1. Editorial Board  
   Page IFC
2. Table of Contents  
   Page A3
3. Information for Readers  
   Page A4

Editorial

4. Financial Support for Research Training and Career Development in Complementary and Alternative Medicine from the National Institutes of Health  
   Pages 483-490  
   Partap S. Khalsa and Nancy J. Pearson

Original Articles

6. Comparison of Posteroanterior Spinal Stiffness Measures to Clinical and Demographic Findings at Baseline in Patients Enrolled in a Clinical Study of Spinal Manipulation for Low Back Pain  
   Pages 493-500  
   Edward F. Owens Jr., James W. DeVocht, M. Ram Gudavalli, David G. Wilder and William C. Meeker

7. The Reproducibility of a Clinical Grading System of Motor Control in Patients with Low Back Pain  
   Pages 501-508  
   Negin Sedaghat, Jane Latimer, Christopher Maher and Trish Wisbey-Roth

8. The Effect of Combining Manual Therapy with Exercise on the Respiratory Function of Normal Individuals: A Randomized Control Trial  
   Pages 509-513  
   Roger M. Engel and Subramanyam Vemulpad
9. Interexaminer Reliability of the Prone Leg Length Analysis Procedure  
   Pages 514-521  
   Michael Schneider, Robert Homonai, Brian Moreland and Anthony Delitto

    Pages 522-526  
    Usama Albastaki, Dimitris Sophocleous, Jan Göthlin and Claude Pierre-Jerome

11. Development of an Evidence-Based Application and Rubric for Evaluating Applicants’ Qualifications for Promotion to Professor  
    Pages 527-535  
    Glenda C. Wiese, Robert E. Percuoco, Joel G. Pickar, Stephen M. Duray, Saeed R. Faruqui, Gilbert O. Schmiedel and Ian D. McLean

Case Reports

    Pages 536-538  
    Ozgur Ozdemir, Tarkan Calisaneller, Erkan Yildirim and Nur Altinors

13. A Case of a Potential Manipulation Responder Whose Back Pain Resolved with Flexion Exercises  
    Pages 539-542  
    Stephen May and Richard Rosedale

14. Early-Onset Multiple Myeloma: An Illustrative Case Report  
    Pages 543-549  
    Rod Kaufman
TABLE OF CONTENTS

EDITORIAL

483 Financial Support for Research Training and Career Development in Complementary and Alternative Medicine From the National Institutes of Health
Parap S. Khalsa, DC, PhD, and Nancy J. Pearson, PhD

491 JMPT Highlights

ORIGINAL ARTICLES

493 Comparison of Posteroanterior Spinal Stiffness Measures to Clinical and Demographic Findings at Baseline in Patients Enrolled in a Clinical Study of Spinal Manipulation for Low Back Pain
Edward F. Owens Jr., MS, DC, James W. DeVocht, DC, PhD, M. Ram Gudavalli, PhD, David G. Wilder, PhD, and William C. Meeker, DC, MPH

501 The Reproducibility of a Clinical Grading System of Motor Control in Patients With Low Back Pain
Negin Sedaghat, BAppSc (Phy)(Hons), Jane Latimer, GradDipAppSc (Manip Phty), PhD, Christopher Maher, GradDipAppSc (Ex&Sport Sci), GradDipAppSc (Manip Phty), PhD, and Trish Wisbey-Roth, MSp.Phty, BApp Sc

509 The Effect of Combining Manual Therapy With Exercise on the Respiratory Function of Normal Individuals: A Randomized Control Trial
Roger M. Engel, DC, DO, and Subramanyam Vemulpad, MSc, PhD

514 Interexaminer Reliability of the Prone Leg Length Analysis Procedure
Michael Schneider, DC, Robert Homonai, DC, Brian Moreland, DC, and Anthony Delitto, PhD, PT

522 Magnetic Resonance Imaging of the Triangular Fibrocartilage Complex Lesions: A Comprehensive Clinicoradiologic Approach and Review of the Literature
Usama Albastaki, MD, Dimitris Sophocleous, MD, Jan Göthlin, MD, PhD, and Claude Pierre-Jerome, MD, PhD

527 Development of an Evidence-Based Application and Rubric for Evaluating Applicants’ Qualifications for Promotion to Professor
Glenda C. Wiese, PhD, Robert E. Percuoco, DC, Joel G. Pickar, DC, PhD, Stephen M. Duray, PhD, Saeed R. Faruqui, PhD, Gilbert O. Schmiedel, DC, and Ian D. McLean, DC

CASE REPORTS

536 Acute Intracranial Subdural Hematoma After Epidural Steroid Injection: A Case Report
Ozgur Ozdemir, MD, Tarkan Calisaneller, MD, Erkan Yildirim, MD, and Nur Altnors, MD

539 A Case of a Potential Manipulation Responder Whose Back Pain Resolved With Flexion Exercises
Stephen May, MSc, and Richard Rosedale, PT, Dip MDT

543 Early-Onset Multiple Myeloma: An Illustrative Case Report
Rod Kaufman, DC

Full-text online access to JMPT is available for all print subscribers. See page 521 for details.
Journal of Manipulative and Physiological Therapeutics

INFORMATION FOR READERS

Questions or Comments?
Correspondence regarding editorial matters should be addressed to the Editor: Claire Johnson, DC, MS Ed, Journal of Manipulative and Physiological Therapeutics, 200 E. Roosevelt Rd, Lombard, IL 60148-4583.

Instructions for Authors
At any time the instructions for authors can be viewed at www.mosby.com/jmpt or can be obtained from the editor. Authors should consult these instructions before submitting manuscripts.

Advertising Representative
Alexandra Leonardo, Elsevier, Inc., 360 Park Avenue South, New York, NY 10010-1710. Phone: 212-633-3649; fax: 212-633-3820; e-mail: e.leonardo@elsevier.com.

Business Communications

Subscriptions
Subscriptions may begin at any time. To enter a subscription to the Journal of Manipulative and Physiological Therapeutics, call 800-654-2452 or 407-345-4000; fax 407-363-9661; or e-mail elsepscs@elsevier.com. Remittances made by check, draft, post office, or express money order should be in US funds, drawn through a US bank, made payable to this Journal, and sent to Elsevier Inc, Subscription Customer Service, 6277 Sea Harbor Dr, Orlando, FL 32887, USA.

2007 US subscription rates: individual, $126.00; institution, $235.00; student and resident, $62.00. Outside of the US and possessions: individual, $166.00; institution, $280.00; student and resident, $83.00. Canadian customers, please add 7% GST to international prices. Prices subject to change without notice. Subscriptions include supplements and full-text online access. 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 letter-head. Orders will be billed at individual rate until proof of status is received.

Single-copy prices will be charged on missing issues older than 3 months (6 months international) from mail date. Back issues generally are available for the previous 5 years.

Indexing/Abstracts
The Journal is indexed or abstracted in Index Medicus, EM-Base/Excerpta Medica, Index to Chiropractic Literature, Current Contents/Clinical Medicine, Russian Academy of Sciences, and BS/CCML Database. Volume index appears in the December issue.

Reprints
To order author reprints, contact Periodical Reprints at 800-325-4177, ext 4350, or 314-453-4350; fax 314-579-3358; e-mail: author.reprints@elsevier.com. To order 100 or more reprints for educational, commercial, or promotional use, contact the Commercial Reprints Department, Elsevier Inc, 360 Park Ave 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 $10.00 per article on the journal Web site, www.mosby.com/jmpt.

Copyright
© 2007 National University of Health Sciences. All rights reserved.

This Journal and the individual contributions contained in it are protected under copyright by National University of Health Sciences, and the following terms and conditions apply to their use:

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 nonprofit 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 online 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 circulations within their institutions. Permission of the Publisher is required for resale or distribution outside the institutions. 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.
FINANCIAL SUPPORT FOR RESEARCH TRAINING AND CAREER DEVELOPMENT IN COMPLEMENTARY AND ALTERNATIVE MEDICINE FROM THE NATIONAL INSTITUTES OF HEALTH

Partap S. Khalsa, DC, PhD, and Nancy J. Pearson, PhD

ABSTRACT

Research careers are a relatively new reality for complementary and alternative medicine (CAM) practitioners (eg, chiropractors, naturopaths, doctors of oriental medicine, etc). Before the establishment in 1998 of the National Center for Complementary and Alternative Medicine (NCCAM) as part of the National Institutes of Health (NIH), there were few funding resources available for those interested in a CAM research career and fewer still feasible paths. Now, however, NCCAM provides a broad array of research training and career development awards for those seeking a long-term career in CAM research. These awards include predoctoral and postdoctoral fellowships, individual career development awards, and institutional training awards. The goal of this article is to provide information about current research training funding opportunities from NCCAM and NIH as a whole that are available to CAM practitioners in the context of the historical challenges of transitioning from a clinical career in CAM practice to a CAM research career.

Key Indexing Terms: Research; Research Support; Fellowships and Scholarships; Complementary Therapies

Resear
careers are a relatively new reality for complementary and alternative medicine (CAM) practitioners (eg, chiropractors, naturopaths, doctors of oriental medicine, etc).¹ Before the establishment in 1998 of the National Center for Complementary and Alternative Medicine (NCCAM) as part of the National Institutes of Health (NIH), there were few resources available for those interested in a CAM research career and fewer still feasible paths.²,³ As a consequence, there were only a few intrepid CAM practitioners who sought the requisite research training to begin a career in CAM research.⁴,⁵ Since that time, and in the last seven years in particular, new funding opportunities as well as formal academic programs in CAM research have become available.⁶ The March 2002 White House Commission on Complementary and Alternative Medicine Policy report,⁷ the 2005 Institute of Medicine of the National Academies report on Complementary and Alternative Medicine in the United States,⁸ and 2 5-year Strategic Plans of NCCAM⁹ all have emphasized the vital importance of training and funding an appropriate cohort of CAM practitioners to be able to conduct rigorous biomedical research. The goal of this article is to provide information about the range of current research training funding opportunities from NCCAM and NIH as a whole that are available to CAM practitioners in the context of the historical challenges of transitioning from a clinical career in CAM practice to a CAM research career.

DEFINITION OF CAM PRACTITIONERS

Complementary and alternative medicine practitioners are trained in healing practices that are not part of conventional (ie, allopathic) medical care. In the United States, most CAM practitioners with terminal doctoral degrees include Doctor of Acupuncture and Oriental Medicine (DAOM), Doctor of Chiropractic (DC), and Doctor of Naturopathy (ND). In
addition, Doctors of Osteopathy (DO) who have been taught and use Osteopathic Manual Medicine are also included by NCCAM as CAM Practitioners with doctoral degrees. There are numerous other CAM professions where a doctoral degree is not the common terminal degree, some of whom are licensed (eg, licensed massage therapist) and others are not (eg, yoga therapist).

Traditional Path for Research Careers

The traditional path to, and de facto standard for, developing a career in biomedical research has been to undergo “research training” by obtaining a doctoral research degree (ie, a PhD) or a doctoral clinical degree (eg, an MD), and then obtain further postdoctoral research training in an advanced research area. This combination of doctoral plus postdoctoral training has generally been seen as critical to successfully developing a subsequent long-term career as an “independent investigator” with the necessary scientific/biomedical knowledge and skills in how to conduct rigorous biomedical research. Recognizing the importance of training medical scientists, in 1964, the NIH began the Medical Scientist Training Program that funds highly qualified new medical students to obtain dual degrees (MD/PhD). In addition, many NIH institutes and centers have supported Mentored Clinical Scientist Development Program Awards, which are institutional programs that provide research training infrastructure for those with clinical doctoral degrees. In recent years, the NIH has explored new avenues to reengineer clinical research training, which is part of the NIH Roadmap Initiative. All of these efforts for developing long-term careers in biomedical research have been based at research-intensive institutions such as conventional universities and health professional schools (eg, medical schools, nursing schools, etc), and these environments have provided a rich opportunity for conventional new biomedical researchers to obtain essential postdoctoral and early research career mentoring.

Challenges for CAM Practitioners

In contrast to the training of conventional medical doctors and biomedical scientists, most institutions that train CAM practitioners (eg, chiropractic schools, naturopathy colleges, etc) have never been directly affiliated with research-intensive universities. The mission of the CAM institutions has primarily been to train CAM clinicians to treat patients, not to conduct basic biomedical or clinical research. Before 1997, few faculty at chiropractic institutions had obtained significant clinical or basic science research funding; the institutions did not have substantive research infrastructure, and the faculty did not have the release time from their teaching and/or clinical responsibilities to pursue research funding. Arguably, chiropractic is one of the CAM professions that has most enthusiastically embraced the need for clinical and basic science research and for developing a research cadre, yet it finds it is still lacking in achieving its research goals. There are still no federally funded CAM dual-degree programs (eg, DC/PhD) comparable to the Medical Scientist Training Program. Furthermore, the CAM professional training has not necessarily been an obvious fit with the research training programs at conventional research intensive universities. Hence, there have been considerable barriers both programmatically and financially for interested CAM practitioners to pursue a research career. With that said, it remains essential to have highly trained scientists who understand CAM to conduct rigorous research in determining efficacy and mechanisms of CAM.

NATIONAL CENTER FOR COMPLEMENTARY AND ALTERNATIVE MEDICINE ADDRESSES THE CHALLENGES

NCCAM is committed to funding the research training and career development of CAM practitioners and scientists who themselves are committed to CAM research. The funding of research training and career development is not just part of NCCAM’s Strategic Plans, it is also an explicit part of the congressional language (Public Law 105-277) that created NCCAM as 1 of the 27 institutes and centers that comprise the NIH. From 2004 to 2006, NCCAM averaged over $9 million dollars annual expenditures for CAM research training and career development awards (Fig 1). There are a number of different funding opportunities that are focused on individuals at different stages of their research training (eg, predoctoral and postdoctoral) and research careers (faculty position at an accredited institution). It
<table>
<thead>
<tr>
<th>Mechanism</th>
<th>FOA</th>
<th>Stage of career</th>
<th>Basic eligibility a</th>
<th>Focus</th>
<th>% Effort</th>
<th>Max salary/stipend</th>
<th>Max research fund</th>
<th>Max years</th>
</tr>
</thead>
<tbody>
<tr>
<td>F31 (NCCAM)</td>
<td>PAR-07-384</td>
<td>Predoctoral</td>
<td>Bachelor’s degree; acceptance into a research doctoral program</td>
<td>Basic or clinical research</td>
<td>100</td>
<td>NRSAb</td>
<td>NRSAb</td>
<td>5</td>
</tr>
<tr>
<td>F31 (diversity)</td>
<td>PA-07-106</td>
<td>Predoctoral</td>
<td>Bachelor’s degree; acceptance into a research doctoral program</td>
<td>Basic or clinical research</td>
<td>100</td>
<td>NRSAb</td>
<td>NRSAb</td>
<td>5</td>
</tr>
<tr>
<td>F32 (NCCAM)</td>
<td>PA-07-319</td>
<td>Postdoctoral</td>
<td>Clinical or research doctoral degree</td>
<td>Basic or clinical research</td>
<td>100</td>
<td>NRSAb</td>
<td>NRSAb</td>
<td>3</td>
</tr>
<tr>
<td>K01</td>
<td>PA-06-001</td>
<td>Newly independent investigator or mid career</td>
<td>Clinical research doctoral degree and independent position</td>
<td>Basic or clinical research with a shift in research</td>
<td>Min 75</td>
<td>$75000</td>
<td>$25000</td>
<td>3-5</td>
</tr>
<tr>
<td>K01 f or CAM practitioners (NCCAM)</td>
<td>PAR-07-003</td>
<td>Newly independent investigator</td>
<td>CAM practitioners with doctoral degrees (see PAR for details)</td>
<td>Basic or clinical research</td>
<td>Min 75</td>
<td>$75000</td>
<td>$25000</td>
<td>3-5</td>
</tr>
<tr>
<td>K07 developmental</td>
<td>PA-00-070</td>
<td>Newly independent investigator</td>
<td>Clinical research doctoral degree and independent position</td>
<td>Curriculum development and clinical or basic research</td>
<td>Min 75</td>
<td>$75000</td>
<td>$25000</td>
<td>5</td>
</tr>
<tr>
<td>K07 leadership</td>
<td>PA-00-070</td>
<td>Senior or mid career investigator</td>
<td>Clinical research doctoral degree and independent mid career or senior position</td>
<td>Curriculum development and clinical or basic research</td>
<td>25 to 50</td>
<td>Institutional salary level to the NIH legislative limit</td>
<td>$25000</td>
<td>2-5</td>
</tr>
<tr>
<td>K08</td>
<td>PA-06-152</td>
<td>Newly independent investigator</td>
<td>Clinical doctorate and independent position</td>
<td>Laboratory or field research</td>
<td>Min 75</td>
<td>$75000</td>
<td>$25000</td>
<td>3-5</td>
</tr>
<tr>
<td>K22 (NCCAM)</td>
<td>PAR-05-129</td>
<td>Postdoctoral</td>
<td>NRSAb postdoctoral training; 1-5 y postdoctoral experience</td>
<td>Career transition; basic or clinical research</td>
<td>Postdoctoral phase: 100 independent phase: minimum 75</td>
<td>Postdoctoral phase: NRSAb; independent phase: $50000</td>
<td>Postdoctoral phase: 25000; independent phase: $150000 direct cost (includes salary but not F&amp;A costs)</td>
<td>Postdoctoral phase: 1; independent phase: 3</td>
</tr>
<tr>
<td>K23</td>
<td>PA-05-143</td>
<td>Newly independent investigator</td>
<td>Clinical doctorate</td>
<td>Patient-oriented research</td>
<td>Min 75</td>
<td>$75000</td>
<td>$25000</td>
<td>3-5</td>
</tr>
<tr>
<td>K24</td>
<td>PA-04-107</td>
<td>Mid career investigator</td>
<td>Clinical doctorate and mid career independent position</td>
<td>Patient-oriented research</td>
<td>25-50</td>
<td>Institutional salary level to the NIH legislative limit</td>
<td>$50000</td>
<td>3-5</td>
</tr>
<tr>
<td>K99/R00</td>
<td>PA-07-297</td>
<td>Postdoctoral</td>
<td>Max 5 y postdoctoral experience</td>
<td>Career transition; basic or clinical research</td>
<td>Postdoctoral phase: 75; independent phase: 75</td>
<td>Postdoctoral phase: $75000; independent phase: institutional salary level</td>
<td>Postdoctoral phase: 25000; independent phase $249000 total cost (includes salary and F&amp;A costs)</td>
<td>Postdoctoral phase: 1-2; independent phase: 3</td>
</tr>
</tbody>
</table>

All F and K awards require that the candidate be a citizen, permanent resident, or noncitizen national of the United States. The one exception is the K99/R00, which does not require U.S. citizenship or permanent resident status. Eligible clinical doctorates for F32 and K awards include MD, DDS, DMD, DO, DC, OD, ND, DVM, PharmD, or the PhD in disciplines such as nursing, speech language pathology, clinical genetics, audiology, and rehabilitation. Research doctorates include PhDs in nonclinical areas. For NCCAM K awards, the applicant must have an independent research position past the postdoctoral stage, with the exception of the K99/R00 and the K22.

PAR indicates Program Announcement with Special Receipt Dates; PA, Program Announcement; FOA, Funding Opportunity Announcement.

a See specific program announcement for complete eligibility requirements.
b National Research Service Award stipend and research allowance information can be found at: http://grants1.nih.gov/training/extramural.htm.
should be noted that NIH does not make grants to individuals but, rather, to institutions who submit a research grant application on behalf of a given investigator. The institution receiving the award has the obligation to provide the necessary and appropriate fiduciary and ethical oversight of all activities funded under the award.

All such grants are made on a competitive, peer-reviewed basis, and investigators are strongly encouraged to contact the relevant program officer at NCCAM for information about what makes an application competitive. Because NIH’s mission is focused on biomedical research, research training, and research career development, NIH does not fund students obtaining solely clinical degrees, whether conventional (eg, MD) or CAM (eg, DAOM, DC, DO, ND). All the funding opportunities described in this article require an individual to have some sort of position (student, postdoctoral, staff, or faculty) at an accredited academic or research institution (CAM or conventional). Information on all NCCAM funding opportunities is described on the NCCAM Web site, and NCCAM research training opportunities can be found on the Training Page of the NCCAM Web site. Also on the NCCAM Training Web site is an interim assessment of the initial 5 years of NCCAM’s research training and career development programs.

**Research Training for CAM Practitioners with Doctoral Degrees**

Complementary and alternative medicine practitioners with doctoral degrees (eg, DAOM, DC, DO, ND) from accredited institutions are eligible to apply for funding for “research training.” In this context, research training implies learning how to rigorously conduct science, whether conventional or CAM. The underlying assumption is that most professional clinical degree programs (ie, both CAM and conventional) train clinicians to treat patients rather than provide rigorous training in the conduct of research. Hence, although a CAM practitioner may have a doctoral degree, they still would need research training. The funding opportunities for research training range from graduate student fellowships, which are called “F31” awards to help fund graduate student research training for those enrolled in a research doctoral program (ie, a PhD program), to postdoctoral fellowships, which are called “F32” awards to help fund more advanced research training in a specific research area (Table 1). Because CAM doctoral degrees are considered by the NIH to be academically equivalent to conventional medical/scientific degrees (eg, MD, PhD), a CAM practitioner with a doctoral degree is eligible to apply for a postdoctoral fellowship (ie, an F32 award). If eligible for postdoctoral fellowships, then the advanced research training may or may not include a component that leads to a research doctoral degree (ie, PhD). The decision to pursue either predoctoral training (via an F31 award) or postdoctoral training (via an F32 award) will depend on many factors including prior academic and research background, stage of clinical career, and personal/family factors.

A new NIH transition program to facilitate segueing from a postdoctoral position to a tenured faculty position is now available through NCCAM. Complementary and alternative medicine practitioners with doctoral degrees in active postdoctoral research training who are no more than 5 years in their postdoctoral research appointment may apply for a “Pathway to Independence” award using the K99/R00 mechanism (Table 1). The primary, long-term goal of this program for NCCAM is to increase and maintain a strong cohort of new and talented NIH-supported CAM independent investigators. This program is targeted to advanced postdoctoral fellows who have sufficient research experience to make them good candidates for tenure-track research faculty positions in 1 to 2 years. It is designed to facilitate a timely transition from a mentored postdoctoral research position to a stable independent tenure-track research faculty position with independent NIH or other independent research support at an earlier stage than is currently the norm. Successful applicants will receive stipend support for up to 2 more years of postdoctoral research studies and a commitment for up to 3 years of salary support for a tenure-track faculty position at a research-intensive institution. Funds for conducting a research project are included in the award.

NCCAM has awarded a number of institutional training grants (T32 awards) to CAM and conventional research institutions, which are listed and described on the Training page of the NCCAM Web site. Depending upon the specific T32 award, these programs recruit CAM practitioners, graduate students, and/or postdoctoral fellows into their CAM research training programs; fund their studies; and provide some tuition reimbursement, fringe benefits, and a stipend.

Complementary and alternative medicine practitioners with doctoral degrees and who are engaged in certain federal-, state-, or privately funded research programs are eligible to apply for the Educational Loan Repayment Program. If awarded, this could substantially reduce or eliminate any prior educational debt incurred to obtain their CAM doctoral degree. This program is designed to help recruit and retain scientists who plan to develop a long-term research career but might otherwise be deterred by the amount of debt they incurred obtaining their doctoral degree.

**Research Training for CAM practitioners without Doctoral Degrees**

For CAM practitioners without doctoral degrees but with bachelor’s degrees, research training could be initiated by matriculating at an academic institution that offers graduate student research training in CAM, such as one of the NCCAM-funded institutional training grant programs (T32), graduate training programs with a CAM focus at other
conventional biomedical research universities (eg, George-town University, Washington, DC), or in a conventional biomedical research university. These CAM practitioners with bachelor’s degrees would also be eligible to apply for a predoctoral fellowships (F31 awards) (Table 1) when matriculated in an accredited biomedical research doctoral program (conventional or CAM). These F31 awards are made on a competitive basis and are considered highly prestigious for the awardee, their mentor, and the institution.

In addition, there are other ways in which graduate students may receive financial support. For example, in many biomedical research graduate science programs in the United States, admitted full-time doctoral students (ie, PhD students) are often financially supported by their academic department, faculty sponsor, or by federally funded or private foundation training grants. Students typically receive some sort of tuition scholarship (or waiver) and are paid a stipend to conduct mentored research. In addition, graduate students may be paid as teaching assistants. For CAM practitioners without bachelor’s degrees, there are a few US institutions that offer bachelor’s degrees with a major or minor in a CAM-related discipline.6 It would be essential to have at least a bachelor’s degree to begin pursuing a research career (in CAM or conventional science).

Research Career Development for CAM Independent Investigators

Complementary and alternative medicine practitioners with doctoral degrees and who have a faculty position at an accredited institution (CAM or conventional) are eligible to apply for research Career Development Awards (ie, the K-award series) (Table 1). To be eligible, the candidate need not have a tenure-track faculty position but any full-time position such as research assistant professor, instructor, clinical assistant professor, and others, as long as it is past the postdoctoral stage and the institution has a long-term commitment to the individual. In general, the K-award series provides salary support for the mentored principal investigator (PI) to enable their “release time” from clinic, administrative, or teaching responsibilities, but typically only provides minimal resources for the conduct of the research project itself. Thus, the bulk of the research support itself typically comes from resources provided by the PI’s mentor (who presumably would already have research support), small grants the PI might also have, and/or institutional funds. Whatever the source of the research support, it must be identified and committed prior to the submission of the research grant application. The specific amount of funding available for research support varies depending on the specific type of K-award (Table 1).

New Opportunity for CAM Practitioners Only

In October 2006, NCCAM released the funding opportunity announcement PAR-07-003—The Bernard Osher Foundation/NCCAM CAM Practitioner Research Career Development Award (K01).31 The specific purpose of this award is to provide research training support for CAM practitioners with clinical doctorates (eg, DAOM, DC, DO, ND), who have had limited opportunities for research training but a strong desire to pursue a career in CAM research. The long-term goal is to encourage more CAM practitioners to enter research careers, thus enriching CAM research through their experience and knowledge of CAM practice. This award will provide support and “protected” time (3, 4, or 5 years) for intensive supervised career development research experience in the biomedical, behavioral, or clinical sciences related to CAM. Awards are not renewable nor are they transferable from one PI to another. Full information about this Funding Opportunity Announcement is available via the grants.gov Web site32 and is also on the NCCAM Training page of the NCCAM Web site.28 This award is designed exclusively for CAM practitioners with doctoral degrees (from accredited institutions) and specifically excludes those with conventional medical and PhD degrees (ie, MD, PhD). Like all NCCAM-funded K01 awards for newly independent investigators, it provides a contribution to salary support up to $75000 annually plus fringe benefits and a small research development fund to enable release time from clinical, administrative, and/or teaching duties at the awardee’s institution to pursue a mentored research career development program. To be competitive, the institution submitting the application on behalf of the investigator must have a well-established record of funded research and research career development activities and qualified research faculty to serve as mentors. The institution must also have a long term commitment to the applicant in providing him/her with some type of position, mentoring, space, and research resources. Collaborative arrangements between CAM and conventional research institutions are possible to facilitate the most appropriate mentoring for a given candidate, and as always, investigators are strongly encouraged to discuss possibilities and options with the appropriate NCCAM Program Officer. To encourage research mentors and institutions to consider providing CAM practitioners with positions and resources for research training, The Bernard Osher Foundation will provide a separate annual salary award of $40000 to the primary mentor of successful applicants. Finally, in some cases, the mentored research may not be directly in CAM per se, as the CAM practitioner presumably already has a strong CAM background. Rather, the CAM practitioner may become mentored/trained in conventional biomedical research that would augment and/or be integrated with their CAM knowledge. Investigators are strongly encouraged to discuss their proposals with the appropriate NCCAM Program Officer before developing and submitting their applications. This Osher/NCCAM K01 award is also distinct from others in the K-award series in that it does not require significant prior research training to be eligible to apply for
it. However, the more prior research training performed would likely make a given application more competitive for actual funding.

**Career Development Award K-series**

The Mentored Research Scientist Development Award is designed for investigators pursuing a shift in focus of their research. It uses the K01 funding mechanism, and its purpose for NCCAM is to provide support and “protected time” (3, 4, or 5 years) for an intensive, supervised CAM career development experience in the biomedical, behavioral, or clinical sciences (Table 1). The proposed CAM career development experience must be in a research area new to the applicant and/or one in which an additional supervised research experience will substantially augment the CAM research capabilities of the investigator. NCCAM uses this award both for newly independent investigators and for midcareer investigators who need some protected time to gain the necessary experience to shift their research career into new areas.

The Mentored Clinical Scientist Research Career Development Award uses the K08 funding mechanism (Table 1) and provides support and “protected time” to individuals with a clinical doctoral degree (either CAM or conventional) for an intensive, supervised research career development experience in the fields of biomedical and behavioral research, including translational research, but not in patient-oriented research per se. Individuals with a clinical doctoral degree interested in pursuing a career in patient-oriented research (POR) should refer to the Mentored Patient-Oriented Research Career Development Award (K23).

The Mentored Patient-Oriented Research Career Development Award uses the K23 funding mechanism, and its purpose for NCCAM is to support the career development of investigators who have made a commitment to focus their research endeavors on CAM POR (Table 1). This mechanism provides support for 3 to 5 years of supervised study and research for clinically trained professionals who have the potential to develop into productive, clinical investigators focusing on CAM patient-oriented research. Applicants must justify the need for a period of mentored research experience and provide a convincing case that the proposed period of support and career development plan will substantially enhance their careers as independent investigators in CAM patient-oriented research. Clinically trained professionals or individuals with a clinical degree who are interested in further career development in biomedical research that is not patient-oriented should refer to the Mentored Clinical Scientist Career Development Award (K08).

The Midcareer Investigator Award in Patient-Oriented Research uses the K24 funding mechanism, and its purpose for NCCAM is to provide support for clinician investigators to allow them protected time to devote to CAM POR and to act as CAM research mentors primarily for clinical residents, clinical fellows, and/or junior clinical faculty (Table 1). This award is primarily intended for clinician investigators who are at a minimum at the associate professor level or are functioning at that rank in an academic setting or equivalent nonacademic setting, and who have an established record of independent, peer-reviewed Federal or private research grant funding in CAM POR. This award is intended to advance both the research and the mentoring endeavors of outstanding CAM patient-oriented investigators. It is expected, for example, that investigators will obtain new or additional independent peer-reviewed funding as the PI for CAM POR and establish and assume leadership roles in collaborative CAM POR programs, and that there will be an increased effort and commitment to mentor beginning clinician investigators in CAM POR to enhance the research productivity of the investigator and increase the pool of well-trained CAM clinical researchers of the future.

The Developmental Academic Career Award (using the K07 funding mechanism) is used by NCCAM to support individuals interested in introducing or improving curricula in a particular CAM scientific field as a means of enhancing the educational or research capacity at the grantee institution (Table 1). The K07 provides up to 5 years of support for more junior candidates who are interested in developing academic and research expertise in a particular field as a way to increase the overall pool of individuals capable of research or teaching in the identified CAM area. During the period of the award, the candidate will become a successful CAM academician in the chosen area. Teaching, curriculum building, research, and leadership skills are to be learned during the tenure of the award. For junior candidates, a mentor is required.

The Leadership Academic Career Award uses the K07 funding mechanism and is used by NCCAM to support individuals interested in introducing or improving curricula in a CAM scientific field as a means of enhancing the CAM educational or research capacity at the grantee institution (Table 1). The award provides from 2 to 5 years of support for more senior individuals with acknowledged CAM scientific expertise and leadership skills, who are interested in improving the CAM curricula and enhancing the CAM research capacity within an academic institution. It is expected that support under this award will increase the visibility and the overall research support or academic capacity for the given field of CAM research within the academic medical/health and research community.

**Institutional Grants**

To ensure that a diverse and highly trained workforce is available to assume leadership roles related to the Nation’s CAM research agenda, NCCAM awards the Ruth L. Kirschstein National Research Service Award (NRSA) Institutional Research Training Grants to eligible institutions using the T32 funding mechanism. For NCCAM, this program supports predoctoral, postdoctoral, and short-term
CAM research training programs at domestic institutions of higher education, including CAM institutions. Awards for T32 institutional NRSA research training grants may be for periods up to 5 years in duration and are renewable. Trainees are required to pursue full-time CAM research training. Only domestic, nonprofit, private or public institutions that have strong and high-quality research programs in the CAM area (s) proposed for research training and have requisite research staff and facilities may apply to NCCAM for grants to support NRSA Institutional research training programs. A history of producing successful research scientists and a faculty with research training experience and substantial research funding are also important characteristics of institutions that are competitive for these awards.

Accredited CAM institutions that have not been major recipients of NIH support are eligible to apply for the Academic Research Enhancement Award (AREA) program using the R15 funding mechanism. The general purpose of the AREA program is to stimulate research at educational institutions that provide baccalaureate or advanced degrees for a significant number of the Nation’s research scientists, but that have not been major recipients of NIH support. Eligible organizations include all public or private institutions and components of institutions such as health professional schools/colleges and other academic components of domestic institutions offering baccalaureate or advanced degrees in the sciences related to health, except those that have received research grants and/or cooperative agreements from the NIH totaling more than $3 million per year (in combined direct and indirect costs) in each of 4 or more of the last 7 years. The purpose of this program for NCCAM is to stimulate CAM research, especially at eligible CAM educational institutions and to create research opportunities for scientists at such institutions that otherwise are unlikely to participate extensively in NIH programs. The AREA grants from NCCAM are intended to support small-scale, CAM-related research projects proposed by faculty members of eligible, domestic institutions. In addition, providing research training for students at these institutions is an important element of AREA awards.

**Research Training and Career Development in CAM for Conventional Scientists**

Most applications that are received by NCCAM for CAM research training awards and for CAM research career development awards are from scientists and physicians with conventional training (eg, PhD and/or MD). Given the substantially larger pool of these scientists, this should not be a surprise. NCCAM welcomes applications from both the CAM and conventional scientific fields to seek funding for promising CAM research training and career development. All of the research training awards (ie, F-series) and the career development awards (K-series, with the exception of the Osher/NCCAM K01) are also available to those scientists with conventional research training who now wish to focus on CAM (Table 1).

Current conventional postdoctoral scientists already supported by F32, T32, or other prestigious postdoctoral awards from whatever sources (Federal, State, or private foundations) and who are proposing a career in CAM research may apply for the Complementary and Alternative Medicine Career Transition Award, which uses the K22 mechanism (Table 1). The goal of this funding opportunity is to provide support for outstanding advanced postdoctoral research scientists during their transition to independence in CAM. The award will provide support for up to 1 year of postdoctoral research training and 3 years of research support as an independent investigator.

**CONCLUSION**

NCCAM has developed an extensive portfolio of CAM training and career development awards that are designed for CAM clinicians, scientists, and institutions. In combination with the NIH Educational Loan Repayment Program, it is now financially and programmatically feasible for clinicians and scientists from the CAM community to pursue a career in CAM research.

**Practical Applications**

- Complementary and alternative medicine research training and career development awards are available from NCCAM.
- Research Career Development Awards for faculty provide salary support.
- Postdoctoral research fellowships provide stipend support.
- Predoctoral research fellowships require matriculation in an accredited research doctoral program.
- Osher/NCCAM CAM Practitioner Award is exclusively for CAM practitioners.

**ACKNOWLEDGMENT**

The authors thank Mr Yancy Bodenstein in NCCAM’s Office of Policy, Planning, and Evaluation for his assistance in obtaining data on the training and career development awards.

**REFERENCES**

27. NCCAM website. [National Center for Complementary and Alternative Medicine, National Institutes of Health]. December 2006; Available at: http://nccam.nih.gov/ [Accessed January 12, 2007].
Acute traumatic wrist injuries

Albastaki et al (p. 522) describe different triangular fibrocartilage complex lesions including the degenerative and the posttraumatic changes seen on magnetic resonance imaging.

Clinical prediction rules

May and Rosedale (p. 539) present a patient who presented with 4 out of the 5 criteria of a clinical prediction rule for responders to manipulation, but instead responded to flexion exercises without manipulation. This case shows that clinical prediction rules do not necessarily describe the only management for some patients.

Manual therapy’s influence on respiratory function

Engel and Vemulpad (p. 509) measured the effect of the combination of chiropractic manual therapy with exercise on normal respiratory function. Manual therapy administered before exercise may create additional tolerance within the respiratory system, permitting more exercise to be undertaken.

Subdural hematoma after steroid injection

Ozdemir et al (p. 536) present a case of subacute intracranial subdural hematoma that developed after epidural steroid injection. Although intracranial subdural hematoma after accidental dural puncture during epidural steroid injection is a rare complication, persistent headache should be evaluated carefully for possible intracranial hematomas for patients who have undergone such a procedure.

Palpatory grading of muscle activity

Sedaghat et al (p. 501) investigated the reproducibility of a palpatory grading system. Although a palpatory method of measuring motor control is clinically practical and readily accessible, this study demonstrated that one particular method is not reproducible. Possible sources of error are identified and evaluated, and modifications are suggested for further research.

Insights into identifying multiple myeloma

Kaufman (p. 543) presents a case study of a patient diagnosed with early manifestations of multiple myeloma. This case illustrates relevant aspects of differential diagnosis; use of laboratory, radiological, and advanced imaging techniques to aid in establishing the diagnosis; and issues of case management.

Opportunities for research funding

Khalsa and Pearson (p. 483) provide information about current research training funding opportunities from the National Center for Complementary and Alternative Medicine (CAM) and the National Institutes of Health as a whole that are available to CAM practitioners in the context of the historical challenges of transitioning from a clinical career in CAM practice to a CAM research career.

How reliable is the leg length check?

Schneider et al (p. 514) explore the interexaminer reliability of the prone leg length analysis as performed according to the Activator Methods and Derifield protocols. The results showed varying levels of reliability for different procedures.

Faculty promotion criteria

Wiese et al (p. 527) describe the tools and processes developed by a Professor Promotion Committee to infuse greater objectivity and confidence into its faculty promotion recommendations.
ORIGINAL ARTICLES
2007 ACC-RAC CONFERENCE AWARD WINNING PAPER

COMPARISON OF POSTEROANTERIOR SPINAL STIFFNESS MEASURES TO CLINICAL AND DEMOGRAPHIC FINDINGS AT BASELINE IN PATIENTS ENROLLED IN A CLINICAL STUDY OF SPINAL MANIPULATION FOR LOW BACK PAIN

Edward F. Owens Jr, MS, DC,a James W. DeVocht, DC, PhD,a M. Ram Gudavalli, PhD,a David G. Wilder, PhD,b and William C. Meeker, DC, MPHc

ABSTRACT

Objective: A system for measuring posterior-to-anterior spinal stiffness (PAS) was developed for use in clinical trials of manipulation for low back pain (LBP). The current report is an analysis of the baseline PAS data, with particular emphasis on relationships between PAS and clinical and demographic characteristics.

Methods: Posterior-to-anterior spinal stiffness measurements were recorded over the spinous processes of the lumbar spines from patients who had LBP. The system uses electronic sensors to record displacement and force, whereas a human operator provides the force of indentation. Clinical and outcome measures were compared with spinal stiffness.

Results: We recruited 192 patients (89 female and 103 male; average age, 40.0 years; SD, 9.4 years). The average Roland-Morris score was 9.7 (SD, 3.2) on a 24-point scale. The Visual Analog Scale pain scores were 55.7 (SD, 20.9) on a 100-mm scale. Stiffness values ranged from 4.16 to 39.68 N/mm (mean, 10.80 N/mm; SD, 3.72 N/mm). Females’ lumbar spines were, on the average, 2 N/mm more compliant than males ($P < .001$).

Conclusions: The PAS system of computer-monitored equipment with human operation performed well in this clinical study of LBP. Spinal stiffness was found to be different between males and females, and age and body mass index were related to PAS. We found no significant relationship between the severity or chronicity of the LBP complaint and spinal stiffness. There was little agreement between the stiff or tender segments identified by the clinicians using palpation and the segment that measured most stiff using the PAS device. (J Manipulative Physiol Ther 2007;30:493-500)

Key Indexing Terms: Chiropractic; Lumbar Spine; Biomechanics; Low Back Pain; Stiffness; Tests and Measurement; Manipulation, Spinal

Spinal stiffness assessments are frequently performed as part of patient evaluation in chiropractic, osteopathy, and physical therapy. Often referred to as posterior-to-anterior spinal stiffness (PAS), in chiropractic, the method might be called motion palpation or joint end-play assessment. The clinician typically will use the palm of his or her hand to press on the spine and feel for restricted movement.

Studies of spinal stiffness or motion palpation in chiropractic and physical therapy most often find limited interexaminer reliability of the assessment.1,2 Because these are commonly used clinical tests, researchers have attempted to develop instruments to help improve the objectivity of the measures. Over the past 20 years, several devices have been reported in the literature.3-11 Such devices typically use a computer-controlled motor to indent the spine or adjacent tissues while monitoring the amount of displacement of the skin surface and the amount of force developed. These devices report very good reliability and accuracy.6

Posterior-to-anterior spinal stiffness measurement is a more complex task than it might seem at first. Investigators have

---

a Associate Professor of Research, Palmer Center for Chiropractic Research, Davenport, Iowa.
b Associate Professor, Departments of Biomedical Engineering and Orthopaedics, Iowa Spine Research Center, University of Iowa.
c President, Palmer College of Chiropractic—West.

Submit request for reprints to: Edward F. Owens Jr, MS, DC, Associate Professor of Research, Palmer Center for Chiropractic Research, 741 Brady Street, Davenport, IA 52803, USA
(e-mail: edward.owens@palmer.edu).

0161-4754/$32.00
Copyright © 2007 by National University of Health Sciences.
doi:10.1016/j.jmpt.2007.07.009

493
identified a host of variables that can influence the measurement. Relative to the patient, variables include patient positioning and the support surface,\textsuperscript{12,13} respiration and abdominal pressure,\textsuperscript{14,15} paraspinal muscle activity,\textsuperscript{16,17} and the thickness of soft tissue between the indenter and the underlying bones.\textsuperscript{5,10,18} Lee et al\textsuperscript{18} suggested that PAS measures may have more to do with the supporting structures than the intrinsic properties of the spine. Relative to the indentation process itself, the variables include the location and direction of load application,\textsuperscript{8,10,19-22} the rate of loading,\textsuperscript{23,24} the magnitude of the peak load used,\textsuperscript{25} and the size of the indenter.\textsuperscript{24} Any system of PAS measurement must control or account for as many of these variables as possible to provide valid measurements. A model has been developed, which includes elements of soft tissue compression at the points of contact between the patient and the supporting surface and under the indenter, as well as skeletal movements.\textsuperscript{18,26}

Interpretation of PAS measurements is also complex. Although clinicians conceive of the test as being indicative of intersegmental motion, particularly at the level of the segment being directly contacted during the test, studies have shown that the movements associated with PAS testing are distributed throughout the lumbar vertebral column and pelvis. When pressing on the midlumbar spine, the whole lumbar spine extends, and there is also rotation of the ribcage and pelvis.\textsuperscript{27-30} Imaging studies with radiography\textsuperscript{30} and magnetic resonance imaging (MRI)\textsuperscript{31,32} have shown that all of the lumbar vertebrae are involved in motion during PAS testing. The movements are generally rotational in nature, producing relative extension of the segments, although flexion can be seen at the lowest lumbar levels when loading is directed to the upper lumbar segments. Powers et al\textsuperscript{32} and Kulig et al\textsuperscript{31} applied posterior to anterior loads to each of the lumbar segments and found that the greatest amount of intersegmental rotations did occur at the segment contacted. These investigators used manually applied loads estimated at between 20- and 25-lb force (9-11 kg).

Posterior-to-anterior spinal stiffness testing shows some promise as either a diagnostic tool or an outcome measure in low back pain (LBP) studies. Neither area has been well investigated. Lumbar PAS over the spinous process has been assessed in nonspecific patients with LBP and found to have no substantial difference with the measure in pain-free controls.\textsuperscript{9,33} These were small studies, however, and did not use matched pairs of patients with LBP and pain-free controls. Using a different type of PAS measuring system, with the PA load contacting the paraspinal muscles rather than the spinous process, Brodeur and DelRe\textsuperscript{3} found a difference in lumbar spine stiffness between male patients with LBP and pain-free male controls. In an animal model of lumbar spine degeneration, Kawchuk et al\textsuperscript{34} have found that PAS measurement is sensitive to disk lesions. Kawchuk’s device is somewhat different from other testing systems in that it uses an ultrasound sensor at the point of contact with the loading probe to detect changes in soft tissue thickness during the test.\textsuperscript{5,35} Hence, the method is better able to detect skeletal movement independent of soft tissue indentation. In a cadaver model, PAS measurements in the thoracic spine correlated with direct measurement of intersegmental flexibility in flexion/extension movements.\textsuperscript{36} Thus, although PAS does seem to provide information about deep lumbar structure, now, the sensitivity and specificity of the measure for detecting spinal joint dysfunction related to LBP in humans are unknown.

Posterior-to-anterior spinal stiffness testing has not been widely used as an outcome measure in clinical trials. In a longitudinal study, spinal stiffness was found to decrease by 8% (1.2 N/mm) when patients were no longer in pain.\textsuperscript{23} However, studies in asymptomatic participants have not shown any changes in PAS in the short term in response to mobilization\textsuperscript{37,38} and the only study to measure PAS in patients with LBP also failed to show any change with a single session of mobilization.\textsuperscript{39} There is a need for studies comparing the results of a course of therapy to changes in biomechanical parameters, such as PAS.

We developed a hand-held device to take measurements of spinal stiffness of patients with nonspecific LBP in a clinical study of the relationship between those patients’ baseline biomechanical and demographic characteristics and their response to spinal manipulation. In earlier studies, we found that the instrument exhibits substantial intraexaminer reliability, with an intraclass correlation coefficient of 0.790.\textsuperscript{40} Repeated testing showed a standard error of measurement of 1.62 N/mm. The current report is an analysis of the PAS data from that clinical study, with particular emphasis on relationships between PAS and clinical and demographic characteristics.

**METHODS**

The project was approved by the college’s institutional review board. Patients were recruited from the local population using print ads, radio announcements, and direct mail. On their initial visit, prospective patients signed an informed consent form and completed a packet of enrollment forms and baseline outcome surveys including the Roland-Morris Questionnaire (RMQ), Medical Outcomes Trust 36-Item Short-Form Survey, Visual Analog Scale (VAS) for pain, Pain Disability Index, and Beck Depression Inventory. Patients were screened for inclusion and exclusion criteria (Fig 1) by a study coordinator and examining doctor of chiropractic. The 2 inclusion criteria were that the LBP was of at least 4 weeks’ duration and the severity was at least 6 on the 24-point RMQ. The physical examination was performed over the course of the study by 3 chiropractors with a minimum of 10 years of experience. The examination included range of motion (ROM) tests, orthopedic tests, and palpatory tenderness and stiffness tests. Range of motion was performed with an electronic dual inclinometer system (Jtech Medical, Salt Lake City, Utah). The clinician located the lumbar spinous processes by palpation with respect to the...
equipment are described fully in a previous report. Briefly, postural sway test, ROM, and PAS. This report describes the on their second visit to the research clinic, which included a 10 indicates the most severe pain).

Tenderness on palpation was scored on a 0 to 10 scale by asking the patient to rate their pain when the clinician pressed on each joint as stiff or not stiff. The clinician pressed on each of the 5 spinous processes and crest of the ilia and the posterior inferior iliac spines (PSIS). The clinician pressed on each of the 5 spinous processes and rated each joint as stiff or not stiff. The clinician pressed on the PSIS to judge the sacroiliac joint stiffness on each side. Tenderness on palpation was scored on a 0 to 10 scale by asking the patient to rate their pain when the clinician pressed on each spinous process or PSIS (0 indicates no pain, 10 indicates the most severe pain).

Enrolled patients had a baseline biomechanics assessment on their second visit to the research clinic, which included a postural sway test, ROM, and PAS. This report describes the PAS results only. The PAS measurement methods and equipment are described fully in a previous report. Briefly, the PAS device consisted of a plastic rod with an inline force transducer (Omegadyne, Inc, LC201-50, Sunbury, Ohio) mounted at the lower end and a position tracking sensor mounted at the upper end. A Polhemus Liberty motion tracking system (Virtalis Group, Manchester, UK) monitored the location of the rod, while a Motion Monitor (Innovative Sports Training, Chicago, Ill) software recorded data on both rod location and force. In use, the device was pressed manually into the soft tissue overlying the lumbar spinous processes of patients, while force and displacement data were recorded. The computer provided audible feedback to the operator when the target force was reached. The lumbar spinous processes were located by palpation and marked with a skin marker. We tested each spinal level from L1 through L5 with 5 cycles of 80 N of compression. If the patient was unable to tolerate 80 N, the target force was decreased to 5 N below his/her pain tolerance; no testing was performed when a patient could not tolerate more than 50 N. The 5-cycle test on one lumbar segment was completed within 5 seconds. Patients were prone on a hard wooden table for the testing and were asked to exhale fully and suspend their breathing during the test.

Data Reduction

Biomechanics technicians who performed the testing and data reduction were blinded to the demographic and clinical data. During data reduction, linear regression was used to calculate the slope of the force-displacement relationship resulting in a stiffness value in newtons per millimeter (N/mm). The regression was only performed on the data in the 20-N range 5 N below the maximum testing load used (eg, for a test load of 80 N, linear regression was performed on the segment from 55 to 75 N). During data reduction, we rejected the stiffness value on a cycle-by-cycle basis if the regression coefficient ($R^2$) was less than 0.90. The data for the first depression were ignored, and an average stiffness value for the test was calculated from cycles 2 to 5. Posterior-to-anterior spinal stiffness testing also resulted in variables related to the performance of the test, including the patient’s most tender lumbar segment, the threshold force that produced pain at that level (if less than 80 N), and the maximum load that was used in the test.

Statistical Analysis

A complete data set was compiled from outcomes surveys, demographic, clinical, and biomechanical data sets, and parsed to SPSS for Windows (version 12; SPSS, Inc, Chicago, Ill). Descriptive statistics were tabulated on the PAS variables, particularly the spinal stiffness at each segment and the threshold sensitivity. Bivariate correlations were explored between the segmental stiffness values and a selection of patient characteristics, including age, sex, and body mass index (BMI), and clinical findings including, RMQ score, VAS, chronicity of complaint, and lumbar ranges of motion in flexion and extension.

The palpation data on stiffness and tenderness were evaluated in an attempt to identify which lumbar segment was found to be most stiff or most tender. Likewise, the PAS data were used to identify which segment had the greatest measured stiffness.

RESULTS

We recruited 192 patients (89 female and 103 male; average age, 40.0 years; SD, 9.4 years) for the study. The

---

**Fig 1. Inclusion and exclusion criteria for the clinical study**

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Age, 21 y or older.</td>
<td>1. Low back pain not meeting Quebec Task Force Diagnostic Classifications 1, 2, or 3; especially LBP associated with frank radiculopathy, altered lower extremity reflex, dermatomal sensory deficit, progressive unilateral muscle weakness or motor loss, symptoms of cauda equina compression, and computed tomography or MRI evidence of anatomical pathology (eg, abnormal disk, lateral or central stenosis).</td>
</tr>
<tr>
<td>2. Idiopathic LBP meeting Quebec Task Force Diagnostic Classifications 1, 2, or 3.</td>
<td>2. Comorbid conditions or general poor health that could significantly complicate the prognosis of LBP, including pregnancy, bleeding disorders, extreme obesity, and clear evidence of narcotic or other drug abuse.</td>
</tr>
<tr>
<td>3. Subacute (onset 4-12 weeks previous) or chronic (onset more than 12 weeks previous) LBP.</td>
<td>3. Major clinical depression defined as scores greater than 29 on the Beck Depression Inventory—Second Edition.</td>
</tr>
<tr>
<td>4. Minimum baseline score of 6 on the 24-item Roland-Morris Disability Questionnaire.</td>
<td>4. Pacemaker (safety issues with equipment in the testing laboratory).</td>
</tr>
<tr>
<td>5. Written informed consent.</td>
<td>5. Use of spinal manipulative care for any reason within the past month.</td>
</tr>
<tr>
<td></td>
<td>6. Unwillingness to postpone use of manual therapies for LBP except receiving disability for a health-related condition.</td>
</tr>
<tr>
<td></td>
<td>7. Use of spinal manipulative care for any reason within the past month.</td>
</tr>
<tr>
<td></td>
<td>8. Inability to read or verbally comprehend English.</td>
</tr>
<tr>
<td></td>
<td>9. Inability to read or verbally comprehend English.</td>
</tr>
</tbody>
</table>
average Roland-Morris score was 9.7 (SD, 3.2) on a 24-point scale. Initial VAS for pain scores were 55.7 (SD, 20.9) on a 100-mm scale, where 100 is the worst pain the patient could imagine. Table 1 shows these baseline descriptive statistics, along with weight, BMI, chronicity of complaint, and lumbar ROM. Most of the variables were normally distributed and could be well represented by mean values and SDs. Chronicity, however, was significantly skewed with nearly half of the patients having had their complaint from between 4 weeks and 4 years. Chronicity is represented by a frequency distribution in Table 1.

Not all patients could tolerate the PAS test because of spinous process sensitivity. Posterior-to-anterior stiffness measures were recorded for 173 patients; however, 11 patients had incomplete tests, where one or more segments were not recorded because of extreme sensitivity or equipment malfunction. Seventy-four percent of patients could withstand 80 N of load without pain. The average time for indentation and retraction of the probe was not the same; indentation averaged 1.16 seconds (SD, 0.39 seconds) and retraction 0.50 seconds (SD, 0.20 seconds). The average rate of loading, although varying across patients, was consistent across all segments, averaging 0.65 cycles/s (SD, 0.18 cycles/s).

Stiffness values ranged from 4.16 to 39.68 N/mm (mean, 10.80 N/mm; SD, 3.72 N/mm). There was a significant difference in spinal stiffness values between females and males (Table 2). Females’ lumbar spines were, on average, 2 N/mm more compliant than males’ ($P < .001$).

There was no difference in PAS with respect to lumbar levels when considering the group as a whole (Table 2). Many patients did, however, exhibit differences of more than 2 N/mm between segments. The mean difference between patients’ maximum and minimum recorded PAS was 5.00 N/mm (SD, 3.97 N/mm). The fifth lumbar had the highest stiffness in 25% of cases, with the other segments each comprising between 11% and 17% (Table 3). Only 17% of patients exhibited less than 2 N/mm difference across the lumbar spine.

Palpatory stiffness and tenderness measures were collected on all 192 patients in the study. Most patients exhibited stiffness and tenderness at multiple segments (Table 4). Although only 11 patients (5.7%) had no segment that the clinician judged as stiff, an almost equal number exhibited stiffness at all 5 lumbar segments. The most frequent finding was stiffness at 2 segments (69, 35.9%). Tenderness to palpation was also found in most patients, except many more had tenderness at several (or all) segments.

Because clinicians only used a 2-level scale to assess spinal joint stiffness (yes or no), it is not possible to derive a correlation to the PAS measure. Instead, we attempted to see if the clinicians were identifying as stiff the same segments as the PAS system measured. First, we examined the clinical data to find the particular segment that showed increased stiffness or tenderness by palpation. In those few cases where

---

**Table 1.** Descriptive statistics on selected baseline measures ($n = 192$)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)</td>
<td>40 (9.41)</td>
</tr>
<tr>
<td>Weight (lb)</td>
<td>195.85 (43.71)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>89.02 (19.87)</td>
</tr>
<tr>
<td>BMI</td>
<td>29.84 (6.69)</td>
</tr>
<tr>
<td>VAS (0-100)</td>
<td>55.65 (20.90)</td>
</tr>
<tr>
<td>Roland-Morris score (0-24)</td>
<td>9.68 (3.20)</td>
</tr>
<tr>
<td>LS active flexion (°)</td>
<td>56.18 (30.72)</td>
</tr>
<tr>
<td>LS active extension (°)</td>
<td>18.28 (8.70)</td>
</tr>
<tr>
<td>Chronicity * (y)</td>
<td>No. of patients</td>
</tr>
<tr>
<td>&lt;1</td>
<td>39</td>
</tr>
<tr>
<td>1-4</td>
<td>50</td>
</tr>
<tr>
<td>5-8</td>
<td>26</td>
</tr>
<tr>
<td>9-12</td>
<td>24</td>
</tr>
<tr>
<td>13-16</td>
<td>26</td>
</tr>
<tr>
<td>&gt;16</td>
<td>27</td>
</tr>
</tbody>
</table>

* Presented as a frequency distribution due to skewness.

**Table 2.** Posterior-to-anterior stiffness measures by lumbar segment and sex

<table>
<thead>
<tr>
<th>Segment</th>
<th>Mean (SD) of PA stiffness (N/mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All patients (n = 162) Female (n = 64) Male (n = 98)</td>
</tr>
<tr>
<td>L1</td>
<td>10.25 (2.98) 9.26 (2.82) 10.97 (2.91)*</td>
</tr>
<tr>
<td>L2</td>
<td>10.45 (3.72) 9.29 (3.62) 11.30 (3.58)*</td>
</tr>
<tr>
<td>L3</td>
<td>10.82 (3.85) 9.38 (2.99) 11.83 (4.07)*</td>
</tr>
<tr>
<td>L4</td>
<td>10.81 (4.24) 9.86 (4.97) 11.48 (3.51)*</td>
</tr>
<tr>
<td>L5</td>
<td>11.12 (3.73) 9.55 (2.74) 12.14 (3.93)*</td>
</tr>
</tbody>
</table>

* Significant differences between sex ($P < .001$).

**Table 3.** The frequency with which stiffness occurred at each lumbar segment as measured with PAS

<table>
<thead>
<tr>
<th>Lumbar segment</th>
<th>% of Patients (n = 173)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;2 N/mm difference</td>
<td>16.8</td>
</tr>
<tr>
<td>L1</td>
<td>15.6</td>
</tr>
<tr>
<td>L2</td>
<td>13.9</td>
</tr>
<tr>
<td>L3</td>
<td>17.3</td>
</tr>
<tr>
<td>L4</td>
<td>11.6</td>
</tr>
<tr>
<td>L5</td>
<td>24.9</td>
</tr>
</tbody>
</table>

The lumbar segment with the highest PAS measure was ranked as stiffest when there was more than 2 N/mm difference between the patient’s most and least stiff segment.

**Table 4.** The number and frequencies of segments that were found either stiff or tender on palpation

<table>
<thead>
<tr>
<th>No. of segments</th>
<th>Palpably stiff, n (%)</th>
<th>Tender on palpation, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>11 (5.7)</td>
<td>5 (2.6)</td>
</tr>
<tr>
<td>1</td>
<td>45 (23.4)</td>
<td>7 (3.6)</td>
</tr>
<tr>
<td>2</td>
<td>69 (35.9)</td>
<td>18 (9.4)</td>
</tr>
<tr>
<td>3</td>
<td>44 (22.9)</td>
<td>16 (8.3)</td>
</tr>
<tr>
<td>4</td>
<td>13 (6.8)</td>
<td>23 (12.0)</td>
</tr>
<tr>
<td>5</td>
<td>10 (5.2)</td>
<td>123 (64.1)</td>
</tr>
<tr>
<td>Total</td>
<td>192</td>
<td>192</td>
</tr>
</tbody>
</table>

The lumbar segment with the highest PAS measure was ranked as stiffest when there was more than 2 N/mm difference between the patient’s most and least stiff segment.
only one segment was noted on palpation, the choice was clear. An attempt was made to reduce the palpation findings to only one segment in those cases where positive findings were found in only 2 or 3 segments and when those segments were contiguous (ie, L3 and L4 or L2-L4 were all found tender). When 2 contiguous segments were found, the most distal was selected as the stiffest segment, and when 3 segments were found, the middle of the 3 was designated. With this method, a single segment was designated as palpably most stiff in 112 patients and as palpably most tender in 166 patients.

The amount of agreement between the PAS and palpation findings was found by subtracting the segment number found by each method for each patient. Hence, we calculated the distance (in segments) between the 3 determinations (Table 5). For example, if a patient was found to be most tender at L1, most stiff by palpation at L5, and had greatest PAS at L4, the PAS vs palpatory stiffness would be 1 (4 − 5), the PAS vs palpable tenderness would be 3 (4 − 1), and the palpable tenderness vs Palpable Stiffness would be 4 (5 − 1).

Table 5. Agreement on designated segment by PAS instrument measurement and manual palpation for stiffness and tenderness

<table>
<thead>
<tr>
<th>Distance (no. of segments away)</th>
<th>PAS vs Palpatory stiffness</th>
<th>PAS vs Palpatory tenderness</th>
<th>Palpatory tenderness vs palpatory stiffness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>82</td>
<td>100.0%</td>
<td>121</td>
</tr>
<tr>
<td>−4.00</td>
<td>8</td>
<td>9.8%</td>
<td>7</td>
</tr>
<tr>
<td>−3.00</td>
<td>19</td>
<td>23.2%</td>
<td>17</td>
</tr>
<tr>
<td>−2.00</td>
<td>16</td>
<td>19.5%</td>
<td>22</td>
</tr>
<tr>
<td>−1.00</td>
<td>10</td>
<td>12.2%</td>
<td>15</td>
</tr>
<tr>
<td>.00</td>
<td>18</td>
<td>22.0%</td>
<td>26</td>
</tr>
<tr>
<td>1.00</td>
<td>10</td>
<td>12.2%</td>
<td>16</td>
</tr>
<tr>
<td>2.00</td>
<td>1</td>
<td>1.2%</td>
<td>9</td>
</tr>
<tr>
<td>3.00</td>
<td>0</td>
<td>0%</td>
<td>8</td>
</tr>
<tr>
<td>4.00</td>
<td>0</td>
<td>0%</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 6. Pearson’s correlation coefficient and level of significance for selected demographic and clinical factors vs PAS value by segment

<table>
<thead>
<tr>
<th>Correlation coefficient (P)</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>L4</th>
<th>L5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.23</td>
<td>0.21</td>
<td>0.13</td>
<td>0.18</td>
<td>0.06</td>
</tr>
<tr>
<td>Weight (lb)</td>
<td>−0.04</td>
<td>−0.07</td>
<td>−0.22</td>
<td>−0.24</td>
<td>−0.26</td>
</tr>
<tr>
<td>BMI</td>
<td>−0.13</td>
<td>−0.17</td>
<td>−0.33</td>
<td>−0.33</td>
<td>−0.40</td>
</tr>
<tr>
<td>Chronicity</td>
<td>0.05</td>
<td>−0.04</td>
<td>0.12</td>
<td>0.03</td>
<td>−0.08</td>
</tr>
<tr>
<td>VAS</td>
<td>0.04</td>
<td>−0.07</td>
<td>0.01</td>
<td>0.01</td>
<td>0.07</td>
</tr>
<tr>
<td>Roland-Morris score</td>
<td>0.09</td>
<td>0.03</td>
<td>0.06</td>
<td>0.08</td>
<td>0.03</td>
</tr>
<tr>
<td>LS active flexion</td>
<td>0.08</td>
<td>0.05</td>
<td>0.02</td>
<td>0.12</td>
<td>0.15</td>
</tr>
<tr>
<td>LS active extension</td>
<td>−0.11</td>
<td>−0.01</td>
<td>−0.10</td>
<td>0.01</td>
<td>0.04</td>
</tr>
</tbody>
</table>

DISCUSSION

The average PAS stiffness measured in our patient population was 3 to 5 N/mm lower than that measured in previous studies that used a fully mechanized testing apparatus. A number of factors could have accounted for this difference, including patient factors and testing methods. Our system does not use a motor to drive the indenter, and so, despite examiner training, there was variability in the rate of loading and inconsistency between loading and unloading rates. The rate of loading in our study, 0.65 cycles/s, was close to the speed (0.50 Hz) used by several investigators in Australia. Their testing showed that rates between 0.5 and 1 Hz produced very similar PAS results. Also, our method of data reduction, using a linear region of the force-displacement curve near the peak load, was styled after that used by Latimer and Lee.

There was, however, a significant difference in the size of the indenting tip used in our study. Our indenter tip is a rectangular tip 720 mm2 in area. Squires et al found that the device most commonly reported had a contoured round shaft with cross-sectional area 314 mm2, whereas the device most commonly reported had a rectangular tip 720 mm2 in area. Squires et al found that a smaller tip (300 mm2) tended to produce larger PAS measures. Some other factor must have been responsible for our generally lower PAS measures.

Our method of PAS measurement uses the force and displacement data of the indenting rod to calculate a stiffness value. Unfortunately, we cannot distinguish the displacement that results from spinal displacement from that due to compression of soft tissue overlying the spinal contact point. The effect of overlying soft tissue is evident in our data as a correlation between BMI and PAS. Patients with a greater BMI are thicker in constitution, which is quite likely a sign of thicker tissues overlying the spine. Compression of tissues between the table and the patient can also contribute to displacement. Increased BMI may have resulted in the lower PAS values seen in our patient population. Lee et al reported a lower mean BMI value in their study (23.8 vs 29.8...
in our sample) but did not suggest a relationship between BMI and PAS. They did, however, note a decrease in PAS at L4 with increasing skinfold thickness over the ilium, a more direct indicator of body fat thickness than BMI.

It is interesting that weight and BMI correlated most strongly with PAS in the lower lumbar segments. Weight is a component of the calculation of BMI, so there should be some correspondence. The greater effect of BMI on lower lumbar PAS scores, though, is probably related to the distribution of body fat. The lower lumbar segments most often have a thicker covering of body fat. It was commonly noted during palpation of spinal landmarks in our study that the L4 and L5 spinous processes were difficult to palpate through the overlying tissues.

Our findings underscore the importance of being able to discriminate the sources of tissue displacement during PAS testing. The ultrasonic indentation system developed by Kawchuk et al has the advantage of being able to measure tissue thickness changes under the indenting head.

The correlation between sex and PAS might be explained by considering BMI. However, BMI was not different between the males and females on our study sample. The correlation between age and PAS was slight but statistically significant, most notably in the upper lumbar spine. It would be interesting to compare the findings in this study with an elderly population to determine if the PAS-age relationship holds true over a wide age range as well.

Active ranges of motion in LS flexion and extension were not related to spinal stiffness. We thought they might be because the action of the PAS test itself induces mild extension at the posterior spinal joints. In a cadaver experiment, Sran et al found that PAS was related to ROM. ROM in the cadaver might be more related to passive ROM in the living human, which we did not measure.

The lack of correlation between indicators of back pain severity (RMQ and VAS) is not surprising in light of the results seen in previous studies. Other authors have found small changes (1.2 N/mm) in patients with LBP on repeat testing when their pain was resolved, but they could not detect any difference in spinal stiffness between patients with LBP and pain-free control subjects. If, as Lee et al suggested and as we have seen in our data, PAS measures depend on patient factors that might not be directly related to spinal stiffness, we could expect significant variance between individuals that might mask differences due to the presence of LBP.

We saw very little difference on the average between PAS measures at different segments of the lumbar spine. This result is different from that of Viner et al who found greater stiffness in the lower lumbar spine. That study was done with healthy pain-free participants rather than patients with LBP as in our study. On the other hand, we saw larger intersegmental differences within patients, in contrast to Viner et al. In our study, the average range between patients’ least and most stiff segments was 5.00 N/mm. Viner et al suggest that any difference between adjacent segments needs to exceed 3.6 N/mm to be greater than that seen in 90% of normal cases. If our system is able to detect intersegmental stiffness differences in patients, this may be a useful tool to determine which segment might be most appropriate for the application of manipulative thrusts. Much more work needs to be done; perhaps looking at the changes that occur with spinal manipulation in a symptomatic patient population will be enlightening.

In our data, we saw very little correspondence between palpation and PAS as indications of segmental stiffness. The palpatory indicator of the stiffest segment and the PAS indicator agreed less than 40% of the time. The main problem with this design as a reliability study is that the clinical palpation tests were not intended to identify a single segment as most stiff or most tender in all patients. If we had seen good agreement, it would suggest the need, perhaps, for a more detailed comparison of palpatory and objective measures of spinal stiffness. From our results, and that of others, there does not seem to be any indication that these measures would agree. A recent study used MRI to measure joint movement during PA spinal loading and compared this measure of spinal stiffness to a clinical test of the patient’s most painful segment. The authors found no association between the degree of intersegmental motion and the level of pain experienced by patients. As Maher and Adams pointed out, spinal stiffness testing as performed by an experienced palpatory indicator is multidimensional, with intersegmental stiffness being perhaps only one dimension evaluated by palpators. Other cues about spinal function are available to the observant human palpator, such as soft tissue consistency, muscular contractions, or viscous properties that our instrument is not intended to detect.

Limitations

Manual palpation was used to locate and mark the spinous processes before PAS testing. The patients were marked in the same prone position that was used for testing so we can be reasonably sure that a spinous process was contacted for the test, but there may be some uncertainty as to the precise segment being tested. None of our palpation methods, whether for landmark location or stiffness or tenderness rating, were tested for reliability before the study. Hence, the lack of agreement between clinical measures and PAS might have been due to inaccuracies in the clinical measures. This is very much the state of the art of palpatory methods for spinal joint function and underscores the importance of developing more objective methods of spinal assessment.

Conclusion

The PAS system of computer-monitored equipment with human operation performed well in this clinical study of
LBP. Spinal stiffness was found to be different between males and females, and age and BMI were related to PAS. We found no significant relationship between the severity or chronicity of the LBP complaint and spinal stiffness. There was little agreement between the stiff or tender segments identified by the clinicians using palpation and the segment that measured most stiff using the PAS device. Future work could be done to measure the sensitivity of PAS measurements in a clinical population of patients with LBP and test whether interventions aimed particularly at changing spine biomechanics affect PAS measurements.

**Practical Applications**

- A system for measuring posterior-to-anterior spinal stiffness performed well in a large clinical study.
- Spinal stiffness was found to depend on patients' sex, age, and BMI.
- We found no significant relationship between the severity or chronicity of the LBP complaint and spinal stiffness.

**ACKNOWLEDGMENT**

The project described was supported by the grant number U19 AT002006 from the National Center for Complementary and Alternative Medicine (NCCAM). The investigation was conducted in a facility constructed with support from the Research Facilities Improvement Program grant number C06 RR15433-01 from the National Center for Research Resources, National Institute of Health. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NCCAM or the National Institutes of Health, Bethesda, MD.

**REFERENCES**

Objective: Over the past decade, instrument and palpation methods for quantifying the activation and recruitment of the transversus abdominis and lumbar multifidus have been proposed. Palpation methods however have recently been described and therefore have been subjected to little evaluation. One such palpation method is the Wisbey-Roth grading system. The recruitment of the transversus abdominis and lumbar multifidus is assessed in a series of functional body positions and movements. The ability to recruit these muscles is quantified by assigning 1 of 6 defined grades. The purpose of this study was to investigate the reproducibility of this grading system.

Methods: A total of 2 meetings and 3 pilot trials were held with raters before commencement of the study to establish an agreed grading system protocol. Interrater reproducibility was investigated using a Latin square repeated measures design. Thirty-four subjects (62% male and 38% female; age range, 15-70 years) with a history of low back pain participated. A total of 4 practicing physiotherapists and 1 sports medicine physician graded subjects using the Wisbey-Roth grading system protocol.

Results: Pair-wise weighted $\kappa$ values ranged from $-0.01$ (95% confidence interval [CI], $-0.33$ to $0.31$) to $0.56$ (95% CI, $0.25$ to $0.87$), with average weighted $\kappa$ being $0.29$. The intraclass correlation coefficient (2,1) was $0.30$ (95% CI, $0.15$ to $0.48$), and the standard error of the measurement was 1.6 units.

Conclusions: The Wisbey-Roth grading system shows fair to poor reproducibility between raters. Therefore, it should not be used to exchange meaningful information between clinicians. Recommendations are made for further research and toward improving its reproducibility. (J Manipulative Physiol Ther 2007;30:501-508)

Key Indexing Terms: Reproducibility of Results; Low Back Pain; Abdominal Muscles; Palpation

It is generally held that the motor control system is responsible for the coordination of the recruitment pattern and activation levels of the lumbopelvic muscles involved in maintaining spinal stability. Specifically, spinal stability has been defined as the control of intervertebral motion, lumbopelvic orientation, and whole-body equilibrium. It is postulated, however, that intervertebral motion beyond physiological limits leads to the onset and/or chronicity of low back pain, which has been termed clinical spinal instability. Although it has been shown that no single muscle dominates in maintaining spinal stability, the specific control of intervertebral motion has been attributed to the local muscles of the trunk. For this reason, it has been suggested that 2 local muscles, the transversus abdominis (TrA) and lumbar multifidus (LM), have segmental stabilizing roles. This is due to observations of their recruitment before the activation of the prime mover of a limb, regardless of the direction of limb movement, and also due to attachments to the lumbar vertebrae and/or thoracolumbar fascia. It is postulated that clinical spinal instability occurs because of dysfunction of the muscles responsible for controlling intervertebral motion, which is ultimately due to alterations in motor control.

Research conducted over the past decade suggests that some patients with low back pain show delayed recruitment of the TrA and LM muscles. In particular, a study investigating experimentally induced low back pain noted that TrA recruitment, unlike other abdominal and LM muscle...
recruitment, was delayed in every participant. Motor control exercise programs aim to address this impairment by teaching the patient how to appropriately recruit these muscles in progressively more challenging positions. There is now also evidence from randomized controlled trials that motor control exercise programs are effective for specific groups of patients with low back pain. As clinicians increasingly prescribe these motor control exercises, the need for clinically appropriate measures of motor control becomes more urgent.

Reproducible and valid measures of motor control allow the clinician to select those patients most likely to benefit from motor control exercise. It may also be used to prescribe appropriate exercises and for the assessment of outcome. Fine-wire and magnetic resonance imaging are gold standard measures of deep trunk muscle recruitment and morphology, respectively. However, they are expensive, not readily available, and fine-wire EMG is invasive, which is not suitable for the clinical setting. More practical methods of measuring TrA and/or LM recruitment have been developed for the clinical setting. Instrumented methods include the pressure biofeedback unit, surface electromyography, and ultrasound imaging. These measures typically evaluate motor control by measuring the activation of the TrA and/or LM muscles. Investigations of the reproducibility and validity of these measures are inconsistent with reported findings ranging from poor to excellent.

In addition to acceptable measurement properties, clinical measures need to be affordable and practical. In our view, current clinical measures of motor control are not sufficiently affordable or practical for routine use in a general clinical setting. For example, although ultrasound imaging can produce reproducible measures of TrA activation, it is probably not a practical measure for the clinical setting because of its relatively high cost and the fact that it requires a highly standardized protocol to provide meaningful measures. The pressure biofeedback unit is an indirect measure of TrA activation and does not allow for testing in more functional positions or movements. Numerous compensatory strategies may also give false-positive results. Therefore, there remains a need for a simple, inexpensive, useful clinical measure of motor control.

Palpation measures, on the other hand, are a practical and inexpensive alternative to existing instruments. Recently, Costa et al. have investigated the intrarater reproducibility of a palpation test of TrA activation and reported moderate reproducibility ($\kappa = 0.52$). The palpation test protocol followed that of Richardson et al., which was also similar to that described by Hides et al. This initial study suggests that a clinical palpation test is not only practical but also provides meaningful information. A palpation measure however would be more useful if it included testing of both the TrA and LM and the ability of the patient to recruited those muscles in functional positions. This would be particularly useful for treatment prescription and assessment of outcome. The Wisbey-Roth grading system is one such palpation measure. Developed in Australia and first described by Wisbey-Roth, this method involves assigning a patient 1 of 6 defined grades (0-5) of motor control. In this grading system, higher grades represent finer motor control and greater stabilizing capacity of the TrA, LM, and pelvic floor muscles. Motor control is measured through manual palpation of the TrA and LM, observation, and verbal cues in a series of functional body positions and movements. Although the grading system shows clinical practicality, its reproducibility has not been evaluated.

Reproducibility is a fundamental requirement of a measurement and can be evaluated using reliability and/or agreement statistics. Although both statistics are applied to repeated measurements, reliability statistics quantify how well an instrument can distinguish patients, whereas agreement statistics focus on how close the repeated scores are. Because the preferred statistic depends upon the application the test user has in mind, it is prudent to describe reproducibility of a measure using both agreement and reliability statistics. The aim of the present study was to investigate the interrater reproducibility of this grading system for patients with current or previous low back pain.

**METHODS**

This study used a Latin square repeated measures design. Raters graded subjects in an anticlockwise direction to account for any effect that may occur due to order of testing, consistent with the Latin square design.

**Subjects**

**Subjects With Low Back Pain.** There were 15 women and 19 men with current or previous low back pain who participated in this study. Subjects were recruited from a private physiotherapy clinic and a medical clinic located in Sydney, Australia. All subjects had experienced an episode of nonspecific low back pain requiring treatment within the 6 months before testing. Most of the subjects were experiencing chronic low back pain, defined as symptoms for greater than 3 months, with 68% experiencing a current episode of low back pain at the time of testing. Subjects were not recruited into this study if the subject had an irritable low back pain condition that would become exacerbated with repeated testing. Characteristics of the low back pain population are reported in Table 1.

**Clinician Raters.** The clinician raters are composed 4 female practicing physiotherapists and 1 male sports medicine medical practitioner. All the raters were experienced in the treatment and management of low back pain and were trained in the use of the Wisbey-Roth grading system. In addition, all raters had used the grading system in clinical practice before commencement of the study. Characteristics of the raters are reported in Table 2.
Informed consent was gained from all participants before entry into the study. Ethical clearance for the study was obtained from the Human Research Ethics Committee, University of Sydney, Australia.

**Materials**

**Wisbey-Roth Grading System.** This grading system requires the testing clinician to assign a score of 0 to 5 depending on the patient’s ability to successfully activate and maintain the contraction of the TrA, LM, and pelvic floor in a series of functional positions. Figure 1 details the procedure used by the clinician raters to determine which grade to assign each subject. Standardized instructions are used to facilitate coactivation of the TrA and LM muscles. The first instruction is, “Breathe in gently then breathe out. When you have breathed out, gently pull the muscles that stop you from urinating slowly up into your pelvis and back towards my fingers that are placed in your back. Keep the muscles on and now continue to breathe comfortably and normally thinking of expanding your lower ribs under your hands.” If this instruction does not elicit an appropriate cocontraction of the muscles, a second instruction is given, “Breathe in gently then breathe out. When you have breathed out, gently pull your lower stomach muscles slowly away from my finger here at the front, back towards my fingers that are placed on your back. Keep the muscles on and now continue to breathe comfortably and normally thinking of expanding your lower ribs under your hands.” Throughout the grades, raters manually palpate the right anterolateral abdominal wall (2 cm medial and 1 cm inferior to the right anterior superior iliac spine) and multifidus (level of right and left L4 transverse processes) to make judgments regarding the activation of those muscles.

Palpation judgments are based on 3 criteria. Firstly, the tester must feel the muscles activate without there being any bulging of the anterolateral abdominal wall or global activity of the trunk muscles. Secondly, the tester must feel that the muscles are activated throughout the duration of the movement or hold required. Thirdly, the tester must feel the muscles relax when the patient is told to relax. Finally, any compensatory movements, such as tilting of the pelvis or rotation of the spine, observed by the testing clinician will be noted as an automatic fail for the grade assessed. A patient passes a grade by fulfilling the palpation criteria while not showing any compensatory behavior. Conversely, a patient fails a grade if the palpation criteria are not fulfilled and/or they show compensatory behavior. Therefore, a grade is allocated to a patient according to the above pass and fail criteria following a specific sequence of testing (Fig 1).

**Procedure**

**Defining the Grading System and Pilot Trials.** Two meetings were held with the investigators and the raters to refine a consensus version of the grading system. Collaboration resulted in the defined Wisbey-Roth grading system protocol. During this time, a flow chart and checklist were designed to document the exact criteria for each grade, and this formed the basis for Figure 1.

Three training sessions were conducted where the raters scored subjects using the established protocol. The first session consisted of 5 participants with a history of LBP. The second and third sessions consisted of 3 participants each with and without a history of low back pain, respectively. Results from these training sessions were evaluated with the raters discussing points of disagreement. After this, the grading system criteria were further refined to improve agreement.

**Power Calculations.** The required sample size was obtained from power tables for the intraclass correlation coefficient (ICC) statistic. The ICC power tables indicate that a sample size of at least 30 participants and 5 raters provides 80% power to detect differences between values of 0.9 and 0.8, between 0.8 and 0.6, between 0.7 and 0.5, between 0.6 and 0.3, between 0.5 and 0.2, between 0.4 and 0.1, and between 0.3 and 0.0.

**Research Procedure.** Five different raters measured 34 subjects with a history of low back pain over 2 days. Each day is composed of three 2-hour testing sessions, which was conducted at a private physiotherapy clinic located in Sydney, Australia. There were 5 to 6 subjects who attended each 2-hour session on one occasion. Participants were given the information sheet, and consent was obtained before
**Fig 1.** Sequence of testing to determine grade allocation using the Wisbey-Roth grading system.

**Legend:**
- Dotted arrow = FAIL
- Solid arrow = PASS
- LL = Lower Limb, UL = Upper Limb
participation in the study. A questionnaire was then completed by the subjects with low back pain in which information concerning anthropometrics, low back pain history, and current status were collected (Table 1). The raters also completed a questionnaire in which data regarding professional history and experience with the grading system were collected (Table 2).

To prevent bias when assigning a grade, raters did not obtain information regarding current low back pain history from the patient, and testing was conducted in individual treatment rooms and/or screened areas. To minimize changes in performance of both raters and patients, raters were also instructed to neither reveal the grade assigned nor provide feedback to patients about their performance. The first rater to grade the patient marked the points for palpation using small round adhesive markers and a black ink pen to control for inconsistencies due to location of palpation points for testing. Finally, in order that the patient would not guess the assigned grade, raters were told to test all participants from grade 0 through to grade 5.

Raters were given a maximum of 15 minutes per assessment. Raters were requested to grade subjects only once and use a maximum of 2 contraction trials to assess each grade for a pass/fail. Volunteer university students and the clinic secretary timed each assessment and gave the raters grading sheets by which to record the grade allocated to each patient. On the grading sheets, raters were instructed to neither reveal the grade assigned nor provide feedback to patients about their performance. The first rater to grade the patient marked the points for palpation using small round adhesive markers and a black ink pen to control for inconsistencies due to location of palpation points for testing. Finally, in order that the patient would not guess the assigned grade, raters were told to test all participants from grade 0 through to grade 5.

Statistical Analyses

Agreement was expressed using the standard error of the measurement (SEM). The SEM was calculated as the square root of the error variance from the ICC (2,1) statistic. In the terminology proposed by de Vet et al, the (2,1) model is referred to as ICC-agreement. Reliability of the grading system was analyzed using weighted $\kappa$ with quadratic weights and the ICC (2,1) statistic. Weighted $\kappa$ with quadratic weights was used to account for the seriousness of the disagreement, so that greater weight was attached to large disagreements on the scale than to small disagreements. We calculated both $\kappa$ and the ICC statistics to accommodate readers who consider the grades from the grading system protocol to represent continuous data and also those who consider them to be ordinal data. This disagreement is of little practical importance because weighted $\kappa$ and the ICC are numerically equivalent when quadratic weights are used.

Unweighted $\kappa$ was used to analyze the interrater agreement of each grade for pass or fail judgments. The scale described by Landis and Koch was also used to test for an effect due to order of testing or rater. Data analysis was performed using SPSS 13.0 (SPSS Inc, Chicago, Ill) and Medcalc 8.0.1.1 (Medcalc, Mariakerke), significance levels were set at .05.

RESULTS

Pair-wise weighted $\kappa$ values are shown in Table 3, with weighted $\kappa$ ranging from $-0.01$ (95% confidence interval [CI], $-0.33$ to 0.31) to 0.56 (95% CI, 0.25 to 0.87). The average weighted $\kappa$ was 0.29. The ICC value across all 5 raters was 0.30 (95% CI, 0.15 to 0.48), and the SEM was 1.6 units. The following were unweighted $\kappa$ values for each grade: grade 0, 0.07 (95% CI, $-0.26$ to 0.40); grade 1, 0.06 (95% CI, $-0.14$ to 0.26); grade 2, 0.23 (95% CI, 0.11 to 0.35); grade 3, 0.30 (95% CI, 0.18 to 0.42); grade 4, 0.30 (95% CI, 0.12 to 0.48); and grade 5, 0.15 (95% CI, $-0.10$ to 0.40). The ANOVA results indicated that there was no effect due to order of testing ($P = .64$), rater ($P = .95$), or an interaction between rater and order of testing ($P = .77$). To assess whether excessive homogeneity could have been responsible for the low reproducibility, we also calculated the frequency with which each of the 6 grades was used. Of the 170 ratings, grade 0 was allocated 45 times; grade 1, 31 times; grade 2, 27 times; grade 3, 17 times; grade 4, 18 times; and grade 5, 32 times.

Discussion

The results of this study clearly show that the current protocol for grading motor control provides ratings of low reproducibility. The ICC and $\kappa$ values align with benchmarks for poor to fair reproducibility. The distribution of grades

<table>
<thead>
<tr>
<th>Rater</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>–</td>
<td>0.32 ($0.01$ to 0.65)</td>
<td>$-0.01$ ($0.33$ to 0.31)</td>
<td>0.30 ($0.01$ to 0.62)</td>
<td>0.11 ($-0.21$ to 0.43)</td>
</tr>
<tr>
<td>2</td>
<td>0.32 ($0.01$ to 0.65)</td>
<td>–</td>
<td>0.20 ($0.11$ to 0.51)</td>
<td>0.39 (0.08 to 0.70)</td>
<td>0.40 (0.09 to 0.71)</td>
</tr>
<tr>
<td>3</td>
<td>$-0.01$ ($0.33$ to 0.31)</td>
<td>0.20 ($0.11$ to 0.51)</td>
<td>–</td>
<td>0.30 ($0.01$ to 0.61)</td>
<td>0.30 ($-0.01$ to 0.61)</td>
</tr>
<tr>
<td>4</td>
<td>0.30 ($0.01$ to 0.62)</td>
<td>0.39 (0.08 to 0.70)</td>
<td>0.30 ($-0.01$ to 0.61)</td>
<td>–</td>
<td>0.56 (0.25 to 0.87)</td>
</tr>
<tr>
<td>5</td>
<td>0.11 ($-0.21$ to 0.43)</td>
<td>0.40 (0.09 to 0.71)</td>
<td>0.30 ($-0.03$ to 0.63)</td>
<td>0.56 (0.25 to 0.87)</td>
<td>–</td>
</tr>
<tr>
<td>Average</td>
<td>0.18</td>
<td>0.33</td>
<td>0.20</td>
<td>0.39</td>
<td>0.34</td>
</tr>
</tbody>
</table>
shows that the low reproducibility is not a consequence of attempting to rate an excessively homogeneous group. The lack of reproducibility cannot be attributed to tester bias because the ANOVA did not indicate that there was an effect due to order of testing (or rater). Furthermore, effort was made to recruit sufficiently experienced raters and allow sufficient training of raters to provide a fair estimate of reproducibility. Lastly, the subjects to be rated were representative of those who would be tested in clinical practice and were not excessively homogeneous. Therefore, it is unlikely that the results of this study are an artifact of poor methodological design but reflect the actual situation with regard to the Wisbey-Roth grading system.

To explore possible sources of disagreement, the inter-rater reproducibility of each grade was analyzed. $\kappa$ analysis indicated that judgments for each grade had similar reproducibility. The error of the overall scale therefore cannot be attributed to any one grade being more difficult to judge than another. This suggests that the error lies within the overall judgments required from the raters, as defined by the criteria of the grading system.

The grading system requires judgments based on manual palpation, which may contribute to the inconsistencies between clinicians. The results of this study are consistent with the results of previous research investigating manual palpation tests of the lumbar spine. Locating spinal levels, in particular L5, has poor interrater reproducibility. This was not a source of error in this study because markers were used so that raters would palpate at exactly the same nominated positions. It is likely that the reproducibility reported in this study would be even lower if raters were required to locate the spinal level for palpation of the skin, subcutaneous tissue, and muscle.

The results also have implications for the task of palpation itself. For example, grade 0 involved palpating for activation of the TrA and LM with the patient in supine position. The reproducibility for this grade was slight ($\kappa = 0.07$). This finding is very different to that reported by Costa et al., where moderate intrarater reproducibility ($\kappa = 0.52$) was reported. These findings considered together suggest that palpation is less reproducible between clinicians, and that superimposing palpation of LM may contribute to increased error. In this study, 68% of subjects experienced greater than mild low back pain before testing (Table 1). This suggests that palpation for deep trunk muscles in patients experiencing low back pain may be more difficult than in patients not experiencing low back pain as in the study by Costa et al. For example, palpation may be particularly difficult with patients who are overweight or who have referred muscle tenderness. To prevent interference with rater judgments, pain was not measured during testing, and therefore, it is not possible to associate pain during testing with poor reproducibility from this study. It is also possible that global extensor and abdominal activity may have been assessed as opposed to the ‘true’ activity of the LM and TrA muscles. A study of the validity of palpation as a test for the activation of these muscles in patients experiencing low back pain is required to confirm this.

Previously, designed measures of motor control involve the use of an instrument. The measurement of TrA thickness using ultrasound imaging has shown excellent intrarater reproducibility ($\text{ICC} >0.93$) in a population with low back pain. Good intra- and interrater reproducibility (weighted $\kappa = 0.61$) of the pressure biofeedback unit was also shown in a healthy population. Therefore, although instruments may be limited in terms of ease of measurement in the clinical setting, they have shown better reproducibility than the manual grading system used here.

A method of improving the reproducibility of the Wisbey-Roth grading system would be either to remove palpation from the current grading protocol or to more explicitly train palpation skills. The latter may be more practical because replacing manual palpation with an instrument such as ultrasound imaging would reduce the clinical practicality of the grading system. Ultrasound imaging however may be used to train the palpatory skills of the clinician. More explicit criteria for palpation may be gained from such a training period, which may improve the reproducibility of the grading system.

Another method to improve the reproducibility of the Wisbey-Roth grading system is to replace the dichotomous scale used with each grade with a scale with more categories. The present scale forces raters to make a pass or fail judgment, and this may not allow sufficient gradation of ability with each grade. Manual grading systems designed to measure the strength of the pelvic floor that have been designed with more than 2 categories for each parameter of contraction have shown acceptable reproducibility. This may be of relevance to improving the reproducibility of the Wisbey-Roth grading system.

Finally, it needs to be considered whether the low reproducibility of the motor control measure argues against its use in the clinical situation. The consequences of error in grading a patient depend upon the seriousness of disagreement between raters. A discrepancy of greater than 3 grades would be serious, as very different clinical judgments would be formed between clinicians. These judgments would be used to prescribe exercise programs, and if there is such a large disagreement between raters, optimal patient outcomes may be compromised.

The grading system used in this study has poor to fair reproducibility, indicating that disagreements of greater than 3 grades are common between raters, which is unacceptable. A finding that aligns with moderate to good reproducibility would be acceptable, as it would indicate that there is less serious disagreement between raters. Therefore, the recommendations made to improve the reproducibility of the Wisbey-Roth grading system should be investigated in the future.
CONCLUSIONS

The findings of this study are that the Wisbey-Roth grading system has low reproducibility and should not be used for exchange of information between clinicians. The poor reproducibility of this palpatory grading system suggests that a source of error may be from the task of palpation itself. Although we have proposed several ways of improving the reproducibility of the grading system, the problem of developing a simple and inexpensive measure of motor control, with acceptable reproducibility and validity, remains.

Practical Applications

- A selected palpatory grading system of motor control showed poor reproducibility in this study.
- This raises many questions for future research:
  - Is palpation a significant source of error in the assessment of motor control?
  - Is palpation of deep trunk muscles a valid technique?
  - Will a different palpatory grading protocol show better reliability?

REFERENCES

34. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968;70:213-20.
THE EFFECT OF COMBINING MANUAL THERAPY WITH EXERCISE ON THE RESPIRATORY FUNCTION OF NORMAL INDIVIDUALS: A RANDOMIZED CONTROL TRIAL

Roger M. Engel, DC, DO, a and Subramanyam Vemulpad, MSc, PhD b

ABSTRACT

Objective: The objective of this study was to explore the effect of combining manual therapy with exercise on respiratory function in normal individuals.

Methods: The study design was a randomized control trial. Forced vital capacity (FVC) and forced expiratory volume in the first second (FEV1) were measured in 20 healthy, nonsmoking individuals before and after 3 interventions: exercise only, chiropractic manual therapy only, and manual therapy followed by exercise. The participants, 18 to 28 years of age, were randomly allocated to a control and 3 intervention groups. Each participant underwent 6 sessions of interventions over a 4-week period.

Results: The exercise only group showed a significant decrease in FVC (P = .002, generalized linear model [GLM]) and FEV1 readings (P = .0002, GLM). The manual therapy only group showed a significant increase in FVC (P = .000, GLM) and FEV1 (P = .001, GLM). The group that received both manual therapy and exercise showed increases in FVC and FEV1 immediately after manual therapy followed by an additional increase after exercise. The overall increase in this group was not statistically significant. Participants in the control group showed no change in FVC or FEV1.

Conclusions: Manual therapy appears to increase the respiratory function of normal individuals. The potential for this intervention administered before exercise to permit additional tolerance within the respiratory system that could allow an extended exercise program than was previously possible is discussed. (J Manipulative Physiol Ther 2007;30:509-513)

Key Indexing Terms: Manual Therapy; Manipulation, Spinal; Respiration; Lung Volume Measurements; Randomized Controlled Trial; Chiropractic

Adequate respiratory function is critical to survival. The cause of over 20% of all deaths in the world is attributed to diseases of the respiratory system, with noncommunicable respiratory conditions responsible for nearly a third of these deaths. 1 This situation is set to deteriorate with the proportion of long-term morbidity attributable to respiratory diseases predicted to increase across many countries by 2020. 2

Spirometry is one method used to evaluate respiratory function. Measurements such as forced vital capacity (FVC) and forced expiratory volume in the first second (FEV1) are common spirometry measures of lung function. Normal adult FVC varies from 1.5 to 5.8 L depending on age, sex, and height, 3 with an FEV1 of less than 80% of the predicted value considered a sign of airflow limitation. 4

‘Normal’ respiratory function typically begins to decline after the age of 25. 5 It is the rate of this decline that is important when attempting to assess future respiratory function. Among the factors contributing to the rate of decline seen in adults is the link to a history of recurrent respiratory tract infections in childhood. 6-8 Screening of respiratory function in apparently healthy adults with a history of recurrent childhood respiratory tract infections could uncover a tendency for a faster rate of decline than would be expected in a similar person without a history of childhood respiratory tract infections. This screening could be a predictor of respiratory function later in life and identify persons who would most likely have a respiratory function at the lower end of the reference range, be asymptomatic, and not present to a health practitioner for anything other than standard respiratory care.

This has led to the suggestion of a possible correlation between childhood respiratory tract infections and the incidence of a number of chronic respiratory conditions that occur later in life such as chronic obstructive pulmonary disease (COPD). 9,10 Chronic obstructive pulmonary disease is characterized by progressive, nonreversible airflow obstruction that does not change markedly over several months. 11 Because there are few presenting symptoms in the
early ‘mild’ stages of this disease,\textsuperscript{4} the timing at which initial diagnosis occurs is critical to the long-term prognosis of the condition. Early detection of an accelerating rate of decline in respiratory function is the cornerstone of current management strategies for improving the prognosis of COPD.\textsuperscript{4,11} In addition to this, preserving normal respiratory function for as long as possible into adulthood in groups identified as being at ‘high risk’ of developing COPD would also benefit strategies aimed at limiting this condition’s morbidity.

In light of the hypotheses that the decline in respiratory function seen in adult chronic respiratory disease may begin as early as childhood and that this decline could be a result of the aftermath of simple respiratory tract infections that leave the system functioning below optimal levels, it is pertinent to explore interventions that have the potential to increase respiratory function in apparently asymptomatic normal individuals.

Monitoring respiratory function in individuals that have no presenting signs or symptoms is currently not part of routine general practice. However, population studies using spirometry to estimate the general prevalence of COPD have shown large variations in rates. In Australia, one study measured 2500 randomly chosen adults older than 18 years and concluded that 24.1% had some degree of COPD;\textsuperscript{12} another study using results from 7 population health surveys over an 18-year period found that 24% of men and 18% of women who were regular smokers and 5% of male and 8% of female nonsmokers had chronic airflow limitation.\textsuperscript{13} In Greece, the rate has been estimated at 8.4% in people older than 35 years.\textsuperscript{14} In the United States, surveillance trends for 1971-2000 estimated the rate at approximately 3% of the total population; however, other data indicated that COPD was underdiagnosed, and that the incidence could be as high as 8% in the United States.\textsuperscript{15}

What these studies draw attention to is the potential for a level of underreporting of COPD in the younger population, the group that traditionally is not considered likely to develop this condition. As outlined above, it is precisely this group that offers the best chance of improving the prognosis of the disease via early detection. What is required is an intervention that has the potential of reversing the long-term negative effects of childhood respiratory tract infections, that is, the cause of any acceleration in the decline in respiratory function.

It has been suggested that manual therapy can improve respiratory function in normal or asymptomatic individuals.\textsuperscript{16,17} Weaknesses in the research design of these studies mean that such evidence is far from convincing. However, the notion that the mobility of the thoracic spine and rib cage can influence respiratory function is an accepted tenet in respiratory anatomy and physiology.\textsuperscript{18}

The purpose of this study was to measure the effect of a combination of manual therapy and exercise on the respiratory function of normal individuals using a randomized control trial.

**METHODS**

In addition to history taking, 50 healthy volunteers between the ages of 18 and 28 years who had been nonsmokers for at least the previous 12 months and who were not currently taking any respiratory medication were subjected to physical examination and spirometry measurements (FVC and FEV\textsubscript{1}) following standard operating protocol for a ‘Spirobank G’ spirometer. From these, 6 volunteers were excluded from the study because of a history of respiratory disease, and 3 were excluded on the grounds of having contraindications to spinal manipulation. Of the remaining 41 volunteers, the 20 volunteers with the lowest FVC readings (7 males and 13 females) were randomly allocated, by the roll of a die, to intervention (die showing up 1, 2, or 3) or control (die showing up 4, 5, or 6) groups. The other 21 volunteers did not take any further part in this trial. Within the intervention group, allocation to groups 1, 2, or 3 (see below) was by way of picking 1 of 15 sealed envelopes. This process ensured that allocation to intervention groups was concealed from both participants and experimenters. Die rolling and envelope selections were conducted by the first author (R.E.).

Participants in group 1 (exercise only [Ex]) underwent a standardized walking treadmill program. This involved adhering to a specific schedule of pace, inclination, and duration. Participants in group 2 (manual therapy only [Mt]) had soft tissue therapy with spinal and rib manipulation administered to their lower cervical, upper and middle thoracic spines, and associated ribs. Participants in group 3 (manual therapy and exercise [MtEx]) had soft tissue therapy with spinal and rib manipulation administered to their lower cervical, upper and middle thoracic spines, and associated ribs followed by the same standardized walking treadmill program as in group 1. Participants in group 4 (control) received neither manual therapy nor a walking program. The exercise component consisted of a 15-minute walking program completed on an electronic treadmill (Healthstream HS2000T/Comet, Selangor Darul Ehsan, Malaysia) with digital time, speed, and inclination readout. The schedule consisted of a predetermined routine of varying speed and inclination settings (Table 1). The spinal manipulation

<table>
<thead>
<tr>
<th>Table 1. Pre-determined schedule of exercise component (groups 1 and 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (min)</td>
</tr>
<tr>
<td>Speed (km/h)</td>
</tr>
<tr>
<td>Inclination setting</td>
</tr>
</tbody>
</table>
component consisted of nonspecific high-velocity low-amplitude manipulation of the lower cervical, upper and middle thoracic spines, and the posterior articulations (costovertebral and costotransverse) of the associated ribs. The soft tissue therapy was applied to the muscles associated with the lower cervical, upper and middle thoracic spines, and the posterior regions of the associated ribs.

Each participant in groups 1, 2, and 3 underwent 6 exercise sessions and/or manual therapy sessions over a 4-week period. Participants in group 4 were assessed at 2-week intervals corresponding to sessions 1, 3, and 6 of the experimental groups. Spirometry measurements were taken 1 minute before and 1 minute after each intervention. For the participants in group 3 (MtEx), spirometry measurements were taken before manual therapy, after manual therapy, and again after completion of the exercise component, that is, the post manual therapy measurement was used as the preexercise measurement for this group. For groups 2 and 3, all soft tissue therapy was administered before any manipulation intervention. For participants in group 4 (control), spirometry readings were taken at the initial consultation and again at 2 and 4 weeks.

This study was conducted from March to May 2005 at a Macquarie University chiropractic outpatient clinic in Sydney, Australia. All interviewing and measurements were performed by 1 of the 6 chiropractic interns associated with this study who was trained and supervised by the first author (R.E.).

Before any volunteer becomes a participant in this study, written informed consent was obtained after all procedures had been fully explained. This study has been approved by the Macquarie University Ethics Review Committee and registered with the Australian Clinical Trial Registry.

In this study, we tested the hypothesis whether chiropractic manual therapy could influence respiratory function. Statistical analysis was performed using the generalized linear model (GLM) method for nested design, wherein the participants and sessions were nested within the interventions, and repeated measurements were nested within the sessions. This meant that for the participants in group 3 (MtEx), who received more than 1 intervention, each postintervention spirometry measurement (FVC1), apart from the initial one at the beginning of each session, became the preintervention reading (FVC0) for the following intervention. This enabled a direct comparison to be made between the effect of components of the combined group (MtEx) vs the single-intervention groups (Mt and Ex). Clinically, the synergistic effect of combining the components may be influenced by the order in which they are administered. The level of statistical significance was set at .05. For this analysis, comparisons were made only between the intervention groups (ie, groups 1, 2, and 3) because the FVC values of the control group were not comparable. Comparison of preintervention FVC (FVC0) and postintervention FVC (FVC1) for each group was made via box plots.

RESULTS

A literature search did not reveal any previous similar studies with standard errors for treatment effect. However, a post hoc analysis revealed that the sample size in our study was adequate to determine both statistical and clinical significance.

Figure 1 depicts the flowchart for this study. It outlines the steps from enrolment through allocation to analysis and includes the number of participants (n) and the number of observations (n’) for all groups. The intervention groups were not matched for sex and age. Group 1 (Ex) is composed of 3 men and 2 women with an age range of 23 ± 3 years; group 2 (Mt), 2 men and 3 women with an age range of 23.5 ± 4.5 years; group 3 (MtEx), 1 man and 4 women with an age range of 20.5 ± 2.5 years; and group 4 (control), 1 man and 4 women with an age range of 20 ± 1 year.

Figure 2 summarizes the overall results of the effect of the interventions on FVC. As can be seen, the control group had lower FVC values. Table 2 outlines the results from the statistical analysis of the regression coefficients for paired values of FVC0 and FVC1 and FEV10 and FEV11 for the intervention groups (ie, groups 1, 2, and 3) using the GLM.

There was a statistically significant decrease in FVC and FEV1 for the Ex group, whereas participants in the Mt group showed a significant increase in FVC and FEV1. Participants in the MtEx group showed an increase in FVC and FEV1 after manual therapy followed by an additional increase after completion of the exercise program. However, the overall increase in this group was not statistically significant. Participants in the control group showed no change in FVC and FEV1 over the study period.

DISCUSSION

The nature and extent of the change in FVC and FEV1 for the 2 groups receiving manual therapy (Mt and MtEx) was similar. These findings are consistent with results of similar studies that examined the effect of manual therapy on normal respiratory function.16,17

The statistically significant decrease in lung function for the Ex group may have been due to the effect of exercise-induced respiratory resistance. Exercise causes smooth muscle relaxation resulting in temporary bronchial dilation.19 This can result in a reduced FVC reading if the measurement is taken soon after the completion of exercise,20 with a suggestion that the response may be reflex in nature.19,21 Because spirometry measurements were taken soon after completion of the exercise component in our study, this may be a possible explanation for the decrease in FVC and FEV1 observed after exercise.

Analysis of the results of the MtEx group opens up the notion of the ability of manual therapy to play a role in modulating normal bodily responses. The nature of the change in lung function after exercise was altered when...
Manual therapy was administered before the exercise. Exercise on its own produced a significant decrease in FVC and FEV₁. When exercise was administered after manual therapy, there was an increase in FVC and FEV₁ attributable to the exercise component. Although this change was not statistically significant, the reversal from a decrease to an increase may be regarded as clinically significant, especially if confirmed by a larger study. Manual therapy appeared to be able to moderate the negative effect of exercise-induced respiratory resistance on lung function.

The mechanism behind the reversal of the short-term negative effects of exercise by manual therapy is unclear. It is possible that the administration of manual therapy to the cervical and thoracic regions before exercising regulates the autonomic nerve supply to the respiratory muscles. This may simply permit a short-term increase in respiratory function that creates additional capacity, which then overrides the negative effect of exercise.

---

**Table 2.** Regression coefficients for the effect of interventions on FVC and FEV₁ using GLM

<table>
<thead>
<tr>
<th>Intervention group (n’)*</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>P</th>
<th>95% confidence intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ex (96)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC</td>
<td>−0.192</td>
<td>0.062</td>
<td>.002</td>
<td>−0.314 to −0.070</td>
</tr>
<tr>
<td>FEV₁</td>
<td>−0.185</td>
<td>0.048</td>
<td>.0002</td>
<td>−0.281 to −0.089</td>
</tr>
<tr>
<td>Mt (112)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC</td>
<td>0.208</td>
<td>0.056</td>
<td>.000</td>
<td>0.097 to 0.318</td>
</tr>
<tr>
<td>FEV₁</td>
<td>0.148</td>
<td>0.044</td>
<td>.001</td>
<td>0.060 to 0.236</td>
</tr>
<tr>
<td>MtEx (120)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC</td>
<td>0.048</td>
<td>0.056</td>
<td>.386</td>
<td>−0.061 to 0.158</td>
</tr>
<tr>
<td>FEV₁</td>
<td>−0.026</td>
<td>0.043</td>
<td>.554</td>
<td>−0.112 to 0.060</td>
</tr>
</tbody>
</table>

* n’ indicates the number of observations. Differences in n’ are due to dropouts in the Ex and Mt groups (see Fig 1).

---

**Fig 1.** Flow diagram showing the number of participants (n) and the number of observations (n’) for each group.

**Fig 2.** Box plots of the distribution of FVC measurements before and after intervention for each group. The FVC was measured before and after each intervention for the intervention groups (i.e., Ex, Mt, and MtEx). These are given as FVC₀ and FVC₁. For the control group, FVC₀ is the measurement at the beginning of the study, and FVC₁ is the measurement at the end of the study.
Reflex bronchial dilation with resultant drop in FVC after exercise could be counterproductive in certain pathological scenarios. The ability for manual therapy to modify the short-term negative effects of exercise may be relevant when considering the design of pulmonary exercise rehabilitation programs for the chronic respiratory patient. Applying manual therapy before exercise may permit additional tolerance within the respiratory system, allowing the patient to undertake an extended exercise program than was previously possible.

This approach may be beneficial when designing strategies aimed at maintaining maximal respiratory function for as long as possible in groups deemed to be at high risk of developing COPD later in life.

The small number of participants does limit the generalizability of the results. Not matching the body types, sex, and age between groups, as well as the narrow range of age group of the participants may have influenced the results of this study. Furthermore, in light of the evidence of exercise-induced respiratory resistance, a better understanding of the effect of exercise on lung volume may have been possible had additional FVC measurements at longer postintervention intervals been included in this study.

CONCLUSION

This study indicates the possible beneficial effect of manual therapy on respiratory function in healthy participants. It also highlights the potential for delivering additional benefit to respiratory function when manual therapy is combined with and precedes mild exercise. In light of these findings, further studies on the effect of manual therapy with and without exercise on lung function are warranted.

Practical Applications

- Chiropractic manual therapy improves lung function
- Chiropractic manual therapy overcomes exercise induced short-term respiratory resistance

ACKNOWLEDGMENT

The authors would like to acknowledge students enrolled in the Master of Chiropractic program at Macquarie University for their assistance in conducting this study, and Professor Don McNeil and his PhD student, Ms Sangdao Wongsai (Department of Statistics, Macquarie University), for their valuable help with the statistical analysis.

REFERENCES


INTEREXAMINER RELIABILITY OF THE PRONE LEG LENGTH ANALYSIS PROCEDURE

Michael Schneider, DC, Robert Homonai, DC, Brian Moreland, DC, and Anthony Delitto, PhD, PT

ABSTRACT

Objective: The purpose of this study was to perform an interexaminer reliability evaluation of the prone leg length analysis procedure.

Methods: Two chiropractors each examined a series of 45 patients with a history of low back pain. Patients were in the prone position, with the knees in both extended and flexed positions, and with the head rotated right and left. The clinicians were asked to determine the side of the short leg with knees extended and if a change in leg length occurred with head rotation or when the knees were flexed. They were also asked to visually judge the amount of leg length differential by categorizing the difference as either less than 0.25, 0.25 to 0.5, 0.5 to 0.75, or more than 0.75 in. The head rotation portion of the test was performed only with patients (n = 22) in whom the leg length differential was determined to be less than 0.25 in.

Results: \( \kappa \) statistics and frequency distributions were calculated for each of the respective observations. Reliability of determining the side of the short leg with knees extended was good at 82% agreement (\( \kappa = 0.65 \)) but fair for determining the amount of leg length difference at 67% agreement (\( \kappa = 0.28 \)). Reliability of the head rotation testing procedure was extremely poor, with only 50% and 45% agreement about the observed change in leg length with the head rotated left and right, respectively (\( \kappa = 0.04 \), \( \kappa = -0.195 \)). There was no significant correlation found between the side of reported pain by the patient and the side of the short leg as noted by either clinician (\( \chi^2 = 0.55 \), \( P = .91 \), and \( \chi^2 = 1.55 \), \( P = .67 \)). All of the patients (100%) were judged to have a leg length difference by both clinicians. When the knees were flexed, there was 93% agreement that the short leg became longer (43/45 cases), with no reported cases of the short leg getting shorter. Calculation of \( \kappa \) statistics was confounded for these last 2 observations because of extremely high prevalence bias.

Conclusions: The results indicate that 2 clinicians show good reliability in determining the side of the short leg in the prone position with knees extended but show poor reliability when determining the precise amount of that leg length difference. The head rotation test for assessing changes in leg length was unreliable in this sample of patients. There does not appear to be any correlation between the side of pain noted by the patient and the side of the short leg as observed by the clinicians; all 45 patients in this sample were found to have a short leg by both clinicians. (J Manipulative Physiol Ther 2007;30:514-521)

Key Indexing Terms: Leg Length Inequality; Low Back Pain; Physical Examination; Observer Variation; Reproducibility of Results; Chiropractic

Some have argued that leg length inequality (LLI) is a clinical indicator of altered muscle and joint biomechanics. A key differentiation is whether or not any observed LLI is purely anatomical or “functional” in nature.

Anatomical LLI is thought to be congenital or acquired and can only be corrected by use of a heel or shoe lift. Functional LLI is considered to be transient and reversible and merely a sign of altered joint biomechanics or asymmetry of muscular contraction. Travell and Simons suggest that functional LLI may be due to either an asymmetrical contraction of the quadratus lumborum muscle or sacroiliac joint dysfunction. Schaeffer suggests that functional LLI is due to posterior innominate rotation at the sacroiliac joint, which causes an apparent “shortening” of that limb as the acetabulum is carried superior and anterior.

Various types of analytic procedures have been developed to analyze and measure LLI, the most standard of which has been plain film radiography. Lawrence published a review of the literature regarding radiographic procedures for detecting anatomical LLI as well as clinical examination methods for detecting functional LLI. Radiography in the standing weight-bearing position (barefoot) with the x-ray beam
directed at the top of the femurs is clearly the gold standard method of measuring LLI, with a documented measurement error of less than 3 mm. Several studies have reported the radiographic procedure for measuring LLI and found that patients with a history of low back pain tend to have a higher incidence of LLI as compared with healthy controls, that osteoarthritis is associated with the hip joint on the longer limb side, and that lumbar osteoarthritis is associated with an LLI.4-8

Although radiography is considered the gold standard for accurate measurement of anatomical LLI, it is generally recognized that radiography exposes patients to ionizing radiation and is not suitable for the purpose of routine postural screening examinations for mechanical dysfunction. Chiropractors have routinely used nonradiographic observational screening methods to determine the presence of LLI, the most common of which is some type of visual inspection of prone leg length. The Derifield-Thompson leg check as published in the 1984 Thompson Technique Reference Manual9 is one of the earliest published descriptions of a chiropractic screening method for LLI.

The Derifield-Thompson leg check procedure was developed to provide chiropractors with a quick clinical procedure to identify an LLI, by having the patient lie prone and observing the side of the “short leg.” Once the short leg side was identified, the clinician would flex the patient’s knees to 90° and observe for any changes in the leg length that purportedly were related to “pelvic syndromes.” These pelvic syndromes were categorized as either Derifield “positive” or “negative”; a positive finding was defined as the short leg observed to lengthen during knee flexion, and a negative finding was defined as the short leg showing no change in length during knee flexion.9 The Derifield test had another component involving head rotation that purportedly could differentiate between cervical or lumbo-pelvic biomechanical dysfunction as the cause of observed LLI.10 It simply suggested that if head rotation made a difference in the length of the observed short leg, the patient had a “cervical syndrome,” which indicated manipulation of the cervical spine to correct the short leg. If the leg length did not change with head rotation, the patient had a pelvic syndrome.

The Activator Methods Chiropractic Technique (AMCT) protocol is another method of prone leg length analysis performed by chiropractors. The AMCT procedure11 of observing for an apparent LLI is basically an adaptation of the first part of the Derifield-Thompson leg check, without incorporating the head rotation part of the test. The short leg as observed in the prone-lying patient is critically important in AMCT because virtually all of the additional tests used to determine joint dysfunction are based on reactions and changes in the relative length of the short leg. There are 3 types of tests (isolation, stress, and pressure) and 1 rule (short-long rule) in the AMCT prone leg check protocol,11 all of which rely on the observation of reactivity of the short leg to various movements or forces applied to the spine or other joints.

The present study was designed to study the interexaminer reliability of prone leg length analysis using 2 chiropractic examiners who had extensive experience with these procedures and were “advanced proficiency certified” by Activator Methods International, LTD (Phoenix, AZ). The focus of this study was to determine the level of reliability of these observational determinations of functional LLI between 2 clinicians and not to validate or correlate these findings with anatomical LLI as determined by radiography. Determination of a reasonable level of reliability is typically considered a prerequisite to further research questions about validity and clinical utility of diagnostic tests or examination procedures.

Methods

Subject Recruitment and Examination Procedures

Forty-five patients with a history of low back pain were recruited from a private chiropractic clinic to volunteer for participation as research subjects in an interexaminer reliability study regarding prone leg length analysis. All of the patients were given a detailed explanation of the study and signed an informed consent document that was approved by the institutional review board of the University of Pittsburgh, Pittsburgh, Pa. They received no compensation for their participation in this study. All volunteers were existing patients at a clinic that uses the AMCT and were receiving a prone leg analysis as part of the normal examination process on their routine office visits.

Inclusion criteria consisted of patient age between 18 and 65 years, history of low back pain, ability to tolerate the prone position, and agreement to participate in this research study. Exclusion criteria consisted of pregnancy, severe symptoms that would preclude the ability to lie prone for a few minutes, previous lumbar surgery, or any red flags of serious illness or pathology noted during the case history. As noted earlier, this study was not designed to validate or correlate the observational findings of functional LLI with any gold standard of anatomical LLI such as standing radiography. Therefore, we did not gather any radiographic data about potential anatomical LLI in these patients.

Each patient was escorted into a treatment room and lowered from the standing position to the prone position by use of a mechanical electric elevation treatment table (Softec Model no. ST-777, Division of Tri W-G, Valley City, ND) to minimize any changes in pelvic or lower extremity position during the transition from standing to the prone position. After the patient was lying prone in a comfortable position, the first clinician would enter the room and perform a prone leg length analysis with the principal investigator recording the results. After the clinician was finished with his analysis, he left the room, and the second clinician entered the room to repeat the process.
The patients remained prone on the examination table between the 2 examinations and were not brought up to the standing position and repositioned. It was thought that by having the patients remain in the same position for both examinations, we could eliminate the potential of repositioning error as a potential confounding variable. Also, the clinicians were not permitted to move or change the patients’ positions on the table during the examination procedure. The 2 clinicians were blinded to the patient’s side of low back pain (if any) and each other’s findings and were not permitted to speak with the patient during the examination process. The patient received no treatment that day until after both clinicians performed the leg length analysis.

The leg length analysis used by the 2 clinicians followed this protocol.

1. Observation of leg length with knees extended, with visual judgment about the side of the short leg and the amount of such difference.
2. Observation of any change in the length of the short leg side after the clinician flexes the patient’s knees to 90°.
3. If the leg length differential is judged to be no more than 0.25 in, the patient was asked to rotate the head to the left, back to center, and then to the right, with another judgment made as to whether or not the short leg changes in length with such rotational movements.

No ruler or tape measure was used in this study; the examiners were asked to eyeball the perceived amount of leg length difference.

Before initiation of the study, the principal investigator met with the 2 clinicians on several occasions to review and practice the leg length analysis, making sure that both clinicians were following the same procedures. In addition to these training sessions, both clinicians were advanced proficiency certified by Activator Methods International, LTD, and well experienced with these protocols. They both routinely used the prone leg length check on all patients in their clinical practices on a daily basis because both routinely used the prone leg length check on all patients.

A data collection form was designed to capture several findings regarding the leg length analysis, including the following information:

- side of low back pain as reported by the patient on the day of examination, recorded as “none,” “right,” “left,” or “central/bilateral”;
- if a short leg was observed with the knees extended (position 1) by each clinician, recorded as “present” or “not present”;
- when a short leg was present, record of the side of the short leg as determined by each clinician in position 1, recorded as left or right;
- amount of leg length difference as visually observed by each clinician in position 1, approximated into 4 categories, less than 0.25, 0.25 to 0.5, 0.5 to 0.75, or more than 0.75 in;
- any changes in the short leg length observed with rotation of the head left and right, recorded as “gets shorter,” “no change,” or “gets longer”; and
- any changes in the short leg length observed when the clinician flexed the patient’s knees to 90° (position 2), recorded as gets shorter, no change, or gets longer.

**Statistical Analysis**

Sample size was calculated using various possible values of coefficients of determination (ρ) or κ values of 0.40 or higher, with the reasoning that any value less than 0.40 would indicate poor reliability and would be clinically irrelevant. Choosing an α level of .05 and β level of 0.20, sample size calculation indicated a minimum sample size of N = 38 subjects to capture a level of reproducibility in the fair to moderate range (κ = 0.40). Therefore, the final sample size of N = 45 had 80% power to detect a significant κ value of 0.40 or higher.

Statistical analysis was performed using the Statistical Package for the Social Sciences version 13.0 statistical software program (SPSS Inc, Chicago, Ill). Because all data were categorical in nature, statistical analyses consisted of percentages of agreement (concordant observations); χ2 cross-tabulations with calculation of relevant P values for determination of any significant correlations between observed side of the short leg and patient-reported side of pain; and κ statistics for the level of agreement between the 2 clinicians for side of the short leg with knees extended, amount of leg length difference, changes in short leg with head rotation, and changes in short leg with knee flexion. All κ values were reported with 95% confidence intervals (CIs). Standards for κ values were interpreted according to the guidelines described by Landis and Koch.

The κ statistic is sensitive to the base rate or proportion of observations and may be unreliable when there is a very high difference between the prevalence of positive and/or negative results. For this reason, we included calculations of the prevalence and bias index for each respective κ value.

Because the magnitude of any κ value is greatly affected by its associated prevalence and bias index values, it is difficult to interpret κ values and raw percentages of agreement without taking into account index values. Higher prevalence index values indicate higher possible chance agreement and will lower the respective κ values, whereas higher bias index values indicate lower chance agreement and will raise the respective κ values. To allow for reasonable interpretation of the magnitude of the κ values, we reported both the
RESULTS

The frequency distributions of the side of reported pain are depicted in Figure 1. Note that there is a wide variety of pain patterns, with most of the patients reporting central pain (n = 17) and almost equal distributions of right- (n = 9) and left-sided (n = 10) pain. A number of patients (n = 9) also reported feeling no pain on the day of examination. This wide variation of the pain patterns allowed for a reasonable exploration of any potential correlation with the side of leg length difference.

$\chi^2$ analysis using cross-tabulation of the side of the short leg with the side of reported pain does not show a statistically significant correlation above chance observation. Two cross-tabulations were performed, 1 for each of the 2 examiners as separate calculations, with the results depicted in Tables 2A and 2B. For the first clinician, $\chi^2 = 0.55$ ($P = .91$), and for the second clinician, $\chi^2 = 1.55$ ($P = .67$). Note that the raw data from these tables for both clinicians show almost equal distribution of right and left short legs with both right and left side of pain, no pain, and central pain, without any apparent pattern or correlation between the variables.

When comparing the 2 examiners’ observations about the side of the short leg, they showed 82% agreement, with a $\kappa$ value of 0.65 that would be interpreted as “good” to “substantial” interexaminer reliability. These results are summarized in Table 3. Both examiners found the presence of LLI in all 45 research subjects, and no subject was judged to have “even legs” by either examiner.

The 2 examiners were asked to determine the approximate amount of leg length difference, using 4 categories of less than 0.25, 0.25 to 0.5, 0.5 to 0.75, and more than 0.75 in. There were no cases of either clinician reporting a difference of more than 0.75 in, and only 2 reported cases of 0.5- to 0.75-in differences, in which the 2 examiners did not agree (discordant cells). There was an overall 62% agreement of the amount of leg length difference, which calculates as a $\kappa$ value of 0.22 that would be interpreted as “fair” reliability. The data for the amount of LLI are summarized in Table 4.

---

**Table 1.** All data from this study put in a singular tabular format to show $\kappa$ values with their associated 95% CIs, raw percentages of agreement, and prevalence and bias index values

<table>
<thead>
<tr>
<th>Clinical procedure</th>
<th>$\kappa$</th>
<th>95% CI</th>
<th>% Agreement</th>
<th>Prevalence index</th>
<th>Bias index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Determination of which is the short leg side</td>
<td>0.65</td>
<td>0.43-0.87</td>
<td>82%</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>Estimation of amount of difference (0.25 in, 0.5 in, etc)</td>
<td>0.22</td>
<td>0.01-0.55</td>
<td>62%</td>
<td>0.31</td>
<td>0.16</td>
</tr>
<tr>
<td>Change in short leg with head rotation to left</td>
<td>0.04</td>
<td>−0.25 to 0.33</td>
<td>50%</td>
<td>0.41</td>
<td>0.36</td>
</tr>
<tr>
<td>Change in short leg with head rotation to right</td>
<td>−0.20</td>
<td>−0.30 to 0.38</td>
<td>45%</td>
<td>0.45</td>
<td>0.41</td>
</tr>
<tr>
<td>Change in short leg with knees flexed</td>
<td>0.0*</td>
<td>−1.0 to 1.0</td>
<td>93%*</td>
<td>0.93</td>
<td>0.00</td>
</tr>
<tr>
<td>Observation of short leg</td>
<td>0.0*</td>
<td>−1.0 to 1.0</td>
<td>100%*</td>
<td>1.0</td>
<td>0.00</td>
</tr>
</tbody>
</table>

* The discrepancy between the 93% and 100% raw agreements and the 0.0 $\kappa$ values found in these 2 rows is explained by the extremely high prevalence index for these observations.

---

**Table 2. A. Side of pain vs side of short leg (examiner 1)**

<table>
<thead>
<tr>
<th>Side of pain</th>
<th>Right</th>
<th>Left</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Right</td>
<td>5</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Left</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Central</td>
<td>10</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>21</td>
<td>45</td>
</tr>
</tbody>
</table>

No significant correlation is found between the side of the short leg and side of pain. $\chi^2 = 0.55$, $P = .91$.

**Table 2. B. Side of pain vs side of short leg (examiner 2)**

<table>
<thead>
<tr>
<th>Side of pain</th>
<th>Right</th>
<th>Left</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Right</td>
<td>3</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Left</td>
<td>6</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>Central</td>
<td>9</td>
<td>8</td>
<td>17</td>
</tr>
<tr>
<td>Total</td>
<td>22</td>
<td>23</td>
<td>45</td>
</tr>
</tbody>
</table>

No significant correlation is found between the side of the short leg and side of pain. $\chi^2 = 1.55$, $P = .67$. 

prevalence and bias index values along with each respective $\kappa$ value in Table 1.
Note that there were 22 cases of agreement on a finding of LLI less than 0.25 in.

The head rotation portion of the testing procedure (Derifield cervical test) was only performed on these 22 patients whom the 2 examiners agreed had a leg length difference of less than 0.25 in. \( \kappa \) values for head rotation to the right and left were calculated separately. For head rotation to the right, agreement was 50%, with \( \kappa = 0.04 \), and for head rotation to the left, agreement was 45.5%, with \( \kappa = -0.19 \). These \( \kappa \) values indicate virtually no inter-examiner reliability above chance observation. The negative \( \kappa \) value reported for head rotation to the left is not a mistake; it simply indicates that the observed agreement was less than 50% (chance observation), which leads to a negative \( \kappa \) value. The data for the head rotation tests (\( n = 22 \)) are listed in Tables 5A and B.

The last 2 aspects of the statistical analysis involved calculations of \( \kappa \) statistics for the second step of the Activator leg check procedure (position 2) involving knee flexion to 90° and observing for any changes on the short leg side and the simple question, “Is some amount of leg length inequality present?”

There was 93% agreement that the short leg “got longer” during knee flexion; however, neither examiner reported any case in which the short leg “got shorter.” Both clinicians reported the presence of a short leg in all 45 patients (100% agreement). In both of these data analyses, calculation of the \( \kappa \) statistic produced a value of 0.0, with a CI spanning from −1.0 to 1.0. These relatively meaningless \( \kappa \) statistics are due to confounding by the extremely high prevalence of one finding, “short leg gets longer.”

Note that both examiners always observed that a short leg was present (prevalence index of 1.0) and never found any cases of equal leg length.

### Table 3. Comparison of side of short leg observed by examiners 1 and 2

<table>
<thead>
<tr>
<th>Short leg side (examiner 2)</th>
<th>Right</th>
<th>Left</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short leg side (examiner 1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>19</td>
<td>5</td>
<td>24</td>
</tr>
<tr>
<td>Left</td>
<td>3</td>
<td>18</td>
<td>21</td>
</tr>
<tr>
<td>Total</td>
<td>22</td>
<td>23</td>
<td>45</td>
</tr>
</tbody>
</table>

Agreement, 82%; \( \kappa = 0.65 \); 95% CI, 0.43 to 0.87.

### Table 4. Comparison of observations of amount of short leg difference by examiners 1 and 2

<table>
<thead>
<tr>
<th>Examiner 2</th>
<th>( \leq 0.25 ) in</th>
<th>&gt;0.25 in, ( &gt;0.5 ) in</th>
<th>( &gt;0.75 ) in</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examiner 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \leq 0.25 ) in</td>
<td>22</td>
<td>11</td>
<td>0</td>
<td>33</td>
</tr>
<tr>
<td>&gt;0.25 in, ( \leq 0.5 ) in</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>&gt;0.5 in, ( \leq 0.75 ) in</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>&gt;0.75 in</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>18</td>
<td>1</td>
<td>45</td>
</tr>
</tbody>
</table>

Agreement, 62%; \( \kappa = 0.22 \); 95% CI, 0.01 to 0.55.

### Table 5. A. Observations of change in short leg during rotation of the head to the left

<table>
<thead>
<tr>
<th>Examiner 2</th>
<th>No change</th>
<th>Shorter</th>
<th>Longer</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examiner 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No change</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Shorter</td>
<td>2</td>
<td>10</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>Longer</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>3</td>
<td>13</td>
<td>6</td>
<td>22</td>
</tr>
</tbody>
</table>

Agreement, 50%; \( \kappa = 0.04 \); 95% CI = −0.25 to 0.33. Note that only those patients with leg length differences less than 0.25 in were examined with the head rotation procedure (\( n = 22 \)).

### Table 5. B. Observations of change in short leg during rotation of the head to the right

<table>
<thead>
<tr>
<th>Examiner 2</th>
<th>No change</th>
<th>Shorter</th>
<th>Longer</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examiner 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No change</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Shorter</td>
<td>3</td>
<td>10</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>Longer</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td>15</td>
<td>3</td>
<td>22</td>
</tr>
</tbody>
</table>

Agreement, 45%; \( \kappa = -0.20 \); 95% CI = −0.30 to 0.38. Note that only those patients with leg length differences less than 0.25 in were examined with the head rotation procedure (\( n = 22 \)).

### Table 6. Observations of change in short leg when the knees are flexed to 90°

<table>
<thead>
<tr>
<th>Examiner 2</th>
<th>Shorter</th>
<th>Longer</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examiner 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No change</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Shorter</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Longer</td>
<td>1</td>
<td>42</td>
<td>43</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td>43</td>
<td>45</td>
</tr>
</tbody>
</table>

Agreement, 93%; \( \kappa = 0.0 \); 95% CI, −1.0 to 1.0. The \( \kappa \) value in this case is zero because of the extremely high prevalence of one finding, “short leg gets longer.”

### Table 7. Observation of a short leg (LLI) by examiners 1 and 2

<table>
<thead>
<tr>
<th>Short leg present?</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examiner 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>45</td>
<td>–</td>
<td>45</td>
</tr>
<tr>
<td>No</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Total</td>
<td>45</td>
<td>–</td>
<td>45</td>
</tr>
</tbody>
</table>

Agreement = 100%; \( \kappa = 0.0 \); 95% CI = −1.0-1.0. The \( \kappa \) value in this case is zero because both examiners always observed that a short leg was present (prevalence index of 1.0) and never found any cases of equal leg length.
flexion portion of the test and for the side of the short leg are listed in Tables 6 and 7, respectively.

Table 7 lists all of the reliability analyses in one large table, including the $\kappa$ values, 95% CIs, raw percentage of agreement, and prevalence and bias index values.

**DISCUSSION**

There are several previous studies in the literature regarding the prone leg length analysis procedure, both with and without the head rotation component. Nguyen et al.\textsuperscript{14} reported 85% agreement between examiners regarding the side of reported short leg, with an unweighted $\kappa$ value of 0.66. Our data analysis revealed 82% agreement on the side of the short leg, with a $\kappa$ value of 0.65. In the Nguyen et al.\textsuperscript{14} study, only 1 examiner reported 2 patients who had equal leg lengths, whereas in our study, neither examiner reported any case in which the legs were of equal length. Fuhr and Osterbauer\textsuperscript{15} reported fair to good $\kappa$ values, ranging from 0.31 to 0.75 for the results of an interexaminer reliability study involving 30 subjects examined in the prone knee-extended position by 4 examiners. The examiners in this study used a 3-point scoring system (right short/even/left short) for LLI observations and then measured the absolute difference with a ruler placed against the heels of the shoes. It was interesting to note that none of examiner pairs ever agreed upon the absence of LLI and that each found that all subjects had some level of LLI.

The data analysis in our study shows good reliability for the determination of the side of the short leg. This aspect of our study that shows relatively good reliability for the prone knee-extended position of leg length analysis ($\kappa = 0.65$) is consistent with previous findings by Fuhr and Osterbauer\textsuperscript{15} and Nguyen et al.\textsuperscript{14}. It was interesting to note that all 45 patients in our study were determined to have a LLI by both examiners, with no patients being found to have equal leg length. This is consistent with the data from the 2 previous studies just cited.

If all patients with low back pain (and who are potentially asymptomatic) indeed have some degree of LLI, it is questionable what clinical utility is to be gained from the observation of a phenomenon that is so universal, regardless of the fact that 2 examiners can reliably agree on the side of the short leg. This question is especially relevant when the data analysis shows that the side of LLI does not appear to be correlated with the side of low back pain. However, from a chiropractic perspective, it could be argued that LLI may be associated with clinical findings other than just pain, such as sacroiliac or lumbar facet joint dysfunction, which are not necessarily symptomatic at the time of examination. This hypothesis was not tested in the present study and would require additional reliability studies that incorporate a different type of research design. This would require using validated methods of determining that such joint dysfunction existed in a population of patients, examining them all for LLI, and then using correlation statistics to test for any significant association between the observed leg length change and the side/level of the joint dysfunction(s).

Some additional literature on the reliability of the prone leg length analysis is also worth reviewing. DeBoer et al.\textsuperscript{16} reported good reliability during performance of an intra- and interexaminer reliability study of the prone leg length check using 40 students each examined twice by 3 different chiropractors in the knees extended and flexed positions. However, their data analysis was later questioned by Danelius.\textsuperscript{17} Rhudy and Burk\textsuperscript{18} performed an interexaminer reliability study on the Thompson leg check procedure using a number of different clinical examiners, including Dr. Thompson himself, and failed to find any significant level of agreement between examiners. They concluded that their results demonstrated a lack of consistency and the need for more objective methods of the evaluating the spine other than the prone leg check.

Shambaugh et al.\textsuperscript{19} performed an interexaminer reliability study of the Derifield-Thompson cervical test with 26 subjects and 5 different examiners and reported that the clinicians could reliably measure an LLI to less than 3 mm using a ruler between the soles of the shoes and could reliably detect changes in LLI when the head was rotated. However, this study was subsequently criticized for improper statistics.\textsuperscript{20} Venn et al.\textsuperscript{21} recruited 60 subjects and compared standing LLI measurements by radiography, supine measurement of LLI with a tape measure, and prone/supine LLI measured categorically for left short/even/right short legs. They concluded that the prone leg length check was unreliable and of questionable value.

Falltrick and Pierson\textsuperscript{22} tested the reliability of the cervical rotation test for LLI using a novel design of having the chiropractic examiners first determine if the cervical spine of the research subjects was “lesioned” or not, to test the hypothesis of LLI being correlated with cervical spine “lesions”. They also performed ipsilateral electrical muscle stimulation to the lumbar paraspinal muscles to simulate muscle spasm, with a drape over the back of the research subjects to blind the examiners who were performing the leg length analysis. The results of this study did not show any significant effect of head rotation on LLI. Subjects who were classified by the examiners as “cervically lesioned” or “not cervically lesioned” were not able to be differentiated based upon any perceived LLI. Lastly, any leg length differentials noted by the examiners did not correlate with the subjects who had the electrical stimulation turned on or off during the LLI examination.

The results of our study also showed very poor reliability for the cervical rotation test. When patients were examined with rotation of the head to each side (Derifield-Thompson cervical test), the results were determined to be completely unreliable in this cohort of patients. The agreement was not much better than chance when the
patients had their heads rotated to the left ($\kappa = 0.04$) and actually less than chance when turning their heads to the right ($\kappa = -0.19$). Considering the absence of evidence for this clinical procedure in the past literature, and confirmed by the results of our study, it seems that this test is of dubious clinical value.

One potential weakness in our study design was that we only performed the head rotation on patients who showed a 0.25-in or less LLI in the prone knee-extended position, making the total number of patients tested with the Derifield procedure about half of the full sample size ($n = 22$). In retrospect, we could have tested all 45 patients with head rotation, regardless of the amount of LLI. However, the 2 clinicians in this study routinely perform the head rotation test only when they perceive the LLI to be 0.25-in or less, and therefore, our study was designed to be pragmatic for these clinicians.

Our power analysis determined a need for a sample size of 38 subjects to detect a $\kappa$ value of 0.40 with 80% power. Therefore, one might argue that our reduced sample size of 22 would lead to the possibility of a type II error (false-negative findings). However, a sample size of 22 patients is not insignificant, but when 2 examiners cannot agree on their observations at a level greater than chance (left rotation, 45%; right rotation, 50%), it still makes this protocol appear to be very unreliable. Furthermore, even if a larger sample size were used and showed a statistically significant $\kappa$ value less than 0.40, this “statistically significant value” would be considered fair to poor reliability and would therefore be clinically insignificant.

This brings up an interesting point regarding key differences between reliability and validity of diagnostic tests in general. The reliability of any diagnostic test is merely the issue of whether or not it will produce the same result when repeated. Clearly, if a test’s results cannot be replicated when performed on 2 separate occasions or when performed by 2 different examiners, the value and usefulness of that test is questionable. Conversely, just because a test is reliable and produces the same result when repeated does not necessarily indicate that the test is clinically useful or a valid diagnostic procedure.

The validity of any screening or diagnostic test is defined as the ability of that test to distinguish between people who have a certain disorder or disease and those who do not. Determining the validity of screening and diagnostic procedures requires testing of large numbers of people with and without a specific disorder, having a gold standard for determining the presence of this disorder, and then comparing the validity of the test in question against the gold standard test. Validity of diagnostic tests is typically reported in terms of sensitivity and specificity: the ability of the test to correctly identify those persons who actually have the disorder or disease (sensitivity) and to correctly identify those who do not have the disorder or disease (specificity).

This study was designed only to analyze the reliability of the prone leg length assessment and not its validity. However, it would seem reasonable that procedures that show poor reliability, such as the head rotation portion of the Derifield-Thompson leg check and eyeballing the amount of LLI, may be unsuitable for further validity testing, for the reasons noted. On the other hand, procedures with good reliability such as the prone leg length analysis with knees extended cannot automatically be considered valid; they still require validity testing to determine their clinical usefulness for detecting spinal joint dysfunction or other biomechanical faults. Lastly, those procedures whose reliabilities are still unknown, such as the seemingly universal presence of a short leg and the knee flexion portion of the prone leg check, require additional research efforts to sort out both reliability and validity issues.

CONCLUSIONS

For the 2 clinicians in this study, the reliability of detecting the side of the short leg with knees extended was good ($\kappa = 0.65$), but reliability was fair ($\kappa = 0.22$) for determining the amount of any such LLI. Rotation of the head during the prone leg analysis (Derifield test) appears to be unreliable. There is no significant correlation between the clinician-observed side of the short leg with the patient-reported side of low back pain. Extremely high prevalence bias confounds reliability regarding the observation of LLI and the change in leg length when knees are flexed. The clinical validity of observation of LLI was not tested in this study and remains under question.

Practical Applications

- For the doctors in this study, the reliability of detecting the side of the short leg with knees extended was good, but reliability was fair for determining the amount of LLI.
- Rotation of the head during the prone leg analysis (Derifield test) appears to be unreliable.
- There was no significant correlation between the clinician-observed side of the short leg with the patient-reported side of low back pain for patients in this study.

REFERENCES


Access to Journal of Manipulative and Physiological Therapeutics Online is available for print subscribers!

Full-text access to Journal of Manipulative and Physiological Therapeutics Online is available for all print subscribers. To activate your individual online subscription, please visit Journal of Manipulative and Physiological Therapeutics Online, point your browser http://www.mosby.com/jmpt, follow the prompts to activate your online access, and follow the instructions. To activate your account, you will need your subscriber account number, which you can find on your mailing label (note: the number of digits in your subscriber account number varies from 6 to 10).

See the example below in which the subscriber account number has been circled:

Sample mailing label

This is your subscription account number

* * * * * * * * * * * * * * * * * * * * * * * * * * 3-DIGIT 001
SJ P1
AUG00 J076 C: 1234567-89 U 05/00 Q:1
J. H. DOE, MD
531 MAIN ST
CENTER CITY, NY 10001-001

Personal subscriptions to Journal of Manipulative and Physiological Therapeutics Online are for individual use only and may not be transferred. Use of Journal of Manipulative and Physiological Therapeutics Online is subject to agreement to the terms and conditions as indicated online.
MAGNETIC RESONANCE IMAGING OF THE TRIANGULAR FIBROCARTILAGE COMPLEX LESIONS: A COMPREHENSIVE CLINICORADIOLOGIC APPROACH AND REVIEW OF THE LITERATURE

Usama Albastaki, MD,a Dimitris Sophocleous, MD,a Jan Göthlin, MD, PhD,b and Claude Pierre-Jerome, MD, PhDc

ABSTRACT

Objective: This article illustrates the frequent lesions of the triangular fibrocartilage complex (TFCC) by means of magnetic resonance imaging.

Methods: We performed a retrospective chart review of the magnetic resonance images of 109 patients from our database. All subjects had history of trauma, and all underwent both radiographic and magnetic resonance imaging examination of the wrist. The changes (degeneration, tears) of the TFCC were assessed.

Results: Ten patients were excluded because of incomplete imaging protocol (4 patients) and low-quality images (6 patients). From the 99 wrists remaining, the TFCC was normal in 30 (30.3%). Degenerative changes were found in 40 (40.4%) wrists. Partial and complete tears were present in 17 (17.1%) and 12 (12.1%) wrists, respectively.

Conclusion: The TFCC lesions in acute traumatic wrists should not be overlooked; they may contribute to wrist pain and disability after treatment of existing bone injuries. (J Manipulative Physiol Ther 2007;30:522-526)

Key Indexing Terms: Wrist, Triangular Fibrocartilage; Magnetic Resonance Imaging

The triangular fibrocartilage complex (TFCC) plays a major mechanical role in the stability of the distal radioulnar joint.1 The TFCC is made of ligaments, fibrocartilage, and capsule attached on the ulnar side of the wrist joint.2 Degenerative disease and tears of the TFCC are the most common lesions. The 2 entities have distinct characteristics on magnetic resonance (MR) images. Degenerative lesions appear as an area of higher signal intensity on both T1- and T2-weighted and fat-saturated images. Tears present with lower signal intensity of the T1-weighted images compared with their correspondent T2 or fat-saturated images, where the intensity of the signal increases. This difference is due to the presence of infiltrated fluid or hematoma within the fibers of the triangular ligament. In cases of complete tears, there is a clear discontinuity of the ligament with associated high signal intensity fluid intercalated between the separated fragments.2,3

Triangular fibrocartilage complex lesions have a high incidence among the general population. By the fifth decade of life, symptomatic perforations are identified in 40% of TFCC studies. By the sixth decade, the numbers increase to 50%. The damages to the TFCC, either degenerative or posttraumatic, are common causes of pain in the ulnar side of the wrist. Clinically, TFCC lesions cause weakness in grip strength, limited work or sport activity, and tenderness over the ballottable area of the ulna. It may present with passive ulnar deviation and dorsal ulnar head subluxation.4-7

The TFCC is not visualized on conventional x-rays; however, damages to the TFCC can be associated with changes on conventional x-rays, such as ulnar variance, ulnar dislocations, cortical sclerosis, or lucent or cystic changes in the attachment points.8-10 The radiologic assessment of TFCC lesions is achieved with magnetic resonance imaging (MRI). Nonenhanced MRI is an excellent noninvasive imaging modality of choice for the examination of the TFCC, which has a bow tie appearance on the coronal images (Fig 1).

The sensitivity and specificity of MRI varies among different studies. Golimbu et al11 reported an MRI accuracy of 95% and a sensitivity of 93% compared with arthrography in detecting TFCC tears. When compared with arthroscopy and arthrotomy, the MR sensitivity reaches 89%. Potter et al12 showed in a prospective study of 77 patients that MRI
had a 100% sensitivity, 90% specificity, and 97% accuracy for detecting TFCC tears. Oneson et al\textsuperscript{13} reported that the sensitivity for detecting central degenerative perforation was 91%, and the sensitivity for detecting radial tears was 100% and 86% for 2 observers. Morley et al\textsuperscript{14} reported that MRI had a sensitivity of 44% for the detection of the TFCC injuries when wrist arthroscopy was used as the standard of reference. Despite its high sensitivity, MRI presents some limitations in the detection of peripheral TFCC tears.\textsuperscript{14}

Conventional arthrography using 3 compartment techniques previously played an important role in the diagnosis of the TFCC lesions. Shih et al\textsuperscript{15} found arthrography sensitivity in diagnosing TFCC lesion to be 83%. However, MRI arthrography has the advantage of showing the exact site of the injury because it also displays the surrounding soft tissue changes and detects other possible causes of the symptoms in case of a normal TFCC. Meier et al\textsuperscript{16} found the MR arthrography sensitivity in the diagnosis of TFCC lesions to be 94% with 89% specificity. Herold et al\textsuperscript{17} found the sensitivity and specificity of this technique in the detection of a TFCC lesion to be 100% and 77%, respectively.

In our study, we used the conventional nonenhanced MRI to detect lesions of the TFCC in patients with persistent wrist pain after trauma.

\textbf{METHODS}

The MR images of 109 patients (62 men and 47 women) were reviewed. The examinations were performed during January 2005 to March 2007. The patients’ age range was 17 to 84 years. All of them had history of trauma and were referred from the emergency unit of the orthopedic department. Ten subjects were excluded because of incomplete imaging protocol (4 patients) and low-quality images (6 patients).

The participants signed a consent form before the MR examination. All patients underwent a radiographic examination of the wrist before the MR examination. The examinations were performed with a 1.5-T magnet (Siemens MAGNETOM Harmony, Berlin, Germany) with a local cylindrical coil. The MR examination protocol consisted of coronal spin echo T1 and T2 weighted with fat saturation. Coronal short tau inversion recovery images were obtained occasionally. The imaging parameters are presented in Table 1.

All the images were retrieved from the database archive system and sent to a dedicated work station in the radiology department. The images were reviewed by 3 radiologists with experience in musculoskeletal MRI. The analysis of the images was done simultaneously, and the final decisions were reached by consensus. We searched for degenerative changes and tears of the TFCC. The lesions were classified as follows:

\begin{itemize}
  \item[I.] Degenerative
  \item[II.] Tears
    \item[IIa.] Partial tear
      \item[IIa1.] Partial tear radial side
      \item[IIa2.] Partial tear ulnar side
    \item[IIb.] Complete tear
      \item[IIb1.] Complete tear radial side
      \item[IIb2.] Complete tear ulnar side
\end{itemize}

\textbf{RESULTS}

We reviewed the MR images of 99 patients. From the 99 wrists, the TFCC was normal in 30 (30.3%) (Fig 1). Degenerative changes (Fig 2) were found in 40 (40.4%) wrists. Partial tears (Fig 3A and B) and complete tears (Fig 4A and B) were present in 17 (17.1%) and 12

\begin{table}
\centering
\caption{Different MRI series used and the different parameters of each series}
\begin{tabular}{|c|c|c|}
\hline
 & Coronal spin echo T1 & Coronal spin echo T2 with fat saturation & Coronal STIR \\
\hline
TR (ms) & 420 & 6170 & 4500 \\
TE (ms) & 14 & 106 & 77 \\
TI (ms) & – & – & 150 \\
Field of view (mm) & 120 & 120 & 120 \\
No. of signal averages & 2 & 3 & 2 \\
Slice thickness (mm) & 3 & 3 & 3 \\
No. of slices & 12 & 25 & 12 \\
Matrix & 512 × 512 & 224 × 320 & 512 × 512 \\
Flip angle & 90° & 150° & 150° \\
Examination time (min:s) & 2:45 & 4:45 & 4:45 \\
\hline
\end{tabular}
\end{table}

\textbf{Fig 1.} Coronal STIR image. The arrow points to the normal intact TFCC with diffuse high signal intensity edema in the distal radius and in the styloid process of ulna.

\textbf{Table 1.} Different MRI series used and the different parameters of each series

\begin{itemize}
  \item[I.] Degenerative
  \item[II.] Tears
    \item[IIa.] Partial tear
      \item[IIa1.] Partial tear radial side
      \item[IIa2.] Partial tear ulnar side
    \item[IIb.] Complete tear
      \item[IIb1.] Complete tear radial side
      \item[IIb2.] Complete tear ulnar side
\end{itemize}
(12.1%) wrists, respectively. The partial tears were located in the radial side in 2 wrists, and they were detected in the ulnar side in 15 wrists. The complete radial-sided tears were found in 4 patients, and the complete ulnar-sided tears were found in 8 patients.

**DISCUSSION**

The TFCC plays a specific role in the stability of the wrist joint. This complex consists of the triangular ligament, the meniscus homologue, and the ulnocarpal ligament. The triangular ligament runs from the medial aspect of the distal radius to the ulnar styloid process. The radial part of the ligament is composed of highly ordered parallel collagen fibers, which explain its hypointense appearance on MRI. The ulnar part of the triangular ligament is formed from 2 distinct laminae, the distal lamina and the proximal lamina. It changes its orientation according to the position of the forearm; it is vertical in neutral position and becomes sagittal in maximum pronation and supination. The radial part is formed from less ordered collagen bundles, being composed of vessels and fat explaining its intermediate signal intensity on MRI. The ulnomeniscus homologue is a triangular soft tissue structure with inhomogeneous intermediate signal intensity in the space between the proximal carpal row and the ulnar styloid process and the articular disk. The ulnocarpal ligament consists of the ulnotriquetral and the ulnolunate ligaments.

Biomechanically, the central portion of the TFCC absorbs 20% of the compressive load transmitted across the wrist. It represents a major stabilizer of the radioulnar joint and provides stability for the ulnocarpal joint. The extensor carpi ulnaris, along with the infratendinous extensor retinaculum and pronator quadratus muscle, represents the dynamic stabilizer of the distal radioulnar joint. The system’s main function is to resist the dorsal ulnar dislocation with full pronation and to prevent the palmar ulnar displacement.
During full supination. The extensor carpi ulnaris might also be used in the case of a chronic TFCC lesion, creating ulnocarpal tether or radioulnar tether.20,21

The mechanisms of injury in the TFCC are related to falls on a pronated hyperextended wrist, twisting with palmar rotation, and forced ulnar deviation, such as a few gymnastics actions. Triangular fibrocartilage complex lesions can also be associated with distal radius fracture and ulnar impaction syndrome.5,22 The degenerative changes might be a part of a normal aging process affecting predominantly the central avascular part of the TFCC. These changes are commonly seen among older individuals and in those with rheumatoid arthritis or osteoarthritis.22

Clinically, the patient with lesions of the TFCC complains of tenderness and a painful click during wrist motion. At a late stage of the TFCC derangement, the ulna will become more prominent during pronation because a depression is noted near the sigmoid notch with supination.

The McMurray’s test is used to detect TFCC lesions. It is based on the passive manipulation of the carpal condyles against the head of the ulna while the wrist is ulnarily deviated. This will produce a painful crepitus, roughness, or an actual snap. Another examination is the piano key test performed by compressing the distal radioulnar joint and palpating the radius and ulna. Easily depressed ulnar protrusion can be felt in a way similar to pressing a piano key.23

The differential diagnosis of TFCC lesions includes a wide range of conditions causing ulnar-sided wrist pain. They include fractures of the ulna and the ulnar carpal bones, chondral tears, osteoarthritis of pisiform-triquetrum joint, rheumatoid arthritis, osteoid osteoma, ganglia, subchondral cysts, and avascular necrosis of the lunate (Kienböck syndrome).8,24 The complications related to TFCC lesions include progressive instability of the distal radioulnar joint, osteoarthritic changes, loss of motion, and reduction of grip strength.15,25

The treatment of TFCC lesions starts conservatively with the administration of anti-inflammatory drugs (eg, nonsteroidal anti-inflammatory drugs) or splinting for more than 3 months. The surgical approach is considered according to the type of the lesion (acute or chronic) and to its location (central or peripheral). Debridement or suture of the fragments are the surgical treatments of choice for the acute injuries. These procedures can be done either arthroscopically or with open surgery.26,27

In cases of a chronic lesion, the type of treatment applied will depend on the presence of arthritis. In the absence of arthritic changes, a repair will be done similar to the treatment applied to the acute lesion. In the case of osteoarthritis, an ulnar osteotomy is done to decrease the loading of the TFCC and to prevent further degenerative changes.28,29

Our study carries some limitations. The number of cases is rather small, and the examination data were retrospectively analyzed. Also, some statistical data, such as the sensitivity and specificity of MRI, were not calculated. However, such information has been previously published from other investigators. Our aim was to assess the incidence of the damages to the TFCC in a population with acute traumatic
wrist. The TFCC lesions should not be overlooked; they can cause persistent wrist pain and disability even after the associated fracture has healed.

**CONCLUSION**

The TFCC lesions in acute traumatic wrists should not be overlooked. They may contribute to wrist pain and disability after treatment of existing bone injuries.

**REFERENCES**

ABSTRACT

Objective: Every promotion committee is challenged by the need to make value judgments on the quantity and quality of peer work. Decisions based upon subjective assessments may not do justice to the applicant’s or the institution’s needs. The purpose of this article is to (1) describe the process a college promotion committee used to increase the objectivity brought to this activity, (2) present the tools developed that facilitated the collection and evaluation of faculty work, and (3) describe their usage in a promotion cycle.

Methods: The Professor Promotion Committee met weekly for 6 months engaging in lengthy and comprehensive discourse to capture the breadth of scholastic and service activities normally engaged in by faculty.

Results: The committee’s work culminated in the development of 4 electronic applications soliciting specific evidence aligned with faculty work and 1 scoring rubric tied directly to the e-applications. More than 55 activities were identified, divided into 4 levels of accomplishment using quantitative and qualitative criteria and weighted according to their centrality to faculty work and relative importance to the institution. Each activity was assigned to one of the following categories: teaching/academic support, scholarship/research, service, and professional development. A consensus score based upon the evidence was used to generate promotion discussions.

Conclusions: The committee believes the online application aids applicants in recognizing the breadth and depth of promotable work. It provides them the opportunity to structure their work in ways that enhance their chances for promotion. The evidence-based rubric helps to reduce subjectivity in the evaluation process. (J Manipulative Physiol Ther 2007;30:527-535)

Key Indexing Terms: Faculty; Chiropractic; Education; Schools; Health Occupations
passionate discussions on who is worthy of promotion.\(^1\)

Decisions based on a purely subjective assessment of applications may not do justice to either the applicants or the needs of the institution. Since 2001, promotion at this chiropractic college has evolved from a time-in-grade quota system that was loosely coupled to faculty productivity to one that is tightly coupled to accomplishments and productivity. From that time, the Professor Promotion Committee (PPC) has struggled to achieve consensus regarding the kinds of evidence evaluated and the level of evidence that represents professorial work.\(^2-8\) Complicating the process was the yearly change in PPC membership, which required a fundamental revisiting of the issues every year with no real consensus development. With the ratification of a collective bargaining agreement (CBA) in 2005, some guidance was afforded all promotion committees in making their recommendations. In addition, PPC membership was stabilized with prescribed revolving terms lasting 3 years. With this stability, the PPC sought to increase the objectivity it brought to promotion decisions.

The aims of this article are 3-fold: (1) to describe the process a college promotion committee used to increase the objectivity brought to this activity, (2) to present the tools developed that facilitated the collection and evaluation of faculty work, and (3) to describe their usage in a promotion cycle.

**METHODS**

Beginning in July 2005, the PPC embarked on a consensus process to identify evidence that could be used to evaluate applicants and to understand, define, and communicate its own value system. The PPC decided early during its deliberations that creating a rubric would provide a means with which to explicitly and objectively identify and measure faculty activities it considered important for promotion. The PPC met weekly for 6 months engaging in lengthy and comprehensive discourse to identify and capture the breadth of scholastic and service activities normally engaged in by faculty.\(^9-11\) A laptop computer, data projector, and a Microsoft Excel (Microsoft Corp, Redmond, Wash) spreadsheet were used to communicate and record those items that a faculty member might produce as evidence of worthiness for promotion. The PPC initially familiarized itself with the provisions of the CBA as it developed promotion criteria. The PPC used previous rank promotion documents as a starting point, as well as examples of productivity demonstrated by faculty promoted since 2001. The 2004 rank promotion documents and other documents created in the 1990s by the college Faculty Promotion and Retention Committee were consulted. The PPC also referred to promotion criteria used at the University of Missouri-Columbia School of Medicine,\(^12\) one institution whose promotion criteria was available on its Web site and accessible via the internet at the time the PPC was examining its own criteria.

**RESULTS**

Deliberations identified more than 55 activities that could produce measurable evidence of faculty work. Each activity was assigned to one of the following categories: teaching/academic support, scholarship/research, service, and professional development. Activities were divided into levels of accomplishment using quantitative and qualitative criteria as appropriate. An integer rating scale ranging from 1 to 4 (1 being the lowest and 4 being the highest rating) was then assigned to each level of accomplishment. This yielded a “scoring rubric.” The PPC was mindful of capturing the breadth of faculty accomplishments by creating scoring rubrics with wide-ranging activities.

Within each category, activities were weighted according to their relative importance to the institution. Weightings were also assigned using a 4-point integer scale. Activities within a category were weighted higher if the PPC deemed the item more important to the institution in terms of promotional worthiness. A lower weighting was assigned to those activities not deemed as central to faculty work or of a lesser magnitude in required effort. Each PPC member declared his/her ranking and an average score per activity was calculated from the 7-member PPC scores. Multiplier weightings were arbitrarily assigned to each range of average scores according to the following scale: 1.00 = 1.0-1.75, 1.25 = 1.76-2.50, 1.50 = 2.51-3.25, and 1.75 = 3.26-4.0. A formula was developed in the Microsoft Excel spreadsheet to automatically calculate the weighted score for each activity by multiplying the PPC’s evidence-based consensus score (see below) for an applicant’s accomplishments by the assigned weighting factor.

Four scoring rubrics were created reflecting the organizational domains of the college’s faculty: classroom teaching, clinic, research, and academic support. The PPC ranked each category and assigned an additional numerical weight to it, depending upon the faculty member’s domain. For the domains of classroom teaching, clinic, and academic support, the categories were weighted accordingly: 50% for primary responsibility, 20% for scholarship, 20% for service, and 10% for professional development. For the research domain, the categories of scholarship and teaching were weighted 50% and 20%, respectively. Across domains, care was taken to assure equal weighting for activities common to all 4 scoring rubrics, for example, chairing a departmental committee for clinic faculty was weighted equally to chairing a departmental committee for research faculty.

Because the rubrics were so wide ranging in activities, it was obvious that no individual faculty member would
provide evidence of engaging in all activities. As a result, the PPC found it more important to evaluate the percentage of maximum points attained within a category rather than a simple raw score. Formulae were created in the Microsoft Excel spreadsheet to calculate percentage scoring within categories. The percentage of points captured both quantitative and qualitative components and allowed the PPC to determine the overall level of productivity in any category. It enabled the PPC to assess the degree to which a faculty member presented a balanced portfolio of productivity across all categories or significant evidence of high levels of productivity in individual categories with less evidence in other areas. In the future, benchmarks will be determined based upon the profiles of percentage scoring in categories demonstrated by those already promoted under this new system.

Four distinct yet similar electronic applications (e-applications), one for each college domain, were created by which faculty could provide, in a point-by-point fashion, the evidence for each activity identified on the relevant scoring rubric. The final page of the e-application provided faculty with space to include any evidence not specifically solicited by the application. E-applications were designed and programmed with Active Server Pages 3.0 via Microsoft FrontPage (Microsoft Corp). Electronic submissions were stored in a Microsoft SQL Server 2000 database. SQL Server 2000 provided secure data storage by enabling storage on a server separate from the web server. In addition, SQL Server 2000 could handle a large number of users and thousands of variables. The web programmer could provide specific users (eg, applicants, PPC members or chairs, and administrators) access to defined portions of the database.

The rubric for the teaching domain was developed first and submitted electronically to all full professors for critical review and feedback using a survey instrument developed by the PPC (See Appendix A). Of the 23 surveys sent to the full professors, 6 were returned. Although the sample size was small, strongest agreement was with the statements, “It is important for the Professor Committee to weigh evidence” (a mean of 4.33 on a 5 point scale) and “A rubric score may be helpful in discerning the quantity and quality of work” (a mean of 4.17 on a 5-point scale). Strongest disagreement was with the statement that “Each major heading is weighted appropriately” (a mean of 3.00 on a 5 point scale), followed by disagreement on the need to benchmark rubric data (a mean of 3.17 on a 5 point scale). Based upon this feedback, the weighting of categories tied to one’s primary assignment was increased from 45% to 50%, and the weighting for the category of scholarship was decreased from 25% to 20%. The survey instrument also allowed faculty to submit other activities in their area of primary assignment they felt should be considered as evidence for promotion. A less formal process was used to solicit feedback from clinic, research, and academic support faculty in developing their corresponding rubrics.

The PPC’s work culminated in February 2006 with the development of two evidence-based instruments—(1) an electronic application (e-application) for each one of the domains of faculty assignment soliciting specific evidence aligned with faculty work and (2) a weighted scoring rubric tied directly to each e-application to help the PPC evaluate the evidence. Each pair of instruments was written specifically to capture a faculty member’s work in his/her designation within the college—classroom teaching, clinic, academic support, and research. Appendix B contains short examples from the e-application and rubric for classroom teaching faculty. Full e-applications and corresponding rubrics can be found in portable document format at www.palmer.edu/promotionrubrics. Faculty members with primary administrative assignments were also eligible to apply for promotion; however, they were evaluated based upon faculty productivity criteria for their domain and not their administrative work.

In an effort to provide as much information as possible about the process and to promote transparency, the PPC held 2 meetings with all eligible faculty interested in applying for promotion to professor. At the meetings, the PPC shared the instructions for filling out the e-applications and distributed copies of the scoring rubrics. The PPC chair also met individually with faculty who had not been able to attend either meeting. An e-application was accessible by password and remained “live” and editable up until the time it was submitted for final review. Once an application was submitted, it was electronically locked and could not be reedited unless unlocked by the PPC chair.

Seven applicants used the newly devised e-applications to apply for promotion. Of the seven applicants, 5 were from a single domain. After the March deadline for receipt of applications had passed, the PPC sent a survey to each of the 7 applicants soliciting feedback on the e-application process (Appendix C). Four responses were received. Strongest agreement was with the statement, “The instructions were helpful” (a mean of 5.0 on a 5-point scale). Strongest disagreement was with the statement, “My primary assignment is adequately represented” (a mean of 3.75 on the 5-point scale). Several applicants had primary administrative assignments, which were not part of the promotion criteria. This may have affected the response mean of the last statement.

From April 2006 through June 2006, the PPC evaluated the 7 applications for professor. Using the e-application in conjunction with the supportive documentation required of each candidate, each PPC member separately reviewed the candidate’s documentation and individually scored the rubric pertaining to the faculty member’s locus of appointment. Approximately 1 hour was needed to review
each application along with its supporting evidence and score the rubric.

After individual PPC member’s work was completed, the PPC met to arrive at a consensus score for each applicant. Activity by activity, the chair solicited ratings from the members; each member was given opportunity to present the rationale with evidence for his/her scores. This became especially important when disparate ratings were given and resulted in a clarification of the criteria used to evaluate the evidence. After discussions, a simple majority vote determined the consensus rating. Consensus ratings were reached for each activity, a process that took approximately 1 hour per candidate. A consensus score for each category and for the entire application was then calculated for each candidate. A value-added outcome of this discursive process was that discussions were appropriately focused on evidence and outcomes rather than applicants during this critical review phase.

To summarize the results of its work, the PPC generated 5 histograms (in Microsoft Excel) for each candidate based upon the major categories of faculty work—primary assignment, service, scholarship, and development—and a composite histogram summing all categories with appropriate weightings. Candidates were grouped according to their domains of responsibility and were evaluated by comparing the histograms within domains. The PPC was careful not to compare productivity across domains (eg, teaching vs research faculty). The PPC considered overall productivity; balance among the various categories of faculty work; distinguished work in 1 or more categories, and specific accomplishments which, for classroom, academic support, and clinic faculty, included a degree of scholarship defined in the CBA as “a sustained record of significant contributions in advancing their area of standard activities and additionally advancing the knowledge and skills of others through scholarly activities.” For research faculty, a higher standard of scholarly output was required with some expectation of mentoring/teaching junior researchers. Candidates whose work showed a balance in productivity, or showed a record of sustained scholarship, or who distinguished themselves significantly in a single category were recommended for promotion.

After arriving at a promotion recommendation, the PPC corresponded with each applicant, informing him/her of its recommendation and the reasons for the recommendation. The chair of the PPC then met with the Vice President for Academic Affairs, explained the PPC’s recommendations, and delivered the consensus rubrics, the e-applications, and the evidence packets. The Vice President for Academic Affairs reviewed the PPC’s recommendations and sent them to the President, who made the final promotion decision.

After the PPC had sent its recommendations to the Vice President for Academic Affairs, the PPC surveyed its members soliciting feedback on the e-applications and use of the rubrics (Appendix D). Of seven possible responses, 6 were received. Strongest agreement was with the statements: “Information from an applicant’s e-application can be transferred logically and efficiently to the rubric”; “The e-application represents the scope of work expected of a faculty member at the Professor rank”; and “The e-application process increases objectivity in the evaluation of candidates for promotion” (a mean of 5.0 on a 5-point scale for each). Strongest disagreement was with the statement, “The e-application is comprehensive enough to capture the breadth and variety of work done by our faculty” (a mean of 4.3 on a 5-point scale).

### DISCUSSION

Developing a new evidence-based method to evaluate faculty productivity for promotion had its challenges. The 7-member PPC was coappointed by the faculty senate and the administration. All members, whether faculty or administrator, hold faculty rank. The goal was to formulate a group that could take into account interests and concerns of both the rank and file faculty and the academic administration in its recommendations. In the past, although the PPC might recommend a faculty member for promotion, the administration denied certain promotions based on insufficient evidence. The membership of the current PPC had worked together for several promotion cycles before embarking on this new project, which had sensitized them to the dynamics of the working group and provided insight into the needs for the tasks ahead.

An initial (and ongoing) challenge was the role and impact of scholarship in promotion decisions. Although scholarship is not required of the faculty within the workload conditions of the CBA, a “sustained record of scholarship” is required for promotion to professor. The college adopted the following definition of scholarship: “Scholarship and creative activity are understood to be intellectual works the significance of which is validated by peers and which is communicated. The principle of peer review and recognition becomes increasingly important as the (CBA) member progresses through the academic ranks. Scholarship emphasizes project-oriented behavior that results in a measurable product or outcome (eg, a publication, written report, manual, or protocol).”

The Boyer model of scholarship was adopted including the scholarship of discovery, the scholarship of integration, the scholarship of application, and the scholarship of teaching. The PPC considered all evidence produced within Boyer’s scholarship definition as acceptable for promotion consideration. Other qualifiers were placed on scholarship, such as a rationale for how the activity or...
behavior supports the mission of the college, documented evidence that the activity was performed and completed, and criteria that establishes the outcome or result met an acceptable standard of quality.

Another challenge was the response rate on surveys of the current full professors. The PPC felt it needed input from the most objective group of faculty regarding promotion. Full professors had no personal risk in helping determine the appropriate criteria for promotion to their rank as they would never be judged by them. Unfortunately, response rates were meager. However, because the PPC was broad in its consideration of what constituted faculty work and the types of evidence representing that work, there was little dissent among PPC members, rank and file faculty members, and administration concerning the content of the rubrics.

The PPC recognized that even the most evidence-laden application requires a subjective recommendation for promotion by a committee. Concern was expressed by several faculty members about the use of scoring rubrics and the reduction of faculty productivity to a number. Raw numbers were not used as a basis for promotion discussions; holistic consideration of the histograms capturing either overall productivity or patterns of productivity was the focus of discussions. The e-application and accompanying rubric infused more objectivity and resulted in deeper investigation and discussion by all PPC members in making such consensus recommendations. The PPC recognizes the need to standardize the instrument over time if it is to be used to set benchmark scores.

Because the criteria and value for evidence had been identified before any candidate applying for promotion, the PPC found it easier to make objective evaluations based on the evidence presented by each candidate. The PPC used data from the rubrics to compose the letters to unsuccessful candidates, giving them feedback on where they might strengthen a future application without being prescriptive.

Areas of the new promotion process needing improvement were identified: (1) some areas of the e-application did not correspond with the rubric and the PPC modified the e-application accordingly; (2) in other instances, the rubric was modified when evidence on the e-application was not considered in the rubric; and (3) the PPC also identified areas where application instructions were inadequate and made appropriate adjustments for future promotion cycles. At each meeting during the evaluation process, areas needing improvement were transcribed by the chairperson. At the end of the promotion cycle, the e-application and rubric will be updated for the following year. The rubrics remain a dynamic tool for gathering and weighing evidence. They can be adjusted to promote greater scoring accuracy or to include unanticipated evidence.

A strength of the e-application is its use as an evidence-based portfolio for faculty work over time that can be submitted at some future date for promotion. An application can be securely saved electronically and edited or appended daily until final submission is required. Unsuccessful applicants can have their application information “released” for future amending, thereby getting a head start on an application for the next promotion cycle. In addition, by identifying the evidence necessary to demonstrate faculty accomplishment, the PPC hopes to influence the professional career paths of junior faculty members as they aspire to be promoted. The college and the chiropractic profession will benefit greatly from such levels of productivity.

In anticipation of the associate professor committee adopting the process and instruments, the e-application has been designed for associate professor as well as for professor applicants. In the e-application, applicants simply check off which committee should be reviewing their submission.

A significant outcome of the process was the trust developed among the members of the PPC in its deliberations. Over a year, the mix of academically productive faculty and administrators proved effective in equitably identifying and weighing appropriate evidence for promotion. Through the process, PPC members found it easier to adjust their opinions and recommendations to reach consensus. An unknown factor and potential challenge for the future is the inevitability of incorporating new PPC members into a structure they did not directly develop. However, because some may have experienced the evidence-based promotion process, they will be familiar with the procedures and expectations on faculty.

Conclusion

The scoring rubric, the e-application tools, and the process through which they were applied helped quantify and clarify the strengths and weaknesses of promotion applications for the rank of professor and infused confidence in the PPC’s recommendations for promotion. Although nothing can take the sting out of not being promoted, a response from the PPC delineating how the applicant was scored and suggesting ways for strengthening a future application should give unsuccessful applicants for promotion an opportunity to focus on activities that will increase their chance for advancement in rank. The PPC has created tools that should help them evaluate all applicants fairly and consistently, provide specific feedback, and shape the future activities of faculty aspiring to reach the rank of professor.

Acknowledgment

The authors wish to acknowledge the technical assistance of Lynn Carber, who designed and programmed the online application using Active Server Pages 3.0.
**APPENDIX A. SURVEY OF PROFESSORS**

**Professor Rubric Survey**

<table>
<thead>
<tr>
<th>The Professor Committee would like your feedback on the rubric developed to evaluate professorial candidates whose primary assignment is teaching. Please indicate the extent to which you agree with each of the following statements by placing an “X” in the appropriate circle (do not write your name anywhere on this form):</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Neither Agree nor Disagree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. It is important for the Professor Committee to weigh evidence of faculty productivity for promotion.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☑</td>
</tr>
<tr>
<td>2. A rubric is a good way to capture evidence of faculty productivity.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☑</td>
</tr>
<tr>
<td>3. A rubric score may be helpful in discerning the quantity and quality of faculty work.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☑</td>
</tr>
<tr>
<td>4. The major headings of Teaching, Research, Service and Professional Development are categories that reflect the full range of faculty productivity toward promotion.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☑</td>
</tr>
<tr>
<td>5. Each major heading is weighted appropriately.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☑</td>
</tr>
<tr>
<td>6. Overall, the category items are weighted appropriately.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☑</td>
</tr>
<tr>
<td>7. Over time, the Professor Committee should benchmark summary rubric scores to determine a recommendation for promotion.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☑</td>
</tr>
</tbody>
</table>

 Comments:
1. Within the major headings, are there any significant activities/categories that were overlooked in this rubric (for Teaching faculty)?
2. General Comments on the Rubric.
APPENDIX B. SAMPLE FROM E-APPLICATION FOR TEACHING

F. **Student Advising** - Identify formal advising activities you have been involved in, i.e., graduate students, special schedule DC students.

<table>
<thead>
<tr>
<th>Academic Advising Activity</th>
<th>Dates</th>
<th>Number of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Evidence:
Verify your formal advising activities through the office that assigns students to you.

G. **Magnitude of Teaching** - Identify the courses for which you have been a lead instructor over the rating period. Include graduate, undergraduate, and Special Programs courses.

<table>
<thead>
<tr>
<th>Course Name</th>
<th>Dates</th>
<th>Credit Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

H. **Learning Outcomes Assessment Beyond Course Grades** - Identify assessment of student learning activities beyond that of simply grading students, i.e., course level assessment.

<table>
<thead>
<tr>
<th>Assessment Activity</th>
<th>Dates</th>
<th>Assessment Cycle: Instructional Change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>&lt; SELECT &gt;</td>
</tr>
</tbody>
</table>

Evidence:
Provide evidence of engagement in the assessment cycle, i.e. identifying course outcomes, measuring the outcomes, interpreting the data, and making changes based upon the data.

---

**Corresponding Sample from teaching Rubric**

<table>
<thead>
<tr>
<th>Teaching</th>
<th>WT</th>
<th>SCORE</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 F. Student Advising</td>
<td>1.25</td>
<td>3</td>
<td>3.75</td>
</tr>
<tr>
<td>1 Formally Advises 1-3 DCP students</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Formally advises 4-6 DCP students; or thesis advisor to 1 grad student</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Formally advises 7-9 DCP students; or #1 + thesis advisor to 2 grad students</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Formally advises 10 or more DCP students; or #2 + thesis advisor to &gt;2 grad students</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| 7 G. Magnitude of Teaching | 1.59 | 2 | 3 |
| 1 Lead of 2 different courses | | | |
| 2 Lead of 3 different courses | | | |
| 3 Lead of 4 different courses | | | |
| 4 Lead of >4 different courses | | | |

| 8 H. Learning Outcomes Assessment Beyond Course Grades | 1.75 | 2 | 3.5 |
| 1 Evidence of data collection on students learning | | | |
| 2 Evidence of data interpretation with planned instructional changes | | | |
| 3 Evidence of instructional changes from data collected and interpreted | | | |
| 4 Evidence of a pattern of data collection, interpretation and instructional changes | | | |
APPENDIX C. E-APPLICATION SURVEY

Electronic Application for Promotion Survey

This survey is intended to solicit feedback on the electronic application for the rank of professor from members of the faculty who applied for promotion in 2006. There is a section for Comments at the bottom of the survey form. Please fill out the following survey and mail it to Glenda Wiese, Chair of the Professor Promotion Committee.

Please indicate your level of agreement with the following statements relative to the e-application for rank promotion to professor.

<table>
<thead>
<tr>
<th>Statement</th>
<th>SA</th>
<th>A</th>
<th>N</th>
<th>D</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>The instructions for filling out the e-application are helpful.</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
</tr>
<tr>
<td>The length of time to complete an e-application is reasonable.</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
</tr>
<tr>
<td>The format of the e-application is a helpful guide in determining the type of evidence expected by the Professor Promotion Committee.</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
</tr>
<tr>
<td>The e-application represents the scope of work expected of a professor.</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
</tr>
<tr>
<td>The e-application is comprehensive enough to capture the breadth of my work.</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
</tr>
<tr>
<td>My primary assignment is adequately represented by the items in the e-application.</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
</tr>
<tr>
<td>The e-application process is superior to my previous experiences with promotion processes.</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
</tr>
<tr>
<td>Overall, I am satisfied with the e-application process.</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
</tr>
</tbody>
</table>

Comments:

APPENDIX D

Promotion Committee Survey on the Use of the e-Application Process and Rubric for Rank Promotion

This survey is intended to solicit feedback from members of the Professor Promotion Committee on the use of the electronic application process and rubric in evaluating applicants for promotion to the rank of Professor. There is a section for Comments at the bottom of the survey form. Please fill out the following survey and mail it to Glenda Wiese, Chair of the Professor Promotion Committee.

Please indicate your level of agreement with the following statements relative to the e-application and rubric for evaluating candidates for rank promotion to Professor.

<table>
<thead>
<tr>
<th>Statement</th>
<th>SA</th>
<th>A</th>
<th>N</th>
<th>D</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information from an applicant’s e-application can be transferred logically and efficiently to the rubric.</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
</tr>
<tr>
<td>The length of time necessary to score evidence from an applicant’s e-application on the rubric is reasonable.</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
</tr>
<tr>
<td>The format of the e-application is a helpful guide in determining if an applicant has presented the type of evidence expected by the Professor Promotion Committee.</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
</tr>
<tr>
<td>The e-application represents the scope of work expected of a faculty member at the Professor rank.</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
</tr>
<tr>
<td>The e-application is comprehensive enough to capture the breadth and variety of work done by our faculty.</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
</tr>
<tr>
<td>The e-application process increases objectivity in the evaluation of candidates for promotion.</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
</tr>
<tr>
<td>The e-application process is superior to my previous experiences with promotion processes.</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
</tr>
<tr>
<td>Overall, I am satisfied with the e-application process.</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
<td>◐</td>
</tr>
</tbody>
</table>

Comments:
REFERENCES

CASE REPORTS

ACUTE INTRACRANIAL SUBDURAL HEMATOMA AFTER EPIDURAL STEROID INJECTION: A CASE REPORT

Ozgur Ozdemir, MD,a Tarkan Calisaneller, MD,b Erkan Yildirim, MD,c and Nur Altinors, MDa

ABSTRACT

Objective: Conservative treatment of lumbar radiculopathy includes bed rest, oral medications, physical therapy, spinal manipulation, mobilization, and epidural steroid injections. Intracranial subdural hematoma after accidental dural puncture is a rare and life-threatening complication of epidural steroid injections. In this report, we present a case of subacute intracranial subdural hematoma that developed after epidural steroid injection.

Clinical Features: A 40-year-old man was admitted to our clinic with severe persistent headache and vomiting for 2 days after epidural steroid injection for right leg pain.

Intervention and Outcome: The patient was hospitalized for epidural steroid injection for right leg pain in our pain clinic and was discharged the same day. Twenty-four hours later, he started having a headache. Despite the use of oral analgesics, his headache worsened, and he began to vomit particularly in the upright position. Magnetic resonance imaging of the brain displayed a right frontal subdural hematoma. The headache was relieved after strict bed rest, intravenous hydration, and analgesics. The patient was discharged with full recovery after 1 week.

Conclusion: Intracranial subdural hematoma after accidental dural puncture during epidural steroid injection is a rare complication. Persistent headache should be evaluated carefully for possible intracranial hematomas. (J Manipulative Physiol Ther 2007;30:536-538)

Key Indexing Terms: Hematoma; Subdural; Intracranial; Injections; Epidural; Steroids

Lumbar radiculopathy is a common problem, and only 10% to 15% of patients require surgery. Most patients get benefit from conservative treatment. Bed rest, oral medications, physical therapy, spinal manipulation, mobilization, and epidural steroid injection (ESI) are widely used as conservative treatments. Although ESIs are usually accepted as safe, complications related to medications (steroids and/or local anesthetics) or accidental dural punctures are reported rarely. Among these complications, postdural puncture headache (PDPH) is an infrequent but well-known complication of epidural anesthesia. On the other hand, intracranial subdural hematoma after accidental dural punctures after attempted epidural anesthesia is seen even more rarely. To our knowledge, this is the first report of an intracranial subdural hematoma after a lumbar ESI. Our aim is to demonstrate the clinical importance of persistent headache after ESI and to emphasize the possibility of intracranial subdural hematoma as a cause of headache.

CASE REPORT

A 40-year-old man was admitted to our hospital with severe headache and vomiting. Results of his neurologic examination were normal. Assuming an upright position significantly worsened the headache. His medical history was unremarkable except that he was hospitalized for a same-day ESI for right leg pain in our pain clinic and was discharged 3 days before admission to hospital for headache. Twenty-four hours after injection is when the headache began. This headache worsened despite the use of oral analgesics. He began to vomit particularly when he was in the upright position.

From his medical records, we noticed that the first attempted epidural insertion of the 18-gauge Tuohy needle (using loss of resistance technique) at the L4-5 level failed owing to dural puncture. The second attempt at the L3-4 level was successful, and 60 mg of triamcinolone diacetate (Bristol Myers Squibb Co, Istanbul, Turkey) was injected.
The patient’s magnetic resonance imaging (MRI) of the brain displayed a right frontal subdural hematoma (Fig 1). His headache was relieved after strict bed rest, intravenous hydration, and analgesics. The patient was discharged with full recovery after 1 week. He was completely free of symptoms during follow-up visit, and cranial MRI showed resolution of the hematoma (Fig 2).

**DISCUSSION**

Epidural steroid injection has an important place in the conservative treatment of lumbosacral radiculopathy resulting from lumbar disk herniation and/or stenosis. Complications of this procedure are rare and reported at a rate of 9.6% by Botwin et al. Nevertheless, major complications including intracranial subdural air, spinal subdural hematoma, and cauda equina syndrome after ESI are reported as isolated cases in the literature. In their recent review of the literature, Abdi et al did not report any cases of intracranial subdural hematoma, and to the best of our knowledge, this is the first report of an intracranial subdural hematoma after ESI.

The proposed pathomechanism for an intracranial subdural hematoma is based on the leakage of cerebrospinal fluid from a punctured dura. This leakage results in intracranial hypotension and subsequent PDPH, mostly after spinal anesthesia. However, PDPH could also be seen after accidental dural puncture during attempted epidural anesthesia. In addition, this intracranial hypotension may cause downward displacement of the brain, traction, and tear of the bridging veins, resulting in intracranial hematoma consequently. Yamashima and Friede described that the thinnest parts of a bridging vein’s walls are in the subdural space, and the thickest are in the subarachnoid portion. This finding suggests that bridging veins are more fragile in the subdural portion, resulting in rupture and bleeding in the subdural space when veins are tractioned. However, there could be other predisposing factors in the development of subdural hematoma such as Valsalva maneuver, brain atrophy, minor trauma, or bleeding disorders. Because PDPH worsens when the patient is upright and improves when recumbent, persistent headache despite bed rest or medication should suggest the possibility of other intracranial abnormalities. Subdural hematomas could also be differentiated from PDPH if there are accompanying focal neurologic deficits, which must be evaluated by computerized tomography or MRI. Treatment of PDPH typically includes bed rest, hydration, and analgesics; but epidural blood patch may be used to prevent further cerebrospinal fluid leakage in patients with persistent headache beyond 24 hours. On the other hand, patients with subdural hematoma require consultation with a neurosurgeon and should be followed for evidence of progressive neurologic deficits. Asymptomatic patients may be treated conservatively, and previous reports suggest subdural hematomas thinner than 5 mm often resolve spontaneously. However, patients with focal neurologic deficits, loss of consciousness, or...
subdural hematoma with midline shift should be treated surgically.\textsuperscript{3,10,11}

**Conclusion**

Although epidural steroids appear to be safe and effective in the conservative treatment of lumbar radiculopathy, intracranial subdural hematomas may occur as a rare complication of ESI. Patients should be followed for the development of subsequent complications in case of accidental dural puncture. Persistent and intractable headache that does not resolve with recumbency could indicate intracranial subdural hematoma, and patients should be considered for cranial imaging.

**Practical Applications**

- Epidural steroid injection has an important place in the conservative medical treatment of lumbar radiculopathy.
- Complications of this procedure are rare. Nevertheless, intracranial subdural hematoma may develop after ESI owing to accidental dural puncture.
- Persistent headache after this procedure should be evaluated for possible intracranial hematoma.

**References**

A CASE OF A POTENTIAL MANIPULATION RESPONDER WHOSE BACK PAIN RESOLVED WITH FLEXION EXERCISES

Stephen May, MSc, and Richard Rosedale, PT, Dip MDT

ABSTRACT

Objective: Researchers have begun to investigate the value of subgrouping patients with back pain to improve clinical outcomes; one method is the development of clinical prediction rules. To be of clinical value, it is important that subgroups identify distinct categories of patients with an associated optimal treatment. This case study raises the suggestion that subgroups identified in this way may not represent distinct categories.

Clinical Features: A patient with sudden-onset back pain, who had 4 of 5 criteria for a clinical prediction rule said to identify responders to manipulation, was successfully treated using repeated flexion in lying exercises.

Outcomes: Pain numeric score and Roland-Morris Disability Questionnaire were used to measure changes in pain and function. Pain score changed from 9/10 to 0/10 and disability score from 19/24 to 0/24 after 1 week and at 1 and 6 months of follow-up.

Conclusion: We have presented a case study that was positive for 4 of 5 items of the clinical prediction rule for manipulation responders, but this patient was successfully treated with flexion exercises. The clinical prediction rule may not represent a discrete subgroup but may include patients who can be effectively managed in other ways. (J Manipulative Physiol Ther 2007;30:539-542)

Key Indexing Terms: Back Pain; Manipulation; Spinal; Exercise Therapy

It has been proposed that the reason for the limited effectiveness demonstrated to date by intervention studies for low back pain has been the failure to identify subgroups most likely to respond to particular interventions. Not surprisingly, there has been continuing interest in determining how best to classify patients with back pain and, furthermore, to determine if classification results in better treatment outcomes. One research group has developed and tested clinical prediction rules (CPRs) based on minimal clinical characteristics, which are purported to define those patients most likely to respond to manipulation or stabilization exercises as opposed to another plan of intervention. In addition to the manipulation and stabilization exercises subgroups, there is also a subgroup within this system for specific exercises. A CPR has not been developed for this group who respond to direction-specific or directional preference (DP) exercises, but it comprises those patients who demonstrate centralization of symptoms in response to repeated movements during evaluation. The DP exercise group is composed of patients who demonstrate centralization, abolition, or decrease in symptoms or increase in range of movement in response to specific repeated movements during evaluation. Centralization is the lasting abolition of distal symptoms or back pain in response to repeated movements.

The CPR for responders to manipulation has the following clinical characteristics: duration of current episode of back pain of less than 16 days; no symptoms distal to the knee; Fear-Avoidance Beliefs Questionnaire work subscale score of less than 19 points; at least 1 segment of lumbar spine classified as hypomobile; at least 1 hip with more than 35° of medial rotation. The intention of this CPR is to aid clinicians in deciding which individual patients should receive manipulation. Further work produced a slimmed down version of the CPR, which only contained the items of duration and minimal referred pain. Although it has been directly stated that CPRs are not meant to replace clinical judgement and should be used to complement clinical reasoning, it may also be argued that the use of CPRs minimizes the clinical reasoning process, reducing decision making to a “tick-box” activity to bypass more complex and high-level reasoning that is often required in clinical practice.

The value of any subgroup is its need for a specific and hopefully optimal treatment; so, to be of clinical utility, they should represent distinct and nonoverlapping categories. George et al claimed that there will be “zero possibility” of centralization in the categories other than the DP subgroup, such as the manipulation subgroup. However, if there was a degree of overlap between the manipulation subgroup and...
the DP group, then the CPR may not be so useful at identifying patients who will respond optimally to manipulation. Although such patients may respond to manipulation, they may equally respond to DP exercises; so, evaluation for DP responses may be as important as the CPR criteria. To illustrate this point, we would like to present the following case study.

**Case Report**

Informed consent was obtained from the patient to publish her case in a medical journal, and anonymity was maintained. The patient, a 40-year-old woman, worked in administration and thus spent most of her time sitting. During her leisure time, she spent more time standing and walking than sitting. She presented with a 1-day history of asymmetric back pain, extending to her buttock crease on her left, and low lumbar pain on her right. She described the onset as occurring during performance of the “bow” in yoga, which is active extension from prone lying. Severity and functional disability was high, but fear avoidance was moderate or low. On a numeric pain scale, she rated her pain as 9/10. She scored 19/24 on the Roland-Morris Disability Questionnaire, 15 14/24 on the Fear-Avoidance Beliefs Questionnaire, 14 14/24 on the Roland-Morris Disability Questionnaire—physical activity subscale, and 0/42 on the Fear-Avoidance Beliefs Questionnaire—work subscale. 15 At the time of her evaluation, she was only able to sit for 2 to 3 minutes. She was not absent from work but minimized sitting time and adapted her normal tasks. She noted the pain was constant except when lying down. The pain was worse when attempting bending, when sitting, when rising from sitting, and when still. The pain improved when she was standing, lying down, or being on the move. Walking and time of day did not influence her symptoms.

The patient reported only 1 previous episode of back pain 7 years before, which had resolved within a month without recurrence until the previous day. She had received no treatment and was in good health apart from the back pain. She was taking a muscle relaxer and nonsteroidal anti-inflammatory medication for the back pain, with little effect.

The patient’s sitting posture was good, with a well-maintained lordosis. A neurologic examination was not deemed necessary because she had only back pain. On examination of her range of movement, it was estimated that there was a major loss of flexion, a moderate loss of extension, and minor losses of left and right side gliding. There was back pain with all single movements, and as she went into forward flexion there was no loss of lordosis. Her hip range of medial rotation was 72° left and 70° right. She was then examined using repeated movements. With repeated extension in standing, there was an increase in her baseline pain, which was worse afterward but had no effect on range of movement. Repeated extension in lying also produced her symptoms, which remained worse afterward and caused a dramatic decrease in lumbar flexion range of movement. In supine lying, she had no symptoms at rest; flexion from supine lying reproduced her back pain. This exercise was repeated with increasing patient overpressure from her hands around her knees. It got easier with repetition and eventually became pain free, and afterward her flexion range was fully restored. Only 1 set of 10 to 15 repetitions was performed. As a mechanically determined DP had been demonstrated with the response to repeated movements, no further examination was deemed necessary.

According to mechanical diagnosis and therapy classifications of McKenzie and May, 16 the patient was categorized as having a derangement, with a DP for flexion exercises. A derangement is characterized by the following findings: in response to repeated movements, pain is centralized, abolished, or decreased with the change in pain location, or decrease or abolition of pain maintained and accompanied or preceded by improvements in the mechanical presentation (range of movement and/or deformity). 16 A DP for flexion exercises describes a patient who is responding in one of these ways to self-mobilization flexion exercises. The patient was advised to repeat 10 flexion exercises while in the supine lying position every 1 to 2 hours, to maintain a neutral posture when sitting and standing, and to avoid positions of extension. For these exercises, from a crook lying posture she would bring her knees to her chest and then apply overpressure with her arms around her knees, with more overpressure applied with each repetition.

The patient returned for review 4 days later; she reported compliance with the exercises. She said she was in much less pain and was moving better. When asked, “[O]n a scale from 0 to 100%, if you had 100% of pain on your initial visit what would it be now?” Daphne answered “15%.” On examination, there was no loss of flexion or extension movement, and during flexion there was a full reversal of the lordosis. With repeated flexion in lying movements, there was no untoward or negative effect, with no reproduction of her symptoms. It was felt that the classification and management strategy were confirmed. She was advised to reduce the frequency of exercise repetition and to do flexion in standing exercises in addition to flexion in lying.

Daphne was reviewed 2 days later. She stated that she was virtually pain free, only experiencing an occasional transient twinge, and was fully functional again. On examination all movements were full range, and none produced her symptoms. She was discharged with advice to maintain a balance between flexion and extension activities, to resume all previous activities, and to change her sitting posture regularly.

At 1-month follow-up, she reported 0/10 on the numeric scale and 0/24 on Roland-Morris Disability Questionnaire. She had resumed all her normal activities, including yoga. She was reviewed at 6 months and again reported 0/10 on the numeric scale and 0/24 on Roland-Morris Disability Questionnaire. Occasionally, over these 6 months, she had experienced a minor episode of pain when doing something “wrong,” such as a movement with
too much extension, but she had been able to completely abolish symptoms independently with “a few” repeated flexion in lying exercises.

**Discussion**

A patient with recent onset of severe back pain associated with high disability responded rapidly to flexion DP exercises. High initial pain severity, functional disability, and restrictions in spinal mobility have been associated with poor long-term outcome. Thus, the existence of DP may be another significant factor that directs the clinician’s treatment despite the presence of otherwise poor prognostic factors or, as in this case, the presence of CPR criteria, which may point toward manipulation as a “most likely” treatment option.

The patient presented had 4 of 5 of the clinical items included in the CPR and both items from the revised version. Symptom duration and range of hip rotation had the highest positive likelihood ratios. The presence of 4 or more of these 5 items was said to raise the chance of success with manipulation from 45% to 95%. According to the slimmed down criteria in the presence of the 2 items, treatment success with manipulation was said to be 85%. In this case study, however, this patient received self-management flexion exercises, which not only resolved her symptoms but also helped her to deal with a few brief recurrences. If this patient had been treated with manipulation by itself, as she appeared to fit the CPR criteria, this management may have failed to provide the patient with her own ability to control and abolish her symptoms when they returned.

Obviously, we cannot exclude the possibility that the patient would have responded equally or even better to manipulation or that additional independent exercises may have been prescribed in addition to manipulation. Her response to DP flexion exercises does not equate to a failure of the manipulation CPR; however, treatment using DP exercises was clearly an alternative option that had the additional benefit of allowing the patient to take control of her symptoms and self-manage a future episode. It may be suggested that the clinically based symptom and mechanical responses are perhaps as useful determinants of management strategies than a set of clinical criteria selected by statistical analysis.

It should be recognized that the patient had very brief symptom duration of 1 day only, whereas the CPR criteria stipulated back pain of less than 16 days. Thus, limitations in the comparability of case study and CPR criteria should be recognized. Furthermore, given the short-term nature of the patient’s symptoms, the positive prognosis of short-duration back pain should also be born in mind. Although the DP exercises appeared to work, and manipulation may also have worked, the pain may have spontaneously resolved given its short-term nature.

One case study is clearly not enough to overthrow a carefully constructed CPR that has attained level 2 status on the levels of validation for CPRs of McGinn et al. However, doubts about certain methodological aspects of its development may be raised. It is recommended that reliability for examination procedures used in CPR attain $k$ values of greater than 0.60, whereas most of the items in this rule did not. Furthermore, it is recommended that at least 10 cases are included for analysis for each variable entered into the initial assessment of variables. In the development of this CPR, 50 variables were included for initial analysis with 71 patients, which was narrowed to 11 variables after univariate regression; but even this final model used limited cases for the number of variables being considered. Failure to provide sufficient cases in the development of a CPR can produce an unstable or biased model; and for this and reasons about different patient and therapist samples, CPRs have been recognized for not performing well when applied to new populations and settings. This is why once a CPR has been derived in one population it must be validated in a different patient population with different clinicians. The manipulation CPR has been so validated, but both derivation and validation studies were performed by specially trained clinicians on US military personnel in military facilities, and the applicability of the results to other patient groups and clinicians is unknown.

The case study thus raises the possibility that patients who fit a CPR for manipulation may also fit a DP response; thus, the subgroups may not be discrete entities. However, a head-to-head comparison would be needed to determine which cluster of clinical findings was most powerful. For instance, a group of patients in whom both the CPR criteria and centralization or mechanically determined DP was elicited could be randomized to receive either manipulation or DP exercises to determine which is more effective.

**Conclusion**

In conclusion, we have presented a case study that was positive for 4 of 5 items of the CPR for manipulation responders. This patient was treated with DP flexion exercises, regaining full painless function that was maintained over at least a 6-month follow-up period. The CPR may not represent a discrete subgroup but may include patients who can be effectively managed in other ways.

**Practical Applications**

- A CPR has been developed to identify patients most likely to respond to manipulation.
- We describe a patient with 4 of 5 of these clinical characteristics who responded to flexion DP exercises after a McKenzie evaluation.
- This case study provides an example of how patients who fit CPRs may actually respond to alternate treatment.
REFERENCES

11. Fritz JM, Childs JD, Flynn TW. Pragmatic application of a clinical prediction rule in primary care to identify patients with low back pain with a good prognosis following a brief spinal manipulation intervention. BMC Fam Pract 2005;6:29 (Available at: http://www.biomedcentral.com/1471-2296/6/29).
**EARLY-ONSET MULTIPLE MYELOMA: AN ILLUSTRATIVE CASE REPORT**

Rod Kaufman, DC

**ABSTRACT**

**Objective:** This case study describes a patient diagnosed with early manifestations of multiple myeloma and illustrates relevant aspects of differential diagnosis and the use of laboratory, radiologic, and advanced imaging techniques to aid in establishing the diagnosis and issues of management.

**Clinical Features:** A 36-year-old male student experienced midback pain that occurred primarily at night in conjunction with fever and unexplained weight loss. Minor trauma induced a significant fracture and an occult fracture in the upper extremity. Physical examination revealed an elevated temperature indicating a fever of undetermined etiology. Plain radiographs revealed diffuse osteoporosis of the thoracic spine. Laboratory tests revealed anemia, hypercalcemia, and abnormal monoclonal paraprotein. Magnetic resonance imaging revealed a fracture with poor healing and an occult fracture in the upper extremity.

**Intervention and Outcome:** The patient was initially assessed for fever of undetermined etiology in association with nocturnal midback pain. Although considered a disease of the geriatric population, subsequent laboratory and radiologic evaluations established a diagnosis of early-onset multiple myeloma. Early recognition and referral with comanagement by an oncologist provided optimum care. Early-onset cases of multiple myeloma tend to have a more favorable response to treatment as compared with cases diagnosed in the geriatric population.

**Conclusion:** Multiple myeloma should be a consideration when a patient presents with nocturnal back pain and fever of undetermined etiology. Differentiating multiple myeloma from other causes of back pain is especially important in making management decisions. With a precise history and physical diagnosis, the diagnosis may be suspected, but confirmation must rely on ancillary investigations. Multiple myeloma is frequently accompanied by a poor prognosis, but early-onset cases generally respond more favorably to interventions. (J Manipulative Physiol Ther 2007;30:543-549)

**Key Indexing Terms:** Multiple Myeloma; Back Pain; Chiropractic

Multiple myeloma is a complex disorder that causes a multitude of clinical symptoms and signs mediated through a variety of mechanisms. It can mimic pain syndromes typically associated with benign etiologies, and therefore, it presents a significant diagnostic challenge to the clinician. It may, for example, be confused with mechanical back pain and present clinical signs and symptoms that result in a delay in the recognition of the entity. In addition, initial laboratory findings and radiographic imaging techniques may be equivocal, and an invasive intervention such as a biopsy is required before a definitive diagnosis is established.

Multiple myeloma is the most common primary bone malignancy, accounting for 10% of all hematologic malignancies and 1% of all cancers. In the United States, there are an estimated 16,000 new cases and more than 11,000 deaths yearly due to multiple myeloma. It is generally a disease of the geriatric population, with the mean age at diagnosis between 65 and 68 years. Approximately 3% of multiple myeloma cases are identified in patients younger than 40 years.

The classic presentation of multiple myeloma is anemia, back pain, and an elevated sedimentation rate in an older man. In addition, pain in the rib cage and pain associated with pathologic fracture, most commonly of the hip, prompt the patient to seek attention. Physical examination may reveal pallor, an elevated temperature, bone tenderness, and soft tissue masses. In addition, spinal cord compression may also occur because of tumor masses and may initiate signs and symptoms of a neurologic deficit.

Clinical signs and symptoms of multiple myeloma are related to the pathophysiology of the disease. In essence, multiple myeloma is a malignancy of plasma cells characterized by bone destruction, replacement of the bone marrow, and paraprotein formation. Replacement of bone marrow initially causes anemia and progresses later to overt bone marrow failure. Destruction of bone results in bone pain, osteoporosis, hypercalcemia, and pathologic fractures.

Although multiple myeloma is a well-defined clinical entity, establishing the diagnosis can be difficult, especially when it appears in a younger individual. However, important clues that alert the clinician to the presence of the entity may

---

Professor, Southern California University of Health Sciences, Whittier, Calif.

Submit requests for reprints to: Rod Kaufman, DC, Professor, 16200 East Amber Valley Drive, Whittier, CA 90604.

Paper submitted October 27, 2006; in revised form April 9, 2007; accepted May 1, 2007.

0161-4754/S32.00

Copyright © 2007 by National University of Health Sciences.

appear in the history and clinical examination of the patient. Furthermore, additional clues that prompt the clinician to include the possibility of multiple myeloma in the differential diagnosis of common pain syndromes may exist in the radiographic imaging of the patient and in routine laboratory tests.

A case of multiple myeloma presenting in a young patient is described to identify important clues in the historical, clinical, and diagnostic procedures necessary to differentially diagnose the disease from other etiologies producing similar signs and symptoms.

CASE REPORT

A 36-year-old male student presented to the teaching clinic of Southern California Health Sciences, Whittier, Calif, complaining of moderate to severe midback pain, worse at night, for 3 months. In addition, the patient complained of feeling tired and excessively warm throughout the day during this period. No initiating factor for the complaint of midback pain could be ascertained. The complaint of midback pain was described as deep and achy and was rated an 8 on a numerical pain scale of 0 to 10.

The patient pointed to the thoracic spine as the principal area of his pain. Despite the relatively large area of spinal pain, the patient could not recall a provocative movement or position that increased the complaint. The patient stated the pain in the midback would awaken him at night, but he would also experience it during the day while either at rest or during movement. Over-the-counter medications such as aspirin and ibuprofen provided some relief of pain during the day but less so at night. The aforementioned medications seemed to be more effective at relieving the complaint of feeling excessively warm but had no effect on the complaint of feeling tired throughout the day.

In general, the patient felt that his condition was worsening, both in the intensity of the pain and in the duration of his complaints. The patient stated he felt pain increasing in severity in his midback for 2 months before his initial consultation. He also stated the pain was more pronounced at night during this period, and he was unable to sleep throughout the night without awakening at midnight and in the early morning. The patient reported feeling excessively warm throughout this period, in conjunction with sensations of fatigue and dizziness. Family history was noncontributory, and a review of systems did not reveal any significant contributing factors.

Although the patient remarked that he had experienced a loss of weight of approximately 12 pounds over 2 weeks before the examination, the recorded weight of 149 pounds with a height of 5 ft 9 in appeared to be within normal limits. The vital signs were within normal limits, with the exception of the temperature and pulse. The body temperature recorded at examination was 103°F, with a pulse rate of 89. The general appearance of the patient was consistent, with good posture without evidence of an antalgic list or lean. His skin color appeared pale, with evidence of diaphoresis. Coordination, gait, and equilibrium appeared to be within normal limits. Gross ranges of motion of the cervical and thoracolumbar spine appeared unrestricted. Evaluation of the motor strength, reflexes, and dermatomes of the upper and lower extremities did not reveal evidence of a deficit. Cervical compression tests were unrevealing. The result of the Valsalva test was negative, and no localization of pain was noted upon percussion of the thoracic spine, although tenderness of the spine upon percussion was evident.

In consideration of the duration of the complaint of pain in the thoracic spine for 3 months without an initiating factor in conjunction with constitutional symptoms of fever and malaise, a radiographic series of the thoracic spine was ordered at the time of the initial visit. Figures 1 and 2 show diffuse osteoporosis without evidence of fracture, dislocation, and gross osteopathology.

A laboratory profile was ordered for similar considerations of pain in the thoracic spine accompanied by constitutional signs and symptoms. The results of laboratory tests revealed a normochromic anemia with rouleau formation evident on the peripheral smear. The erythrocyte sedimentation rate was markedly elevated at 90 mm/h, Westergren. The white blood cell count was within normal limits.
limits. Elevated values of blood urea nitrogen, serum creatinine, serum calcium, and serum uric acid appeared on the chemistry profile.

The key finding in the laboratory analysis of rouleau formation in the peripheral smear mandated the requisition of serum protein electrophoresis.8,9 Additional laboratory tests, occurring within a week of the initial visit, revealed a finding of a paraprotein on serum electrophoresis, consisting of a monoclonal spike in the \( \gamma \)-globulin region.

Two days earlier, the patient fell and injured his right wrist. Although the patient described his fall as "trivial", palpation revealed marked tenderness in the anatomical snuffbox. Tenderness in the anatomical snuffbox is suggestive of scaphoid fracture. Figure 3 demonstrates a fracture through the waist of the scaphoid. Subsequent radiographs obtained over a period of 6 weeks demonstrated poor healing of the fracture, so a magnetic resonance imaging (MRI) study was performed to fully assess the fracture. Figure 4A and B demonstrate T1-weighted axial and coronal images revealing an occult fracture of the os triquetrum and an unhealed fracture of the scaphoid.

The patient was informed of a diagnostic impression of multiple myeloma at the health center within 2 weeks of his initial visit. He was advised to seek consultation with an oncologist, and an appropriate referral was provided. The oncologist confirmed the diagnostic impression of multiple myeloma with a biopsy. Induction therapy reportedly consisted of Thal-Dex, a thalidomide/dexamethasone compound, in preparation for autologous peripheral stem cell transplantation. The patient has reportedly responded well to the aforementioned treatment. The oncologist noted a relatively quick diagnosis, and the young age of the patient contributed to a positive outlook in this case. In addition, an outcome measure ordered by the oncologist revealed a positive outlook. A myeloma cell-labeling index (plasma cell-labeling index) indicates how fast cancer cells are growing. A high index can predict a more rapid accumulation of cancer cells and a worse outlook.10 Fortunately, in this case, the opposite is true, and the patient reportedly had a low index upon follow-up examinations performed by the oncologist.

The patient received care at the health center for the fracture of the scaphoid. He was immobilized in a splint at the time of his initial x-rays, and an immediate consultation with an orthopedic surgeon was recommended. Follow-up communication with the orthopedic surgeon revealed that the patient eventually received a graft of bone for the unhealed fracture of the scaphoid. This occurred approximately 12 weeks after diagnosis and after a trial of immobilization. The orthopedic surgeon reported the fracture of the os triquetrum healed without complications.

**Discussion**

This case reflects the challenges involved in diagnosing the onset of multiple myeloma occurring at an early age. In an analysis of 178 cases of multiple myeloma occurring over a period of 7 years (1993-1999), Usha et al11 uncovered only 14 cases in patients younger than 4011. Such cases generally involved men as the predominating sex, with typical
presenting complaints of backache, pain in the pelvis, and weakness, accompanied by anemia and an elevated sedimentation rate.

Although the precise etiology of multiple myeloma is unknown, certain risk factors have been identified (Fig 5). Reidal and Pottern have identified the following 7 risk factors. (1) Age is the most significant multiple myeloma risk factor. Two thirds of people diagnosed with this cancer are older than 65 years. (2) Sex is also a significant risk factor, and men are 50% more likely to develop multiple myeloma than women. 3. Race is a risk factor, and multiple myeloma is more than twice as common among black Americans as white Americans. (4) Exposure to radioactivity has been suggested as a risk factor but accounts for a very small number of cases. (5) Family history also appears to be a risk factor for multiple myeloma. If people have a sibling or parent with multiple myeloma, their chance of developing it is nearly 4 times that of the general population. This is quite rare, however, and most patients have no affected relatives. (6) Occupational exposure of workers to petroleum-related industries may result in a risk of developing this cancer, but more research is needed in this area. (7) Obesity is also recognized as a risk factor for multiple myeloma.

Although the cause of multiple myeloma is unknown, an understanding of the pathophysiology of the disease has advanced theoretical explanations for the development of this cancer. Plasma cells, which transformed into malignant cells in myeloma, are a subset of B cells. It is the latter that are the producers of humoral immunity factors known as antibodies. An individual plasma cell can produce antibody molecules of only a single immunoglobulin to combine with a single antigen and is therefore known as monoclonal. If malignant transformation occurs in a single plasma cell, its clones produce only a single type of immunoglobulin, and electrophoresis demonstrates a monoclonal peak corresponding to this particular immunoglobulin.

Current theories involve chronic antigenic stimulation of a plasma cell, which results in transformation and the development of myeloma. However, once the plasma cell is transformed, it is known to produce innumerable clones, which spread hematogenously to other myelogenous areas. Once there, these neoplastic cells form sheets that replace the normal bone marrow. In addition, the myeloma cells produce

Fig 4. A, T1-weighted MRI axial image of wrist demonstrating a fracture of the scaphoid (arrow 1) and a fracture of the os triquetrum (arrow 2). B, T1-weighted MRI coronal image of wrist demonstrating a fracture of the scaphoid (arrow 1) and a fracture of the os triquetrum (arrow 2).

Fig 5. Risk factors associated with multiple myeloma.
osteoclast-stimulating factor, a cytokine that results in bone marrow destruction.\textsuperscript{13}

Concurrent with the aforementioned cascade of malignant transformations of plasma cells is the appearance of plasma cell–activating factor interleukin 6 within the bone marrow.\textsuperscript{13} This results in plasma cell proliferation. Finally, osteoblastic response in myeloma tends to be suppressed, resulting in severe demineralization and bone destruction characteristic of the disease.\textsuperscript{13} In addition to multiple risk factors and an uncertain etiology, the clinician may encounter different types of multiple myeloma. For example, monoclonal gammopathy of unknown significance is the presence of a paraprotein, but without signs and symptoms.\textsuperscript{14} Smoldering myeloma is a type of premeloma and represents a classification of myeloma between monoclonal gammopathy of unknown significance and myeloma. It presents with a paraprotein but does not have associated signs and symptoms.\textsuperscript{15}

Fewer than 10% of patients present with a single myelomatous lesion, a plasmacytoma, appearing as bubbly expansions of a single bone, often with the ribs or posterior elements of the spine, and occasionally associated with a soft tissue mass.\textsuperscript{16} A rare form of myeloma known as POEMS (polyneuropathy, organomegaly, endocrinopathy, monoclonal gammopathy, and skin changes) syndrome may demonstrate sclerotic lesions on x-rays, but this is responsible for fewer than 1\% of cases.\textsuperscript{17} The classic appearance of multiple myeloma is that of multiple, well-circumscribed, lytic, punched-out lesions in the skull, spine, and pelvis.\textsuperscript{18} In addition, multiple myeloma may present as a diffusely demineralized state of bone.\textsuperscript{18}

Confirmation of a diagnosis of multiple myeloma is dependent on the use of bone marrow morphological examination and electrophoretic analysis of the monoclonal paraprotein, which constitutes the current gold standard of techniques available to the clinician.\textsuperscript{19} The clinician may also use radiographic and advanced imaging techniques in assessing the presence and severity of multiple myeloma. Currently, conventional skeletal surveys, MRI, and positron emission tomographic/computed tomographic examinations are the most useful instruments.\textsuperscript{20} Magnetic resonance imaging remains the instrument of choice in staging tumors because it can depict local spread of tumors into surrounding tissue.\textsuperscript{21} It is not entirely specific for assessing multiple myeloma because it may incorrectly designate the stage or degree of severity of the disease.\textsuperscript{22} Positron emission tomography is similar in sensitivity to MRI/computed tomographic scans for detecting lesions and for distinguishing postsurgical scarring from tumor reoccurrence but offers a higher degree of specificity.\textsuperscript{22} Unfortunately, higher costs and limited availability for positron emission tomographic scanning currently inhibit accessibility in the clinical arena.

Multiple myeloma is staged to determine how much the cancer has advanced. Staging is important for determining treatment options and prognosis. Although many staging systems have been proposed for assessing multiple myeloma, 2 are commonly used in clinical practice. First proposed in 1975, the Durie-Salmon staging system is based on 4 factors.\textsuperscript{23}

1. The amount of abnormal monoclonal immunoglobulin in the blood or urine. Large amounts of monoclonal immunoglobulins indicate that many malignant plasma cells are present and are producing abnormal protein.
2. The amount of calcium in the blood. High levels of calcium in the blood are indicative of advanced bone damage.
3. The severity of bone damage based on x-rays. Multiple areas of bone destruction seen on x-rays indicate an advanced stage of multiple myeloma.
4. The amount of hemoglobin in the blood. Low hemoglobin levels indicate that myeloma cells occupy much of the bone marrow and that not enough space is left for the normal blood cell–producing marrow cells.

The Durie-Salmon staging system incorporates 3 stages for classifying the extent of the multiple myeloma.\textsuperscript{23}

Stage 1: A relatively small number of myeloma cells are found. All of the following features must be present: (a) hemoglobin level only slightly above normal (above 10 g/dL), (b) no x-rays appearing normal or showing only 1 area of bone damage, (c) normal blood calcium levels (less than 12 mg/dL), and (d) a relatively small amount of monoclonal immunoglobulin in blood or urine.

Stage 2: A moderate number of myeloma cells are present. Features are between stages 1 and 3.

Stage 3: A large number of myeloma cells are present. One or more of the following features must be present: (a) hemoglobin level quite low (below 8.5 g/dL), (b) high blood calcium level (above 12 mg/dL), (c) 3 or more areas of bone destroyed by the cancer, and (d) large amount of monoclonal immunoglobulin in blood or urine.

The Durie-Salmon system is still in use but has largely been superseded by the International Staging System.\textsuperscript{24} This staging system classifies patients into 1 of 3 stages according to laboratory levels of serum β2-microglobulin and serum albumin.\textsuperscript{24}

Stage 1 is characterized by a serum β2-microglobulin level of less than 3.5 mg/L and serum albumin level above 3.5 g/L. It represents the mildest form of the disease and is associated with a median survival period of 62 months.

Stage 2 is neither stage 1 or 3, meaning either that the β2-microglobulin level is between 3.5 and 5.5 mg/L regardless of the albumin level or that the albumin level is below 3.5 mg/L, whereas the β2-microglobulin level is less than 3.5 mg/L. This stage is associated with a declining median survival period of 44 months.
Stage 3 is characterized by serum $\beta_2$-microglobulin level higher than 5.5 mg/L and serum albumin level above 3.5 mg/L. This stage is associated with advanced disease and a median survival period of 29 months.

Multiple myeloma may present with various complications relative to the disease and treatment. Although it is beyond the scope of this article to delineate and discuss every potential complication, some of the principle affectations need to be recognized in clinical practice. These complications include spinal cord compression and paralysis resulting from collapse of the vertebral body and various other pathologic fractures associated with significant disease-induced osteoporosis. In addition, common affectations complicating multiple myeloma include infections occurring from compromise of the immune system, either as a result of the disease or as an adverse reaction from drug therapy. Patients may also seek attention because of complications resulting in kidney impairment and renal failure.

Unexpected symptoms may arise from the complication known as hyperviscosity syndrome, including vertigo, nausea, visual disturbances, mucosal bleeding, and alterations in mental status. In addition to impairment of the immune system, either as a result of the disease or as an adverse reaction from drug therapy. Patients may also seek attention because of complications resulting in kidney impairment and renal failure.

Unexplained weight loss. The patient in this case noted that rest did not relieve his pain.

The patient also exhibited diffuse osteoporosis as evidenced by a thoracic x-ray series. Although this might be expected in a postmenopausal woman, it is unusual in a man 36 years of age. Therefore, a further work-up should be initiated to determine the cause. In this case, laboratory investigations were revealing, and a routine blood count with differential and a sedimentation rate provided significant clues. Although the white blood count was within normal limits, a relative increase would have indicated the probability of an infection. It is now recognized that neoplastic disorders have replaced infections as the most common cause of fevers of undetermined etiology. In addition, the sedimentation rate was markedly elevated, and the latter may be encountered in infection, inflammatory entities such as arthritis, and malignancy. Without any significant history of arthritis or inflammatory disease and no elevation in the white cell count, a malignancy should be considered. The presence of rouleau formation on the peripheral smear should provide a substantial clue to the possibility of a malignancy such as multiple myeloma, and further laboratory investigations are warranted in requesting the identification of a paraprotein with electrophoresis. Confirmation of the presence of multiple myeloma is dependent upon biopsy and bone marrow aspiration.

The patient presented with a secondary finding of a fracture of the scaphoid and an occult fracture of the os triquetrum. It is unclear if these findings are directly related to the primary pathology of multiple myeloma or have manifested from a mild traumatic episode involving bones predisposed to fracture by generalized osteoporosis. The incidence of involvement of the hand in multiple myeloma is 2.9%, with bone abnormalities appearing most commonly as lytic lesions.

Treatment of multiple myeloma generally consists of drug therapy or autologous stem cell transplantation, with the latter prolonging disease-free survival and overall survival.
According to the consulting oncologist in this case, the patient has responded favorably to drug therapy and autologous stem cell transplantation. The prognosis in multiple myeloma can be based, in part, upon the stage of the disease and the age of the patient. The patient in this case was diagnosed as being in stage 1 by the consulting oncologist. The prognosis and response to treatment is usually improved in patients younger than 60 years. Thus, the prognosis for the patient in this case is considered to be good for stage 1 multiple myeloma.

CONCLUSION

A high level of clinical suspicion is needed to diagnose multiple myeloma at an early age. Failure to consider the diagnosis in the presence of clinical red flags may lead to medicolegal pitfalls in practice. These pitfalls include failure to diagnose impending pathologic fractures and failure to detect subtle myeloma lesions. Routine laboratory tests and plain film radiography can assist the practitioner in providing important clues in the diagnosis of multiple myeloma, but bone marrow aspiration and biopsy are ultimately required to confirm it. Prognostic factors can be established by staging multiple myeloma. In addition, early diagnosis of multiple myeloma in a relatively young patient provides for optimum care and an improved prognosis.

REFERENCES