2</td>
<td>2002</td>
</tr>
<tr>
<td>2</td>
<td>2004</td>
</tr>
<tr>
<td>3</td>
<td>2001</td>
</tr>
<tr>
<td>3</td>
<td>2002</td>
</tr>
<tr>
<td>3</td>
<td>2004</td>
</tr>
<tr>
<td>4</td>
<td>2001</td>
</tr>
<tr>
<td>4</td>
<td>2002</td>
</tr>
<tr>
<td>4</td>
<td>2004</td>
</tr>
<tr>
<td>5</td>
<td>2001</td>
</tr>
<tr>
<td>5</td>
<td>2002</td>
</tr>
<tr>
<td>5</td>
<td>2004</td>
</tr>
</tbody>
</table>

$P = .012$ $P = .059$
applicable to their education, effectiveness of imparting content, and perceived results (that is, were they able to integrate knowledge gained into their clinical practice in a positive way). On a 5-point scale the results of this rating were all 4.5 or higher. Notable comments included a significant improvement in comfort with the equipment and troubleshooting.

Comment

There is mounting evidence in the surgical literature to support the use of skills labs outside of the operating room for resident training.\textsuperscript{11-15} Cundiff found that a curriculum consisting of didactics as well as laparoscopic skills training in a trainer and pig lab setting resulted in significant improvement in ob-gyn residents’ performance in laparoscopic surgery.\textsuperscript{13} Similarly, Coleman and Muller\textsuperscript{11} found a significant improvement in technical skill and operative performance in a group of upper level ob-gyn residents after a 4-week curriculum of intensive laparoscopic skills training. At the University of Washington, Goff et al have demonstrated similar effectiveness with a comprehensive surgical skills curriculum for ob-gyn residents.\textsuperscript{9,10,27,29}

Procedures involving endoscopy lend themselves particularly well to standardization via models. For example, the intrauterine cavity can be simulated on a computer or in an inanimate model for resection with striking similarity to human tissue. By using models for objective surgical skills assessment, several advantages are gained. First, residents can be allowed to operate completely independently. There is an opportunity to make surgical decisions and demonstrate judgment without encountering the obvious ethical dilemmas the same would arouse on a human patient. Another advantage is the ability to repeat the same procedure as many times as necessary with continuous feedback until adequacy of skill and comfort with the equipment are reached. There are economic advantages to this method, as well as the models are portable, it does not prolong operating room time, and the same equipment can be used repeatedly. Thus, an individual resident can actually perform as the primary surgeon many more of a given procedure with this method. Competence is more quickly obtained while expending fewer resources.

Historically, evaluation of residents’ surgical skills has been subjective evaluation by an individual attending sporadically throughout the rotation or upon completion of it. Often this assessment does not occur at the time of a given surgery and therefore may lack effectiveness and objectivity caused by poor recollection of individual events by the attending. Because of variations in attending personal style and individual patient factors, subjective resident evaluation has been shown to be inconsistent and difficult to standardize, as well as lacking reliability and validity.\textsuperscript{5} There are extensive data in both the general surgery and gynecology literature supporting objective evaluation of surgical skills as both reliable and valid.\textsuperscript{5,10,14,16-18} Much of the research in establishing OSATS for ob-gyn residents has been accomplished at our institution.\textsuperscript{9,27,29} In addition, the reliability and validity of this instrument has been shown to be quite high, even when the providers of the exam were blinded to resident year of training and other personal details.\textsuperscript{27} Given this high reliability and validity, OSATS is applicable to other standardized exam scenarios, including national board exams and other high stakes exams.

While the instruments used in this study had high reliability and validity, interrater reliability on the global scale was significantly lower than seen previously. The global scale used in this study has been previously validated extensively by Reznick,\textsuperscript{5} and UW studies.\textsuperscript{9,14,27,29} Examiners from the institution whose residents did not receive the curriculum were not consistently the same examiners throughout the study, and they did not receive as much training in the use of the form as those at the University of Washington. Therefore, the low interrater reliability on the global instrument in this study could be caused by inadequate training in the use of the global form by new examiners. The task specific checklist developed for evaluation of OSATS in this study showed high interrater reliability between blinded and unblinded examiners. Construct validity of the instrument used for the hysteroscopy OSATS was also good. We were able to show a significant increase in baseline knowledge and technical skill by advancing years in both trained and control residents.

The curriculum developed for this study was shown to be effective as, overall, those residents provided the curriculum did significantly better on both the global scale evaluation and the task-specific checklist at the time of OSATS. When the individual PGY years were broken down, there continued to be noteworthy trends in improved acquisition of hysteroscopic knowledge and skill per postgraduate year, although this only became statistically significant when the junior residents and senior residents were compared as groups between trained residents and controls. Additional years of evaluation will increase the number of residents and, thus, the power of this study, which should improve our ability to show statistical significance on an individual basis.

All residents were surveyed upon completion of each year of the study. There was an overwhelmingly positive response to the curriculum across the years of the study.

Residents specifically noted an improved comfort with hysteroscopic instrumentation as well as practical use. They added that this comfort and familiarity translated to increased independence in the operating room which increased with PGY year.
The issue of feasibility is also important to consider in order to use the curriculum and OSATS at other institutions. The curriculum and OSATS outlined above are portable and, therefore, feasible to conduct not only in our institution but at others as well.

Instruments used for the curriculum are already available at any institution providing hysteroscopic procedures, and the models used in the OSATS are inexpensive, reusable, and replaceable. Thus, this curriculum and evaluation method is an excellent option for training in hysteroscopy, an important skill for any practicing obstetrician gynecologist.

References

Measurement of endometrial stripe thickness by obstetrics and gynecology residents

Daniel M. Breitkopf, MD,a,* Edward R. Smith, PhD,a William N. P. Herbert, MDb

Departments of Obstetrics and Gynecology, University of Texas Medical Branch, Galveston, TX; University of Virginia, Charlottesville, VA

Received for publication February 21, 2005; revised August 8, 2005; accepted August 11, 2005

Objective: How well do obstetrics and gynecology residents measure the endometrial stripe by transvaginal sonography?

Study design: Three obstetrics and gynecology residents at the University of Texas Medical Branch from each year level were tested for their ability to perform endometrial stripe measurements. Measurements of endometrial stripe thickness within 1 mm of the reference value that was obtained by an expert sonographer were deemed to be correct.

Results: Residents correctly measured the endometrial stripe in 14 of 24 cases (58%). Postgraduate year–4 residents correctly measured endometrial stripe thickness in 5 of 6 cases (83%); lower level residents were correct in 9 of 18 cases (50%). The most common error that was noted was an incorrect image plane in 25 of 47 attempts (53%). Fewer errors were committed by upper level as compared with lower level residents (P < .05).

Conclusion: By the end of residency, residents can accurately measure the endometrial stripe using transvaginal sonography. Emphasis should be placed on the improvement of the measurement skills.

Competence in pelvic sonography is vital for practicing obstetrician/gynecologists. Many practicing obstetrician/gynecologists perform pelvic sonography in their offices for a variety of indications. Residency training programs in obstetrics and gynecology are required to document resident experience in pelvic sonography for accreditation by the Accreditation Council for Graduate Medical Education. There is evidence that the number of sonography examinations that are performed does not correlate necessarily with competence. Hertzberg et al1 evaluated 10 radiology residents after they had performed 200 sonography cases. Anatomic landmarks were found correctly in only 56% of cases, although clinically significant errors were made in 1 of every 2 cases. The experience of the resident did not correlate with performance. On the basis of these findings, it is apparent that competency in sonography must be measured specifically, instead of assumed, on the basis of experience alone.

The measurement of endometrial stripe thickness is used to determine whether an abnormality that requires further evaluation is needed and to direct therapy for abnormal uterine bleeding. Inaccuracy in measurement

* Reprint requests: Daniel Breitkopf, MD, 301 University Blvd, Route 0587, Galveston, TX 77555-0587.
E-mail: dmbreitk@utmb.edu

0002-9378/$ - see front matter © 2005 Mosby, Inc. All rights reserved.
doi:10.1016/j.ajog.2005.08.030
may lead to the performance of unnecessary procedures or misdiagnosis. To date, no studies have examined specifically the performance of residents at measuring endometrial stripe thickness. This study seeks to fill this gap in knowledge by demonstrating either the need for further education or that current residency training is adequate to learn this important skill. The objective of this investigation was to measure the performance of obstetrics and gynecology residents in the sonographic measurement of endometrial stripe thickness.

Materials and methods

Using an institutional review board–approved protocol, 3 residents from each year level in the obstetrics and gynecology program at the University of Texas Medical Branch (postgraduate year [PGY] 1-4) were selected randomly. Residents were approached, and consent was obtained by the co-investigator who is not involved in their evaluation or promotion decisions (E.R.S.). Data from this study were not used in the evaluation or promotion decisions regarding the resident physicians who were involved.

The obstetrics and gynecology residency program consists of 8 residents per year level who obtain the vast majority of their clinical experience at a large tertiary care academic medical center. Formal training in transvaginal sonography consists of didactic lectures given 1 to 2 times each year and direct supervision in a specialized pelvic sonography clinic. The clinic experience is once each week for 2 6-week rotations in the first, third, and fourth years of training. Residents perform pelvic sonography and sonohysterography with an experienced faculty member in an American Institute of Ultrasound in Medicine–accredited sonography laboratory. The mean number of vaginal sonograms that were performed by the 2004 class of residency graduates from the program was 136.

Each resident performed a transvaginal sonogram on 2 normal volunteers, specifically obtaining endometrial stripe measurements. For the first sonogram, 1 subject was imaged by all residents. For the second sonogram, residents scanned 1 of 2 different subjects. For each subject, the measurement of endometrial stripe thickness was obtained on 2 separate images, and the values were averaged for analysis. The measurements that were obtained were compared with the value that was obtained by an expert sonographer. Measurements of endometrial stripe thickness within 1 mm of the reference value that was obtained by the expert sonographer were deemed to be correct. Several published reports confirm the interobserver variability of endometrial stripe measurements to be within 1 mm.2-4

Subjects who were scanned were recruited from the standardized patient program at the University of Texas Medical Branch. These patients participate in the Gynecologic Exam Teaching Associate program to instruct second-year medical students in the performance of pelvic examinations. After giving informed consent, subjects had a transvaginal sonogram by 4 different examiners in 1 day (3 resident examinations, 1 faculty examination each day).

The time to perform the sonogram and still and video images of the endometrial stripe measurement that were coded with subject and examiner numbers were recorded. Subjects rated their discomfort with the examination by using a visual analogue scale. Residents’ perception of their own competence in measuring the endometrial stripe was measured by a visual analogue scale. The linear visual analogue scales utilized for both the discomfort and competence measurements were 100 mm in length.

Still images that were obtained by the residents were reviewed by the expert sonographer to determine errors that were committed in the measurement of the endometrial stripe by the American Institute of Ultrasound in Medicine Standards.5 Errors were classified in the following manner: incorrect image plane, uterus not imaged, cervical canal echo measured, and incorrect caliper placement. Image plane was deemed correct if the endometrial echo could be seen in longitudinal/sagittal plane in its entirety as defined.

Data were analyzed with SigmaStat software (version 3.0; SPSS, Inc, Chicago, IL). Mean differences between resident and expert measurements of endometrial stripe thickness, time to perform the examinations, and pain scale ratings were compared with the use of the t-test or Mann-Whitney Rank Sum test, as appropriate. The proportion of correct measurements was compared with the use of the chi-squared test. For the chi-squared test, cells were combined when the proportion of expected cell frequencies < 5 was > 25%. One-way analysis of variance was used to compare data across year levels. The Pearson product moment correlation was calculated to examine the relationship between the difference of endometrial stripe measurement from the reference value and number of vaginal sonograms that were performed. A probability value of < .05 was considered significant.

Results

All residents completed the assigned endometrial stripe measurements in August and September of 2004. One PGY 1 resident did not take 2 endometrial stripe measurements on 1 of the subjects. Three perimenopausal women from the standardized patient program were scanned during the study. PGY 1 residents did not have clinical or didactic experience in vaginal sonography at the time of the study. The mean number of sonograms that were performed during residency for the PGY 2, 3, and 4 residents studied was 58.
In 14 of 24 attempts, residents measured the endometrial stripe thickness within 1 mm of the expertly obtained value. Table I compares the performance of the residents by PGY level. Differences between performance by year level were not statistically significant for mean time to complete the examination, proportion within 1 mm of reference, mean visual analogue pain score, or mean difference in endometrial stripe from standard by analysis of variance. There was not a significant correlation between the number of vaginal sonograms that were performed and the difference of endometrial stripe measurement from the reference value ($r = -0.18; P = .58$). The residents as a group took longer to complete the examination than did the expert reference examiner ($P < .001$, Mann Whitney rank sum test). The mean pain scores were not significantly different between residents and the expert.

Errors in imaging and measurement were classified by the examination of the still images by the reference sonographer (D.M.B.) and are presented in Table II. Upper level residents were significantly less likely to make errors in imaging and measurement than lower level residents ($P < .05$). Errors in obtaining the correct imaging plane were the most common and persisted throughout the year levels. Errors in visualization of the uterus/endometrium, measurement of the cervix, or incorrect caliper placement became less common with increasing resident experience.

Residents’ perception of their competence, as measured by the visual analogue scale, in performing pelvic sonography increased from 23 mm for the first-year residents to 69 mm for the fourth-year residents, but this change did not reach statistical significance.

**Comment**

Significant difficulties were encountered by residents in the measurements of the endometrial stripe thickness. Most fourth-year residents were able to measure the endometrial stripe accurately and with fewer technique errors than junior residents. Time to perform the examination was longer for the residents; however, the subjects did not note more discomfort.

The present study examined a sample of residents from 1 large university program. Although residency curriculum and clinical rotations are set mainly by national standards, interprogram differences in sonography education are likely to exist. Thus, the applicability to other residency programs may be limited. The study used a limited number of normal subjects, which should reduce differences in the level of examination difficulty between residents. However, because only subjects with normal endometrial stripes were included, the potential for error in measurement was limited. We did not determine the ability of residents to distinguish

---

**Table I** Performance data for residents by year level

<table>
<thead>
<tr>
<th>Group</th>
<th>Measurements within 1 mm of reference value (n)</th>
<th>Endometrial stripe measured (n)</th>
<th>Mean endometrial stripe difference from reference (mm)*</th>
<th>Mean time to complete examination (sec)*</th>
<th>Mean pain score* (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGY 1</td>
<td>6 (50%)</td>
<td>11 (64%)</td>
<td>0.08 ± 0.07</td>
<td>206 ± 79</td>
<td>3.3 ± 1.6</td>
</tr>
<tr>
<td>PGY 2</td>
<td>6 (67%)</td>
<td>12 (67%)</td>
<td>0.19 ± 0.21</td>
<td>232 ± 82</td>
<td>4.0 ± 2.0</td>
</tr>
<tr>
<td>PGY 3</td>
<td>6 (40%)</td>
<td>12 (83%)</td>
<td>0.15 ± 0.06</td>
<td>198 ± 110</td>
<td>3.5 ± 1.0</td>
</tr>
<tr>
<td>PGY 4</td>
<td>6 (83%)</td>
<td>12 (100%)</td>
<td>0.07 ± 0.06</td>
<td>175 ± 32</td>
<td>3.1 ± 2.1</td>
</tr>
<tr>
<td>All residents</td>
<td>24 (58%)</td>
<td>48 (77%)</td>
<td>0.12 ± 0.12</td>
<td>204 ± 79†</td>
<td>3.5 ± 1.7</td>
</tr>
<tr>
<td>Reference/expert</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>85 ± 53</td>
<td>3.3 ± 0.8</td>
</tr>
</tbody>
</table>

N/A, Not applicable.

* Data are given as mean ± SD.
† Time to complete examination significantly longer than reference; $P < .001$, Mann Whitney rank sum test.

**Table II** Classification of errors made in the measurement of endometrial stripe thickness*

<table>
<thead>
<tr>
<th>PGY level</th>
<th>Total measurement attempts (n)</th>
<th>No errors made (n)†</th>
<th>Error in image plane (n)</th>
<th>Uterus and/or endometrial stripe not imaged (n)</th>
<th>Calipers placed too wide (n)</th>
<th>Cervical canal echo measured (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

* The row totals are not equal to measurement attempts as multiple errors could be made in a single measurement attempt.
† $P < .05$ for lower level residents (PGYs 1 and 2) versus upper level residents (PGYs 3 and 4) by chi-squared test.
between normal and abnormal endometrial stripes. The inclusion of a larger number of residents would improve the statistical power and might lead to the ascertainment of significant differences that were not seen with the present sample.

This is the first published investigation to examine the performance of obstetrics and gynecology residents in the measurement of endometrial stripe thickness. As mentioned earlier, Hertzberg et al\(^1\) tested radiology residents in their sonographic abilities; however, skills that are specific to pelvic sonography were not analyzed separately. Few other reports of trainee performance in sonography exist and are limited mostly to abdominal sonography in trauma by emergency medicine and general surgery.\(^6,7\) Parker et al\(^8\) examined the performance of obstetrics and gynecology residents and nurse practitioners in the performance of sonohysterography with surgical findings as a reference. No significant differences were seen in true-positive rates between examinations that were performed by the nurse practitioners, second- or third-year residents, or fellows. The results were limited by the retrospective study design, by the fact that the examinations were not performed independently by the provider, and by the fact that the examinations with negative/normal findings were not included.

Although specific certification is not required for physicians to perform sonographic examinations, several professional organizations offer voluntary accreditation of sonography practices. The accreditation programs of the American College of Radiology and the American Institute of Ultrasound in Medicine require physicians to have a specified number of sonographic examinations but do not require competency testing. Our data indicate that, at least for graduating residents, significant difficulties exist in the measurement of the endometrial stripe and that competency specific testing may be needed to assure accuracy.

An analysis of the errors that were committed by the residents indicates a need for specific education in obtaining the correct image plane and identifying the uterus accurately. A checklist of specific requirements to obtain an accurate endometrial stripe may be helpful in aiding the proper measurement technique. It is encouraging that, despite a significant number of errors, the endometrial stripe was measured within 1 mm of the reference value in most cases. Obtaining the “right answer” by using the wrong method, however, may increase the likelihood of error in future attempts.

Our data, although specific to 1 residency training program, indicate the need for the specific assessment of sonographic skills in the measurement of the endometrial stripe. Residency programs should consider testing their trainees to determine the effectiveness of their curriculum with regards to this vital skill for the practicing obstetrician/gynecologist.

**Acknowledgments**

We thank Julie Griffice, RN, Anne Llana, RDMS, and Billie Rosenberger, RDMS, for their assistance in coordinating the data collection.

**References**

Impact of 1996 Residency Review Committee obstetrics-gynecology primary care requirements on residency training and surgical procedures

Margaret W. Chu, MD, a,b Martha J. Rall, MS, a,c Linda M. Frazier, MD, MPH, a,c,* Douglas V. Horbelt, MD, a,b,c Travis W. Stembridge, MD a,b

Department of Obstetrics and Gynecology, University of Kansas School of Medicine–Wichita a; Wesley Medical Center b; Women’s Research Institute, c Wichita, KS

Received for publication February 22, 2005; revised August 10, 2005; accepted August 18, 2005

KEY WORDS
Primary care
Obstetrics-gynecology
Residency training
Attitudes

Objective: This study evaluated the impact of required primary care rotations in obstetrics-gynecology residency training after 1996.

Study design: A questionnaire was sent to the 1994 to 2003 graduates from 1 residency program, and records of surgical procedures completed during residency were analyzed.

Results: Thirty-nine of 46 graduates participated in the study (response rate 85%). Required primary care training was associated with increased confidence in providing primary care (81.5% versus 54.5%, P = .12) but less agreement that obstetrics-gynecology is a primary care specialty (21.4% versus 45.5%, P = .23). Abdominal hysterectomies and vaginal hysterectomies per resident did not decrease (127 versus 113, P = 0.149, and 55 versus 48, P = .06, respectively). Adjusted for temporal trends, cesarean sections per resident decreased (366 versus 321, P = .009).

Conclusion: Residents maintained adequate rates of major inpatient surgical procedures after implementation of required primary care training. There was a tendency for residents who graduated after 1996 to have less favorable attitudes about primary care.

© 2005 Mosby, Inc. All rights reserved.

For many women of reproductive age, obstetrician-gynecologists serve as their principal source of primary and preventive health care. Therefore, to educate residents training in obstetrics and gynecology primary health care conditions, the Residency Review Committee for Obstetrics and Gynecology (OB/GYN) implemented new requirements for OB/GYN residency training programs. The requirements, effective January 1, 1996, mandate a minimum of 6 months of primary care training, including rotations in emergency medicine, family practice, and/or internal medicine and geriatrics. The length of residency training remained 4 years.

In a 1995 baseline survey, 87% of OB/GYN residents considered OB/GYN to be primary care and 85% planned to practice primary care after residency.1 Another survey in 1998 showed that 93% of practicing obstetricians and gynecologists routinely saw primary care patients in their practice.2 A review of the literature showed no published studies assessing how well graduates
believe that the new primary care curriculum prepares them for the role of primary care provider. No published studies were located evaluating whether the primary care training requirements may have an adverse impact on the number of surgical procedures done during residency.

The goals of this study were to: (1) evaluate beliefs among OB/GYN residency graduates about serving as a primary health care provider, (2) assess whether obstetrician-gynecologists who participated in primary care rotations during residency feel adequately prepared to provide primary care, and (3) determine whether the primary care training requirement may affect surgical training experience.

**Material and methods**

This study involved resident physicians who graduated between 1994 and 2003 from the OB/GYN residency program at the University of Kansas School of Medicine–Wichita.

Prior to 1996, residents in this program could choose from a list of electives, which included primary care training such as internal medicine, and non–primary care training such as pathology, anesthesiology, sonography, and dermatology. After 1996, residents rotated through 6 required months of family practice and/or internal medicine, geriatrics, and emergency medicine.

The study was approved by the Institutional Review Board of the Wichita Medical Research and Education Foundation and the University of Kansas School of Medicine, Wichita, KS. Each physician who completed all 4 years of training at this program was contacted by mail with a consent form explaining the study. Consent-forming former residents completed a questionnaire designed to assess how they manage primary care conditions among their patients, attitudes about primary care, and the residency training they received. Primary care was defined as “general health care (ie, diseases other than those of the reproductive tract),”.

Data from the department’s inpatient surgical procedure database were used to compute the total number of selected procedures performed by each consenting graduate over the 4 years of residency training. Surgical procedures were captured in this database by an interface with the hospital’s computerized medical record system and were validated monthly by each resident during training. The 3 most common surgeries performed during OB/GYN residency, cesarean section, abdominal hysterectomy, and vaginal hysterectomy, were evaluated. Eligible surgical procedures were cases in which the resident had “complete” involvement in the procedure as defined by the Obstetrics Gynecology Residency Review Board.

**Statistical analysis**

Data were analyzed using the Statistical Package for the Social Sciences (version 11.5, SPSS Inc., Chicago, IL) and SAS (version 8.01, SAS Institute, Cary, NC). Residents who graduated in 1994 to 1996 were compared with residents who graduated in 1997 to 2003. Because the incidence of cesarean sections changed over the years during the time frame of the study, we adjusted the number of cesarean sections performed by each resident to the total number of cesarean sections performed at the program’s teaching hospital as follows. The ratio of cesarean sections done by the resident to the total number of cesarean sections at the hospital was calculated for each half-year and then multiplied by the average number of cesarean sections at the hospital during 1994 to 2003. Means were compared using independent-samples t tests, and proportions were compared using 2 tests or Fisher exact tests. If data were skewed, medians are presented and the appropriate nonparametric tests were used to assess statistical significance. All tests were 2 tailed using an alpha of 0.05.

**Results**

Thirteen residents graduated in 1994 to 1996 (earlier graduates), and 33 graduated in 1997 to 2003 (later graduates). Eleven earlier graduates and 28 later graduates responded (response rates 84.6% versus 84.5%, respectively, P = 1.0). Earlier graduates were older than later graduates (median 40 versus 34.5 years, respectively, P < .001) and were less likely to have participated in family practice or internal medicine rotations during residency (36.4% versus 96.4, respectively, P < .001).

When comparing earlier to later graduates, there were no significant differences in gender (male, 54.5% versus 46.4%, respectively, P = .73), practice location (small town 54.5% versus 32.1%, respectively; urban/military, 18.2% versus 35.7%, respectively; suburban, 27.3% versus 32.1%, respectively, P = .39), or portion of practice devoted to obstetrics as compared with gynecology (median 50%, range 30% to 65% versus median 60%, range 15% to 75%, respectively, P = .24).

**Provision of primary care services**

All respondents reported being currently in general OB/GYN practice. Most respondents said they usually treat patients younger than 18 years of age (71.1%), although 7.9% treat these individuals only if pregnant, and 10.5% refer them to other physicians. Women in geriatric age groups were treated by 61.1% of the respondents, whereas 27.8% said they refer these women. When graduates were asked to categorize how they typically care for various health problems (Table), urinary tract infections, psychosocial problems, and tobacco use were most commonly treated, whereas lipid disorders, cardiovascular disorders, and breast disorders were most commonly referred. Several questions assessed whether the respondents usually provide or order
age-appropriate screening of several types as part of a well-woman examination. The proportion of graduates reporting that they always or almost always provide screening was 100% for breast examination, 100% for mammography, 94.9% for flexible sigmoidoscopy or colonoscopy, 89.7% for fecal occult blood testing, 82.1% for cholesterol testing, and 56.4% for diabetes screening. There were no significant differences between earlier and later graduates for performing or ordering these reported screening procedures.

The physicians were asked to estimate the proportion of patients for whom they were the woman's primary health care provider. When comparing earlier to later graduates, there was no difference for nonpregnant patients (median 10%, 0% to 70% versus median 25%, 0% to 60%, respectively, \(P = .92\)) or for pregnant patients (median 90%, 0% to 100% versus median 75%, 0% to 100%, respectively, \(P = .23\)).

### Preparation for primary care

Most graduates felt they were well prepared or somewhat well prepared to care for urinary tract infections (100%), immunizations (80.6%), tobacco use (73.5%), thyroid disease (73.5%), psychosocial problems (72.7%), diabetes (71.4%), pediatric or adolescent patients (70.6%), and hypertension (68.6%). The health problems for which graduates most often felt poorly or somewhat poorly prepared to care for were dermatologic disorders (40.0%), lipid disorders (36.4%), cardiovascular disorders (36.4%), obesity (31.4%), breast disorders (23.7%), and gastrointestinal disorders (22.2%).

There was a tendency for the later graduates to have less favorable attitudes about primary care, but most differences were not statistically significant. Fewer of the earlier than the later graduates reported that their residency prepared them at least adequately to provide primary care (54.5% versus 81.5%, respectively, \(P = .12\)). When asked, “Do you consider OB/GYN to be a primary care specialty,” 45.5% of earlier graduates agreed, compared with only 21.4% of later graduates (\(P = .23\)). Earlier graduates were more likely than later graduates to believe that residency training should be more than 4 years in duration to accommodate primary care rotations (30% versus 0%, respectively, \(P = .01\)).

### Impact on inpatient surgical procedures

There were no statistical differences between the earlier and later graduates in the average number of abdominal hysterectomies each resident performed (mean 127, median 132, range 84 to 153 versus mean 113.4, median 112, range 63 to 189, respectively, \(P = .149\)) or in the average number of vaginal hysterectomies performed (mean 55, median 58, range 39 to 77 versus mean 48, median 48, range 32 to 71, respectively, \(P = .06\)). A greater absolute number of cesarean sections were performed by the earlier graduates (mean 388, median 394, range 305 to 455 versus mean 286, median 278, range 207 to 410, respectively, \(P < .0001\)). After the number of cesarean sections performed by each resident was adjusted for the total number of cesarean sections performed every 6 months at the teaching hospital, the magnitude of the difference between the earlier and later graduates was less prominent (adjusted mean 366 versus 321 respectively, \(P = .009\)).

### Comment

The goal of the new Residency Review Committee requirement implemented in 1996 was to prepare OB/
GYN residents adequately as primary care providers. This study suggests that residency graduates provide screening services for women presenting for well-woman examinations, but they refer out certain complex primary care disorders such as management of abnormal lipid profiles. Obstetrician-gynecologists who graduated after 1996 provided about the same amount of primary care to their patients as those who graduated in 1996 or earlier. Graduates after 1996 had a tendency to believe they were better prepared to provide primary care but not to consider OB/GYN a primary care specialty, although these differences were not statistically significant. As some of our graduates wrote: “[I] don’t feel we should be primary care; otherwise we are family practice without pediatrics/males”; or “[I’m] not interested in managing all health problems”; and “[I] consider [my]self as specialist and consultant, not PCP [primary care physician].”

Despite the fact that 60.2% of residency program directors once believed that educational deficiencies would develop as the result of the 6 months primary care training, our study found only a small, nonstatistically significant decrease in the number of abdominal and vaginal hysterectomies performed by earlier and later graduates. Although the adjusted number of cesarean sections performed per resident decreased after 1996, the number of procedures remained adequate for training purposes. The post-1996 graduates from this program completed more of these procedures than the national means for all residency programs reported by Accreditation Council for Graduate Medical Education for 1997 to 2002, which were 85 abdominal hysterectomies, 35 vaginal hysterectomies, and 175 cesarean sections per resident.

This study has several limitations. Our sample included only physicians who graduated from 1 residency program, and this may limit generalizability. We had limited power to detect statistically significant differences in attitudes toward primary care. If confirmed, our findings would suggest somewhat paradoxically that more primary care training is associated with less favorable attitudes about primary care. Because only a before-after study design could be used (rather than a randomized trial), this difference in attitudes may reflect temporal changes in the practice of OB/GYN nationally, rather than a direct effect of primary care training. Our small sample size also precluded analysis of all but the 3 most common inpatient surgical procedures. Surgical procedures performed in office gynecology settings were not captured by the hospital data used for the study. Thus, we were not able to assess whether some of the less common inpatient or outpatient surgical procedures were sacrificed when residents had off-service commitments.

On the other hand, in the earlier and later periods, the response rates were essentially identical, and the career paths of graduates were very similar. This suggests that response bias was unlikely. Our survey suggests several hypotheses about the self-perceived competencies of OB/GYN physicians in providing primary care services. Because 6 months of a 4-year residency is a substantial length of time, larger studies may be worthwhile to further investigate these issues.

As OB/GYN physicians strive to fulfill their roles as primary care givers, they must maintain and update their knowledge about care of ambulatory health problems. In an age of exponentially increasing medical information, staying on top of both primary care issues and obstetrical-gynecological issues is a challenge to a busy OB/GYN physician. As a graduate wrote: “Knowing when to screen is one thing, but staying abreast of the various treatments is another.”

References

Improved performance and student satisfaction after implementation of a problem-based preclinical obstetrics and gynecology curriculum

Petra M. Casey, MD,a,* Diane Magrane, MD,b Timothy G. Lesnick, MSc

Department of Obstetrics and Gynecology, Mayo Clinic College of Medicinea; Faculty Development and Leadership Programs,b Association of American Medical Colleges; Department of Biostatistics, Mayo Clinic,c Rochester, MN

Received for publication February 17, 2005; revised May 20, 2005; accepted July 18, 2005

KEY WORDS
Problem-based learning  
Curriculum development  
Obstetrics and gynecology course

Objective: This study was undertaken to assess student performance and satisfaction as a preclinical obstetrics and gynecology course changed from a didactic to a problem-based format.

Study design: We prospectively compared examination scores and course evaluations given to 162 second-year medical students over 4 years: 2 years before and 2 years after the curricular change, to assess student performance and satisfaction with learning. We used analysis of variance for the analysis of student performance and likelihood-ratio test for analysis of student satisfaction.

Results: Mean examination scores increased from 79.1% to 84% after implementation of the problem-based curriculum ($P < .0001$). The data showed statistically significant improvement in satisfaction with course content, learning objectives, learning resources, instructional methods, and course examinations.

Conclusion: Students’ satisfaction and performance in a preclinical obstetrics and gynecology course improved significantly when problem-based learning methods were introduced.

Problem-based learning (PBL) represents the evolution from passive learning to an active, student-centered learning style.1-3 PBL involves a group of students (typically 6-10) and a faculty facilitator who guides learning via discussion.1 The PBL approach has been found effective in improving performance in a variety of specialties.4-7 Surveyed students have indicated that, compared with didactic teaching, PBL promoted a more positive attitude toward learning and a more stimulating environment for clinical knowledge acquisition.4,8,9 This educational method has been used in clinical and preclinical medical education7,10 as well as nonreproductive women’s health issues.11,12

The obstetrics and gynecology curriculum for students at Mayo Clinic College of Medicine includes a 26-session obstetrics and gynecology course in the second year, followed by a 6-week clinical clerkship in the third year. Second-year students have a dense curriculum of basic sciences and often perceive the introduction to obstetrics...
and gynecology as a distraction from learning as they prepare for the United States Medical Licensing Examination (USMLE).

The curricular change described in this study formalized an effort to engage the students in learning, to promote information acquisition, and to integrate basic science with clinical medicine. The effect of PBL on examination scores and satisfaction with various aspects of introductory obstetrics and gynecology has not been described in a prospective study in publication. Further, there are no studies investigating this effect over several years or in the setting where the curriculum includes nonreproductive women’s health issues.

Materials and methods

All second-year medical students (n = 162) enrolled in introductory obstetrics and gynecology at Mayo Clinic College of Medicine during four academic years participated in this study. There were 40 students during 2000-2001 (year 1), 42 during 2001-2002 (year 2), 41 during 2002-2003 (year 3), and 39 during 2003-2004 (year 4). For PBL, students were divided into 4 groups of 10, whereas all students attended lectures as 1 group. The project was exempted by Mayo Clinic Institutional Review Board.

The course was taught by clinical faculty including general obstetrician-gynecologists, maternal-fetal medicine specialists, and gynecologic oncologists. A domestic violence coordinator from Mayo Department of Social Services facilitated the domestic violence unit. The faculty was allotted 3 hours of preparation per hour of PBL discussion. All faculty completed a 4-hour facilitation workshop before course participation. This workshop described PBL and allowed each faculty member to facilitate a group discussion in a role-playing exercise.

The new curriculum was introduced over 4 academic years. Traditional didactic curriculum was taught during year 1. Two PBL units were introduced during year 2, while other topics were taught with lectures. During year 3, 66% of lectures were converted into PBL, followed by a complete problem-based curriculum during year 4. Domestic violence was added as a women’s health issue not adequately taught elsewhere in the curriculum.

The units were modeled after Association of Professors of Gynecology and Obstetrics (APGO) Guide to Implementing a Problem-based Learning Curriculum and Teaching Guide to Psychosocial Issues.13,14 Course content was available to students and faculty on a Web site and a printed syllabus. One author (P.M.C.) selected 26 topics in obstetrics and gynecology, which followed subject areas in the course textbook.15 Each PBL unit consisted of APGO Medical Student Educational Objectives,16 textbook reading corresponding to the learning objectives and additional content (American College of Obstetricians and Gynecologists [ACOG] technical bulletins, journal articles, prescribing information and Powerpoint presentations narrated using Camtasia [Camtasia Studio 2.1, TechSmith, Okemos, MI]). Multiple-choice self-quizzes assessed subject mastery after self-study. Students then participated in a clinical case-based discussion. Course evaluations communicated student feedback. The time requirement for each unit design was about 15 hours.

Four examinations assessed knowledge acquisition. During year 1, examinations were administered via paper and pencil as they had been before the study. To determine whether testing method would impact examination scores, 1 examination during year 2 was administered with the use of WebCT software (WebCT Campus Edition 3.8, WebCT Inc, Lynnfield, MA), whereas the others were traditional paper-and-pencil examinations. We tested for a difference between combined paper test scores and the WebCT test scores in year 2 using paired t tests. In subsequent years, all 4 examinations were Web based. To create questions of consistent quality and discriminatory value, we used APGO test bank’s multiple choice questions pertinent to the topics.17 The examination questions were intermixed, but otherwise kept consistent from year to year, allowing for meaningful comparison of scores.

We tested composite examination scores for normality using the Kolmogorov-Smirnov goodness-of-fit test, and compared mean final grades using a 1-way analysis of variance (ANOVA). Pairwise differences between the years were assessed using 3 post hoc tests (Tukey, least significant difference, and Scheffé) to adjust for multiple comparisons and cover the range from liberal (likely to find differences) to conservative (unlikely to find differences).

Student satisfaction was measured by Likert scale responses to 17 questions on a course evaluation. These measured change in satisfaction with scheduling, learning objectives, course content, instructional methods, examinations, and integration of the course into the second year curriculum as related to the teaching method. Likert scale responses of 5, 6, or 7 indicated agreement, 4 was neutral, and 1, 2, or 3 indicated disagreement with each statement. The proportions of responders who agreed with each item were compared between the years using likelihood-ratio tests for logistic
regression trend, in which the predictor variable was the proportion of PBL in the curriculum.

Medical College Admission Test (MCAT) and USMLE scores for each class of students were tested using regression analysis to uncover differences in students’ learning abilities and test taking skills.

Table II showed the mean composite examination scores in each academic year. The scores were normally distributed with significant trends (ANOVA $P < .0001$). Pair-wise comparisons of the scores during the study period using tests of mean difference showed significant differences between years 1 and 3, 1 and 4, 2 and 3, and 2 and 4. The same method revealed no significant differences in the mean scores between years 1 and 2 and between years 3 and 4.

Table II represented students’ responses to 17 survey questions encompassing 7 general categories: course organization/scheduling, learning objectives, course content, learning resources, small group format, examinations, and integration into the rest of the curriculum. Varying numbers of students responded to each question each year. Significant improvement in satisfaction was seen in five areas: learning objectives (2 items), course content (2 items), learning resources (2 of 3 items), small group format (1 of 3 items), and examinations (2 of 4 items). There was no significant improvement in course integration and organization, or in the remaining items in the other categories.

To test whether the change from paper-and-pencil to Web-based examinations had an effect on performance, we compared the aggregate scores of each of the 4 classes by examination administration method. The average score of the paper examinations during year 2 was $78.5\% \pm 5.5\%$ and the mean score on the WebCT examination was $78.6\% \pm 6.5\%$. These means did not differ by regression analysis ($P = .928$). MCAT and USMLE scores of students enrolled in the study were
comparing with investigate trends in students’ academic abilities reflected by standardized examinations. No significant trend over the years was found on either 1 of the examinations. Mean MCAT scores were 33.25 ± 0.50, \( P = .74 \) and mean USMLE scores were 232.525 ± 4.461, \( P = .85 \) by regression analysis.

Comment

Our study showed a consistent improvement in examination scores and student satisfaction in preclinical obstetrics and gynecology after implementation of PBL. It supported PBL as a valuable tool for knowledge acquisition in this specialty. Further, it controlled for biases introduced by different examination methods and student academic abilities. The latter was measured by comparing MCAT and USMLE scores. Greater academic strength resulting in higher examination scores could have been associated with better satisfaction regardless of teaching method. We found no significant trends in these parameters.

The robust improvement in written examination scores after extensive implementation of PBL confirmed that despite some difficulties with PBL teaching, our students recalled more information with this method. Years 1 and 2 curricula represented primarily didactic teaching, whereas years 3 and 4 curricula utilized primarily PBL. As expected, the mean examination scores for year 2, when only 2 PBL units were presented, did not differ from year 1. The improvement in examination scores followed the adoption of PBL for the majority of the course and was sustained during the complete PBL implementation. Thus, the observed change was not likely because of enthusiasm for a new learning method.

The interpretation of student satisfaction with introductory obstetrics and gynecology, was more complex. The students communicated that the didactic format did not provide adequate learning and that they preferred the PBL format taught in other courses. Once the PBL format was implemented, the students indicated improved satisfaction with many aspects of the course. The most dramatic improvement was seen when years 1 and 4 (all didactic vs all PBL) were compared.

Areas showing improved satisfaction included the relationship of learning objectives to concepts and appropriate level of detail and learning objective clarity, the relationship of basic science and clinical medicine apparent in well-organized course content, Web site and course material quality, facilitation of conceptual understanding during class activities and the testing of important concepts and appropriate detail. There were no deleterious effects of PBL on any parameters measured.

Interestingly, some satisfaction items contradicted each other. For example, as a result of classroom activities, students reported an improvement in understanding of important concepts, but not of the relationship between basic science and medical practice. On the other hand, they indicated that course content related basic science to practice better with the PBL format. This may have been artifactual because of the sample size of responders or variable response rate.

The preference for small group instruction over lectures item required the most careful interpretation. How well prepared were the students to make this decision? Scores from year 3, when the students had a direct basis for comparison led to the most positive rating of the PBL curriculum. During years 1 and 4, all material was either lecture or PBL providing the students with only indirect basis to compare the 2 methods via PBL teaching in other courses. Some students simply preferred lectures, shown to facilitate short-term learning appropriate for multiple-choice assessment. Thus, lectures may have represented an advantage within the curricular density of the second year and a better match for the examination method. Interestingly, though the students did not track the number of hours needed for study, the workload was not perceptibly increased with PBL perhaps because we dedicated 1 hour of study time per hour of discussion.

Some of the aspects showing significant change represented course organization rather than the teaching method. These included the scheduling and learning resources items.

Lastly, certain course evaluation items did not show improvement. It was reassuring that the proportion of positive responses to the course schedule, textbook, gynecologic examination instruction, and examination method, purposely kept constant during the study, remained stable. This validated our survey instrument. Similarly, the consistently low scores on the integration item suggested that the emphasis on USMLE overshadowed clinical courses regardless of instruction method. Overall, the curricular change to PBL teaching led to a dramatically increased satisfaction with learning in obstetrics and gynecology.

This study also presented an alternative for curricular development. Previously described strategies included PBL and didactic courses in parallel, abrupt transitions from one teaching method to another and a pilot PBL course in the setting of didactics. We offered a model where gradual implementation of a curricular format was undertaken over several years. The advantages of this model included continuous feedback from students and faculty for identification of areas needing adjustment, increased faculty comfort level with the new teaching method on a smaller scale before full implementation and the monitoring of scores and satisfaction parameters for unanticipated negative outcomes.

The largest challenge with implementing a PBL course within our small obstetrics and gynecology department was securing time for course design, faculty development/
training and course participation without negatively impacting patient care. PBL format required 4 times the number of faculty hours to teach the same material. The APGO Guide to Implementing a Problem-based Learning Curriculum and Teaching Guide to Psychosocial Issues provided obstetric and gynecologic cases and a chapter on “Violence against Women” with learning issues. The use of these resources considerably reduced time for PBL unit design. By using existing APGO cases enabled 1 author (P.M.C.) to devote time to course content, examination, and course evaluation development. Consistent design of the units allowed various faculty members to effectively facilitate any PBL unit. However, multiple teaching sessions per subject and the occasional faculty substitution introduced some content variability, a study limitation. Although we made every effort to maximize student contact with a core of proven educators, throughout the study, there were occasional unavoidable changes. We received full support of department and medical school leadership for additional faculty time.

Finally, given our relatively low and variable response rates to the course evaluations, did the evaluations accurately reflect the class viewpoint? Further, because the evaluations followed the last examination, was the students’ satisfaction with the course altered by their impression of the examination? At the Mayo Clinic School of Medicine, students completed the course evaluation as a prerequisite to the release of grades. They were not required to answer every item and could submit verbatim comments alone without the Likert scale evaluation. This explains the lower response rates in some years. A collaborative multicenter study with a larger sample size could address this bias provided a comparable curriculum could be located.

In summary, the robust improvement in examination scores and student satisfaction with implementation of PBL in preclinical obstetrics and gynecology should encourage others to apply this method to their settings. Our study presented a gradual implementation of curricular change, multiyear performance and satisfaction data analysis and extensive use of APGO materials in course design. These aspects of our study may be readily adopted elsewhere.

Acknowledgment

We gratefully acknowledge the mentorship of Dr Andrew E. Good, and a Faculty Development Grant from Mayo Clinic College of Medicine.

References

Letters to the Editors

Risk assessment for neonatal respiratory distress syndrome with FLM II combined with gestational age

To the Editors: The article by Parvin et al1 provides the clearest set of data to date that incorporate gestational age into the evaluation of fetal lung maturity.

The authors of this study failed to mention our response in the letters to the editor section in regards to differences in quantitative fluorescence polarization-based fetal lung maturity assay (TDx-FLM) I versus TDx-FLM II.2 Despite a change in reagents, the actual numeric values on paired specimens that were tested contemporaneously were nearly identical (±4.0%).

Differences in the prevalence rate of respiratory distress syndrome (RDS) have led to several different recommended cutoff points, regardless of the version of FLM test kits that were used.2,3 A large part of the difference in the prevalence rate can be explained simply on the basis of the differences in the gestational ages that were being tested for fetal lung maturity in the published studies. Indeed, the prevalence of RDS changes dramatically between 30 and 36 weeks of gestation.4 On the basis of gestational age alone, most fetuses <30 weeks of gestation will be at risk of RDS, whereas few fetuses after 36 weeks of gestation will experience RDS. Obtaining FLM results at 39+ and <30 weeks of gestation is arguably not very useful.

Even within their study, the prevalence of RDS was different among the 3 groups. A selection bias toward including patients with RDS resulted from an increased number of RDS cases between 32 and 36 weeks of gestation in group 3. Bias towards including cases with RDS illustrates the potential problem with retrospective studies. Furthermore, the inclusion of patients at <30 and >39 weeks of gestation is not useful and potentially skews the overall prevalence. Eliminating the cases <30 weeks of gestation leaves approximately 50 cases of RDS distributed over 10 weeks of gestation. There appears to be only 1 case of RDS after 37 weeks of gestation of 131 patients who were tested. Notably, the number of patients who were enrolled between 30 and 34 weeks of gestation was small (2, 6, 14, 30, and 49, respectively). Owing to the higher incidence of RDS during those gestational ages (55%, 40%, 35%, 25%, and 15%, respectively), a larger number of patients at those gestational ages may provide more meaningful data.

Although odds ratios are statistically more applicable, given the potential differences in prevalence between studies, they are much more difficult to use conceptually when risks of RDS are being discussed with patients and colleagues.

The collection of a large number of gestational age-specific data between 30 and 36 weeks with gestational age-specific analysis would provide better risk estimates.

Michael G. Pinette, MD*
Joseph Wax, MD
*Maine Medical Center
Department of Obstetrics and Gynecology
Division of Maternal-Fetal Medicine
887 Congress St, Suite 200
Portland, ME 04102
E-mail: cartia@mmc.org

References


0002-9378/S - see front matter © 2005 Mosby, Inc. All rights reserved.
doi:10.1016/j.ajog.2005.05.022
Reply

To the Editors: We appreciate the acknowledgement by Drs Pinette and Wax that our article provides the clearest set of data to date to incorporate gestational age into the evaluation of fetal lung maturity. We agree that overall prevalence of RDS that is observed at a particular institution will depend on the distribution of the gestational ages that are tested. We discussed this point in our article. We also agree that estimates of a single TDx-FLM II cutoff for all gestational ages will be influenced by the distribution of the gestational ages that are tested. The purpose of our study was not to estimate a single TDx-FLM II cutoff but to model RDS risk as a function of TDx-FLM II ratio and gestational age. Two of our study sites collected all available data for physician-ordered TDx-FLM II results; the third site sought additional cases of RDS, independently of gestational age at the time of fetal lung maturity testing. As stated in our article, there was no statistically significant difference among study sites in the rate of change of RDS risk as a function of gestational age, but we did find a significantly higher overall prevalence of RDS at the third study site; both observations are consistent with the third study site’s identification of additional cases of RDS, independent of gestational age. Data based on consecutive series of subjects allow unbiased estimation of risk. Therefore, in addition to presenting tables of odds ratios that reflect only the rate of change of risk (but as Drs Pinette and Wax point out are also more difficult conceptually), we also were able to present tables of absolute risk by fitting a separate baseline prevalence term for the third study site and adjusting for their “biased” overall prevalence.

Last, they state that they found only a 4% difference between the TDx-FLM I and II in a study of 47 samples, which suggests that the 2 tests are “nearly identical.”

Two larger studies suggest that this is not the case. One of our laboratories (J.F.C.) participated in the multicenter testing of TDx-FLM I versus FLM II (n = 778 patients). The following regression data were from this study: slope, 0.742; Y intercept, 1.54 (r = 0.962), which resulted in FLM II ratios that were between 10% and 25% lower than FLM I ratios. A study by Moxness et al (n = 63 patients) reported the following data: slope, 0.9; Y intercept, −0.9 (r = 0.99), which resulted in FLM II ratios that were between 10% and 20% lower than FLM I ratios. Together these data suggest the 2 assays are not “nearly identical” and the differences should not be ignored.

C. A. Parvin, PhD
L. A. Kaplan, PhD
J. F. Chapman, DrPH
T. G. McManamon, PhD
A. M. Gronowski, PhD*

*Department of Pathology and Immunology
Division of Laboratory Medicine
Washington University School of Medicine
660 S. Euclid, Box 8118
St. Louis, MO 63110
E-mail: gronowski@pathology.wustl.edu

References


Lymphatic mapping for endometrial cancer: Is hysteroscopic injection a safe technique for sentinel lymph node biopsy?

To the Editors: We read with interest the article by Raspagliesi et al demonstrating the relevance of intraoperative lymphatic mapping using hysteroscopic injections of tracers in patients with endometrial cancer. In a series of 18 patients, they injected radiocolloid and blue dye subendometrially around the lesion during hysteroscopy, and detected sentinel nodes (SN), by laparotomy, in all cases. One previous study has described
laparotomic identification of SNs under hysteroscopic guidance in patients with endometrial cancer. \(^2\) They injected radiocolloid under the endometrium during hysteroscopy, and detected SNs, by laparotomy, in 82% of cases. \(^2\)

This article raises several concerns. The first is the choice of injection sites. Two injection sites—subserosal intraoperative myometrial and pericervical—have been previously used to detect SN in women with endometrial cancer. The former approach, using blue dye, was initially used because it reflected the anatomic drainage of the corpus uteri. \(^3\) However, this technique is controversial, as combined detection is not possible, and the number and location of myometrial injection sites have not yet been standardized. The latter approach, which permits combined detection with patent blue and radiocolloid, is well standardized, simple and reproducible; it yields a higher SN detection rate, but does not reflect the anatomic drainage of the tumor. \(^4\) Hence, intrauterine injection, mimicking the natural lymphatic drainage of endometrial cancer, is an attractive alternative. Raspagliesi et al \(^1\) injected tracer under hysteroscopic guidance. In our experience, patent blue injection under hysteroscopic guidance was difficult, as the endometrium was obscured by dye diffusion. Moreover, injection “under the endometrium” probably corresponds in fact to injection in the superficial myometrium.

The risk of cancer cell dissemination into the abdominal cavity must also be discussed. Despite the controversial prognostic significance of positive cytology, previous studies have shown a risk of tumor cell dissemination and metastasis after hysteroscopy. To avoid this risk, we injected patent blue under hysteroscopic guidance at the beginning of surgery, after coagulating the Fallopian tubes.

Raspagliesi et al \(^1\) and Niikura et al \(^2\) confirmed the relevance of the SN biopsy in patients with endometrial cancer, but did not obtain a higher SN identification rate and bilateral SN detection rate than in previous studies using dual labels injected pericervically. \(^4\)

Finally, contrary to the laparoscopic SN approach, the procedure used by Raspagliesi et al, \(^1\) involving preoperative hysterectomy with uterine distension and laparotomy, is not compatible with the concept of minimally invasive surgery, which is particularly beneficial for these often obese and elderly women with underlying general health disorders.

Emmanuel Barranger, MD*
Serge Uzan, MD
Emile Darai, MD, PhD
*Department of Gynecology and Obstetrics
Hôpital Tenon
4 rue de la Chine
75020, AP-HP
Paris, France
E-mail: emmanuel.barranger@tnn.ap-hop-paris.fr

References

Reply

To the Editors: We appreciated very much the comments of Dr Barranger et al regarding our article entitled “Hysteroscopic injection of tracers in sentinel node detection of endometrial cancer: A feasibility study.”

Regarding the site of injection, we named it the sub-endometrial layer because the patient is awake during the hysteroscopic injection. We avoided the myometrial injection because it elicits pain caused by the presence of neural fibres in this stratum. Moreover, the myometrial injection increases significantly the rate of intravascular leakage that reduces the nodal uptake of the tracers.

Visual impairment during hysteroscopic injection is common during blue dye (BD) injection because of its leakage from the injection site. However, it is limited to a few seconds when a continuous flow hysteroscope...
Evidence does not support cervical preservation

To the Editors: Just when we thought the case for cervical preservation during hysterectomy has been settled after recently published studies, we were surprised to see a review article defending supracervical hysterectomy.1 We believe the fundamental controversy regarding laparoscopic supracervical hysterectomy is not whether it is superior to laparoscopic-assisted vaginal hysterectomy or endometrial resection. The ultimate issue with any supracervical hysterectomy technique is whether preservation of the cervix has any advantage or not. Because all the theoretical benefits of laparoscopic supracervical hysterectomy are associated with saving the cervix, level I or II evidence from 4 recent well-designed studies comparing total either abdominal hysterectomy or vaginal hysterectomy with subtotal abdominal hysterectomy should suffice to answer this question for laparoscopic supracervical hysterectomy, as well. Briefly, all of these studies found no statistically or clinically significant advantage to cervical preservation.2-5 Moreover, Okaro et al suggested that it might even have long-term disadvantages. They reported that 16 of 70 (22%) women who were followed for a mean of 66 months after laparoscopic supracerclival hysterectomy required trachelectomy.6

Finally, in any area of medicine, it is the surgeon’s goal to develop the least invasive, the fastest, the least complicated, and the most effective operative techniques with the shortest hospital stay and lowest cost.7 In the case of hysterectomy, it is hard to beat the vaginal route from all of these perspectives.

Oz Harmanli, MD
Stephen A. Metz, MD
Department of Obstetrics and Gynecology
Baystate Medical Center
Springfield, MA
E-mail: oz.harmanli@bhs.org

References

2. Roovers JP, van der Bom JG, van der Vaart CH, Heintz AP. Hysterectomy and sexual wellbeing: prospective observational study of

Reply

To the Editors: I appreciate Drs Harmanli and Metz’s response to my review of laparoscopic supracervical hysterectomy (LSH). The article was a review of the existing LSH literature and was not intended as a “defense” of supracervical hysterectomy nor as a comprehensive comparison of hysterectomy techniques. I agree that vaginal hysterectomy is an effective hysterectomy technique with minimal morbidity. However, some patients either for surgical reasons or personal preferences are not candidates for a vaginal hysterectomy. Therefore, it is imperative that we investigate less morbid techniques than abdominal hysterectomy.

I disagree with the assertion that “all the theoretical benefits of LSH are associated with saving the cervix.” After an LSH, most patients return to normal work and personal activities within 2 weeks of the procedure (Tables I, II, and III), as opposed to the standard 6 weeks after total abdominal or vaginal hysterectomy. The earlier return-to-function results in significant financial savings related to lost time at work or costs of childcare during recovery. Furthermore, although the recent trials referenced in your letter have advanced our knowledge of abdominal supracervical hysterectomy versus abdominal hysterectomy, these procedures involve an abdominal incision that results in significant postoperative pain and infectious risk. LSH eliminates the need for an abdominal incision and therefore may lower this morbidity.

Finally, the trials referenced in your letter showed no clinical or statistical difference between supracervical and total hysterectomy; however, they found no strong disadvantages to cervical preservation. The 22% subsequent trachelectomy rate reported by Okaro et al is the highest published in the literature to date. LSH typically has a 5% postoperative complication rate (Tables I-III), which compares favorably with vaginal hysterectomy, and most patients undergoing LSH do not require further therapy. As stated in the review, all patients considering a LSH should be aware of the potential for postoperative vaginal bleeding, cervical dysplasia, or the rare need for trachelectomy.

In conclusion, I concur that vaginal hysterectomy should be our goal when it is acceptable to the patient and is technically feasible. However, for those patients who are not candidates, I submit that LSH is a promising technique given its previously enumerated benefits. LSH still requires further study to determine its appropriate role in our surgical armamentarium.

Todd R. Jenkins, MD
Carolina Medical Center
Obstetric and Gynecology
1000 Blythe Blvd
PO Box 32861
Charlotte, NC 28232
E-mail: todd.r.jenkins@carolinashealthcare.org

References

Fetal monitoring with the ST analyser: Need for a long-term follow-up of the infants

To the Editors: The recent publication of Devaitis et al showed that when attending physicians were unaware to the ST analysis information, only using available cardiotocography (CTG) as per current practice, the clinical STAN guidelines had a sensitivity of less than 50% for metabolic acidemia at birth (BD >12 mmol/L). Similar results were reported by other perinatologists with the combination of CTG and STAN: sensitivity of 30% for Haberstich et al and 46% for Kwee et al for the same degree of metabolic acidosis.

Two questions have to be answered before introducing this technology for routine use in the labor wards, whereas its cost is twice higher than conventional CTG: Does the STAN lower operative deliveries in comparison with CTG only? In the Swedish multicentric study, it is obvious that ST analysis plus CTG lowers the rate of operative deliveries for fetal distress (relative risk 0.83 [95% CI 0.69-0.99]).

Is the neonatal outcome better with the STAN? In the CTG-ST group of the Swedish trial versus CTG alone, metabolic acidosis was effectively lowered (relative risk 0.47 [0.25-0.86]) when defining fetal asphyxia from a pH <7.05 and a BD_{ecf} >12 mmol/L, but no significant differences between the groups were found regarding Apgar scores, admissions to neonatal intensive care, or neonatal encephalopathy. In fetal lambs, the rising T/QRS reflects a catecholamine surge, activation of beta adrenoreceptors, myocardial glycogenolysis and metabolic acidosis, whereas the appearance of biphasic and negative waveforms are late markers for severe decompensation.

From a complementary analysis of the Swedish trial conducted by Noren et al, it is also obvious that the STAN events are associated with a marked metabolic acidosis (mean pH at 7.07 and BD_{ecf} at 10.2 mmol/L for biphasic ST + T/QRS rise and mean pH at 7.04 and BD_{ecf} at 10.2 mmol/L for T/QRS rise only).

The real issue is: must we wait for this degree of metabolic acidosis before intervening? In a recent paper by Victory et al, mean arterial pH in the umbilical cord at birth was 7.24 ± 0.07 and BE was at −5.6 ± 3.0 mmol/L. The values of pH (7.05) and of BE_{ecf} (−12 mmol/L) selected by Amer-Wählin et al are lower than 2 SD, as defined by Victory et al, who claimed that with range values of umbilical cord blood gases at 3 SD to 2 SD below the mean there was a substantial increase in the incidence of Apgar scores less than 7 at 5 minutes (×10), for NICU admissions (×4), and for assisted ventilation (×3).

So, further long-term neurodevelopmental studies of the infants monitored by STAN+CTG are needed before attesting that fetal brain cells are not more sensitive to metabolic acidosis than the fetal myocardium and before validating the STAN technology for standard intrapartum fetal surveillance.

Georges J. J. Boog, MD*
*Department of OBGYN
University of Nantes
Regional Perinatal Center
38, Avenue Jean Monnet
44093 Nantes cedex 1
France
E-mail: georges.boog@chu-nantes.fr

References
To the Editor: Dr Boog has made important comments related to our article reporting the first North American clinical study on the ability of the STAN clinical guidelines to predict metabolic acidemia during labor. We also appreciate bringing to our attention the article from Kwee et al,\(^1\), which was not published at the time of submission. We reported an incidence of poor ECG quality of 11% of the tracing time that is almost identical to the incidence of 10% of technical failure reported by Kwee et al,\(^1\) which further supports 1 of our conclusions that these technical issues need to be addressed if this new technology is to have clinical application.

The ultimate goal of fetal health surveillance is to minimize the risk of intrapartum asphyxia with its potential neurologic sequelae and to keep perinatal morbidity and mortality to a minimum. The Canadian guidelines indicate a threshold for intervention during labor should be an umbilical artery pH falling below 7.15 with a base deficit less than \(-12\) mmol/L. These values are close to \(-2\) SD from population-based umbilical artery cord gases we previously published (pH < 7.13 and base excess < \(-10.3\) mmol/L, respectively).\(^2\) Above these thresholds, perinatal morbidity and mortality in a normally formed neonate should therefore be minimized but not completely eliminated. It is also important that a fall in pH below these thresholds can occur fairly rapidly because pH is a nonlinear logarithmic measure of the hydrogen ions concentration. Victory et al\(^3\) also demonstrated an exponential rapid increase in neonatal morbidity when the umbilical artery pH fell from \(-2\) SD to \(-3\) SD, which further supports the fact that an intervention threshold for umbilical artery pH of 7.15 is a prudent choice as a primary outcome variable as used in our study. The Swedish randomized trial evaluating the STAN monitor used a relatively low cutoff for umbilical artery pH of 7.05 with a base excess in the extracellular fluid of less than \(-12\) mmol/L. Because no follow-up data on long-term neurologic outcome from the Swedish randomized trial are available, until such information is published, a higher more prudent threshold for an umbilical artery pH as we elected to choose for our validation trial should be used. Finally, it is increasingly evident that the STAN technology within its current state cannot be applied yet for standard intrapartum fetal health surveillance because both our validation trial and the European validation trial\(^1\) demonstrated a sensitivity below 50% for the detection of clinically relevant onsetting fetal metabolic acidosis during labor with a technical failure rate of 10%.

Robert Gagnon, MD
Kristina L. Dervaitis
Department of Obstetrics and Gynecology
University of Western Ontario
St. Joseph’s Health Care London
268 Grosvenor Street, London, ON, N6A 4V2
Canada
E-mail: rgagnon@uwo.ca

References
Communications

Communications regarding editorial management should be addressed to the appropriate Editor. For mailing address, telephone and fax numbers, e-mail address, and detailed instructions on manuscript submission, see Information for Authors in the first issue of each volume and on our Web site at www.ajog.org.

Address business communications to Periodical Publisher, Mosby, Inc, 11830 Westline Industrial Dr., St. Louis, MO 63146-3318.

Customer service (orders, claims, online, change of address): Elsevier Periodicals Customer Service, 6277 Sea Harbor Drive, Orlando, FL 32887-4800. Tel: (800) 654-2452 (U.S. and Canada); (407) 345-4000 (outside U.S. and Canada). Fax: (800) 225-4030 (U.S. and Canada); (407) 363-9661 (outside U.S. and Canada). E-mail: elseps@elsevier.com.

Address changes must be submitted four weeks in advance.

Yearly subscription rates: United States and possessions: Individual $212; Institution $470; Student/Resident $106. All other countries (prices include airspeed delivery): Individual $260; Institution $522; Student/Resident $130. Single Issues $44. To receive student/resident rate, orders must be accompanied by name of affiliated institution, date of term and the signature of program/residency coordinator on institution letterhead. Orders will be billed at the individual rate until proof of status is received. Current prices are in effect for back volumes and back issues.

Further information on this journal is available from the Publisher or from this journal’s website (http://www.ajog.org). Information on other Elsevier products is available through Elsevier’s website (http://www.elsevier.com).

Advertising information. Advertising orders and inquiries can be sent to: USA, Canada, and South America, Elsevier Inc, John Maney, 360 Park Avenue S, New York, NY 10021; phone: 212-633-3855; fax: 212-633-3820; e-mail: j.maney@elsevier.com. Classified advertising orders and inquiries can be sent to Traci Peppers, Elsevier, 360 Park Avenue South, New York, NY 10010; phone (212)633-3766. Europe and the rest of the world, Julie Toop; phone +44 (0) 1865 843016; fax +44 (0) 1865 843976; e-mail media@elsevier.com.

Author inquiries

For inquiries relating to accepted articles (including electronic submission where available), please visit Elsevier's Author Gateway at http://authors.elsevier.com. The Author Gateway also provides the facility to track accepted articles and set up e-mail alerts to inform you of when an article’s status has changed, as well as detailed artwork guidelines, copyright information, frequently asked questions, and more. Please see Information for Authors for individual journal. Contact details for questions arising after acceptance of an article, especially those relating to proofs, are provided after registration of an article for publication.

The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).

Reprints. For queries about author offprints, e-mail author-support@elsevier.com. To order 100 or more reprints for educational, commercial, or promotional use, contact the Commercial Reprints Department, Elsevier Inc., 360 Park Avenue South, New York, NY 10010-1710. Fax: (212) 462-1935; e-mail reprints@elsevier.com. Reprints of single articles available online may be obtained by purchasing Pay-Per-View access for $30 per article on the journal website, www.ajog.org.

Copyright 2005 Mosby, Inc. All rights reserved.


Microform edition available from ProQuest Information and Learning, 300 N Zeeb Rd, Ann Arbor, MI 48106-1346; telephone (734)761-4700 or (800)521-0600.

Photocopying: Single photocopies of single articles may be made for personal use as allowed by national copyright laws. Permission of the Publisher and payment of a fee is required for all other photocopying, including multiple or systematic copying, copying for advertising or promotional purposes, resale, and all forms of document delivery. Special rates are available for educational institutions that wish to make photocopies for non-profit educational classroom use.

Permissions may be sought directly from Elsevier’s Rights Department in Philadelphia, PA, USA: phone (+1) 215 239-3804, fax (+1) 215 239-3805, e-mail healthpermissions@elsevier.com. Requests may also be completed on-line via the Elsevier homepage (http://www.elsevier.com/locate/permissions).

In the USA, users may clear permissions and make payments through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA; phone: (+1) (978) 7508400, fax: (+1) (978) 7504744, and in the UK through the Copyright Licensing Agency Rapid Clearance Service (CLARCS), 90 Tottenham Court Road, London W1P 0LP, UK; phone: (44) 20 7631 5555; fax: (44) 20 7631 5500. Other countries may have a local reprographic rights agency for payments.

Derivative works: Subscribers may reproduce tables of contents or prepare lists of articles including abstracts for internal circulation within their institutions. Permission of the Publisher is required for resale or distribution outside the institution.

Permission of the Publisher is required for all other derivative works, including compilations and translations.

Electronic storage or usage: Permission of the Publisher is required to store or use electronically any material contained in this journal, including any article or part of an article.

Except as outlined above, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior written permission of the Publisher.

Address permissions requests to: Elsevier Rights Department, at the fax and e-mail addresses noted above.

Notice: No responsibility is assumed by the Publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made.

Although all advertising material is expected to conform to ethical (medical) standards, inclusion in this publication does not constitute a guarantee or endorsement of the quality or value of such product or of the claims made of it by its manufacturer.