# Table of Contents

## AAPM&R PRESIDENTIAL ADDRESS

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>549</td>
<td>Benjamin Franklin: The First Physiatrist?</td>
<td>Steve M. Gnatz, MD, MHA</td>
</tr>
</tbody>
</table>

## WALTER J. ZEITER LECTURE

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>551</td>
<td>Brain Injury Research: Lessons for Reinventing the Future. The 38th Zeiter Lecture</td>
<td>Ross D. Zafonte, DO</td>
</tr>
</tbody>
</table>

## ORIGINAL ARTICLES

### ORIGINAL ARTICLES

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>555</td>
<td>Mirror Therapy Enhances Lower-Extremity Motor Recovery and Motor Functioning After Stroke: A Randomized Controlled Trial</td>
<td>Serap Sutbeyaz, MD, Gunes Yavazer, MD, PhD, Nebahat Sezer, MD, B. Fasun Koseoglu, MD</td>
</tr>
<tr>
<td>560</td>
<td>How Gender Impacts Career Development and Leadership in Rehabilitation Medicine: A Report From the AAPM&amp;R Research Committee</td>
<td>Amy K. Wagner, MD, Jacinta McElligott, MD, Leighton Chan, MD, MPH, Eugene P. Wagner II, PhD, Neil A. Segal, MD, Lynn H. Gerber, MD</td>
</tr>
<tr>
<td>569</td>
<td>Reduced Longitudinal Excursion of the Median Nerve in Carpal Tunnel Syndrome</td>
<td>Alan D. Hough, PhD, Ann P. Moore, PhD, Mark P. Jones, PhD</td>
</tr>
<tr>
<td>577</td>
<td>The Role of the Back Rx Exercise Program in Diskogenic Low Back Pain: A Prospective Randomized Trial</td>
<td>Vijay B. Vad, MD, Atul L. Bhat, MD, Yasir Tarabichi</td>
</tr>
<tr>
<td>583</td>
<td>Association of Mobility Limitations With Health Care Satisfaction and Use of Preventive Care: A Survey of Medicare Beneficiaries</td>
<td>Jeanne M. Hoffman, PhD, Anne Shamway-Cook, PhD, PT, Kathryn M. Yorkston, PhD, Marcia A. Ciol, PhD, Brian J. Dudgen, PhD, Leighton Chan, MD, MPH</td>
</tr>
<tr>
<td>589</td>
<td>What Patient Attributes Are Associated With Thoughts of Suing a Physician?</td>
<td>David A. Fishbain, MD, Daniel Bruns, PsyD, J. Mark Disorbio, EdD, John E. Lewis, PhD</td>
</tr>
<tr>
<td>597</td>
<td>Cancer Screening Behaviors Among Canadian Women Living With Physical Disabilities</td>
<td>Nicole S. Cooper, MSc, Karen K. Yoshida, PhD</td>
</tr>
<tr>
<td>604</td>
<td>Is Stair Climb Power a Clinically Relevant Measure of Leg Power Impairments in At-Risk Older Adults?</td>
<td>Jonathan F. Bean, MD, MS, Dan K. Kiely, MPH, Sharon LaRose, BS, Joda Alain, BA, Walter R. Frontera, MD, PhD</td>
</tr>
<tr>
<td>610</td>
<td>Exercise Testing and Training in a Cancer Rehabilitation Program: The Advantage of the Steep Ramp Test</td>
<td>Ingrid C. De Backer, MSc, Goof Schep, MD, PhD, Adwin Hoogeveen, MD, PhD, Gerard Vreugdenhil, MD, PhD, Arnold D. Kester, PhD, Eric van Breda, PhD</td>
</tr>
<tr>
<td>617</td>
<td>Recovery of Function in Skeletal Muscle Following 2 Different Contraction-Induced Injuries</td>
<td>Richard M. Lovering, PT, PhD, Joseph A. Roche, PT, Robert J. Bloch, PhD, Patrick G. De Deyne, MPT, PhD</td>
</tr>
</tbody>
</table>
Table of Contents (continued)

626  Test-Retest Reliability of Isokinetic Knee Extension and Flexion
     Gisela Sole, BSc(Physio), MSc(Med), Jonas Hamrén, BSc(Physio), Stephan Milosavljevic, BAppSc(Physio), PhD,
     Helen Nicholson, MD, S. John Sullivan, PhD

632  Reliability of the Clinical Outcome Variables Scale When Administered Via Telephone to
     Assess Mobility in People With Spinal Cord Injury
     Ruth N. Barker, PhD, Delena I. Amsiers, MPhys, Melissa D. Kendall, MHumSrv, Kiley J. Pershouse, BSc(Wk,
     Terry P. Haines, PhD

638  Symptom Burden in Persons With Spinal Cord Injury
     Mark P. Jensen, PhD, Carrie M. Kuehn, MA, MPH, Dagmar Amtmann, Diane D. Cardenas, MD, MHA

646  Reliability and Validity of the Incontinence Quality of Life Questionnaire in Patients With
     Neurogenic Urinary Incontinence
     Brigitte Schurch, MD, Pierre Denys, MD, Chris M. Kozma, PhD, Pat Ray Reese, PhD, Terra Slaton, MS, Rich Barron, MS

653  Effect of Orthoses on Postural Stability in Asymptomatic Subjects With Rearfoot Malalignment
     During a 6-Week Acclimation Period
     Carl G. Mattacola, PhD, ATC, Maureen K. Dwyer, MS, ATC, Amy K. Miller, MS, ATC, Tim L. Uhl, PhD, ATC,
     Jean L. McCrorey, PhD, Terry R. Malone, EEd, PT, ATC

661  Fluid Collections in Amputee Stumps: A Common Phenomenon
     Rajiv Singh, MRCP, John Hunter, FRCP, Alistair Philip, PhD

664  Evidence of the Psychometric Qualities of a Simplified Version of the Activities-specific
     Balance Confidence Scale for Community-Dwelling Seniors
     Johanne Filiatrault, OT, MSc, Lise Gauvin, PhD, Michel Fournier, MA, Manon Parisien, OT, MSc, Yvonne Robitaille, PhD,
     Sophie Lafont, PhD, Helene Corriveau, PT, PhD, Lucie Richard, PhD

REVIEW ARTICLE

673  The Effects of Tai Chi on Bone Mineral Density in Postmenopausal Women: A Systematic
     Review
     Peter M. Wayne, PhD, Douglas P. Kiel, David E. Krebs, PhD, Roger B. Davis, ScD,
     Jacqueline Savetsky-German, MPH, MAOM, Maureen Connelly, MD, Julie E. Buring, ScD

BRIEF REPORT

681  Prevalence of Female Athlete Triad Characteristics in a Club Triathlon Team
     Anne Z. Hoch, DO, John E. Stavrakos, MD, Jane E. Schimke, AAS

COMMENTARY

683  The Work That Remains at the Intersection of Gender and Career Development
     Janet Bickel

DEPARTMENTS

687  Letters to the Editor
     Do We Really Want to Suggest 4 Transforaminal Epidural Steroid Injections Prior to Surgery?

688  Corrections

ORGANIZATION NEWS—AAPM&R

689  Journal-Based CME Evaluation and Application
The ideas and opinions expressed in *Archives of Physical Medicine and Rehabilitation* do not necessarily reflect those of the American Academy of Physical Medicine and Rehabilitation, the American Congress of Rehabilitation Medicine, the Editor, or the Publisher. Publication of an advertisement or other product mention in *Archives of Physical Medicine and Rehabilitation* should not be construed as an endorsement of the product or the manufacturer’s claims. Readers are encouraged to contact the manufacturer with any questions about the features or limitations of the products mentioned. The American Academy of Physical Medicine and Rehabilitation, the American Congress of Rehabilitation Medicine, and the Publisher do not assume any responsibility for any injury and/or damage to persons or property arising out of or related to any use of the material contained in this periodical. The reader is advised to check the appropriate medical literature and the product information currently provided by the manufacturer of each drug to be administered to verify the dosage, the method and duration of administration, or contraindications. It is the responsibility of the treating physician or other health care professional, relying on independent experience and knowledge of the patient, to determine drug dosages and the best treatment for the patient.

The contents of *Archives of Physical Medicine and Rehabilitation* are indexed in Index Medicus/MEDLINE, Excerpta Medical/EMBASE, Current Contents/Clinical Medicine, Science Citation Index, Citation Alert, BIOSIS, and CINAHL.
## Archives of Physical Medicine and Rehabilitation

Official Journal of the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

**MISSION STATEMENT:** The mission of the Archives of Physical Medicine and Rehabilitation is to disseminate information, with the ultimate goal of furthering the art and science of the practice of physical medicine and rehabilitation and interdisciplinary rehabilitation, and improving the health and welfare of persons with disabilities.

### EDITOR-IN-CHIEF
Jeffrey R. Basford, MD, PhD (Academy)

*Rochester, MN*

### ACADEMY EDITOR
Leighton Chan, MD, MPH

*Bethesda, MD*

**Managing Editor**
Michael A. Vasko, MA

**Editorial Assistant**
Kristen R. Golem

Michael L. Boninger, MD (Academy)

*Pittsburgh, PA*

Bruce Caplan, PhD, ABPP (Congress)

*Wynnewood, PA*

Martin D. Hoffman, MD (Academy)

*Mather, CA*

Kenneth M. Jaffe, MD (Congress)

*Seattle, WA*

David D. Kilmer, MD (Academy)

*Sacramento, CA*

David E. Krebs, DPT, PhD (Congress)

*Boston, MA*

Jay M. Meythaler, MD, JD (Congress)

*Detroit, MI*

Jonathan F. Bean, MD, MS (Academy)

*Boston, MA*

Brenda J. Brouwer, PhD (Congress)

*Kingston, ON*

John Chae, MD (Academy)

*Cleveland, OH*

John DeLuca, PhD, ABPP (Congress)

*West Orange, NJ*

Alberto Esquenazi, MD (Academy)

*Philadelphia, PA*

Robert G. Frank, PhD, ABClinP (Congress)

*Gainesville, FL*

Lynn H. Gerber, MD (Academy)

*Fairfax, VA*

Helen Hoenig, MD, MPH, OT (Congress)

*Durham, NC*

Mark P. Jensen, PhD (Congress)

*Seattle, WA*

### CONGRESS EDITOR
Allen W. Heinemann, PhD, ABPP

*Chicago, IL*

**Editorial Coordinators**
Carol J. Manow

*Lupe Soto*

**Editorial Administrator**
Karen K. Parks

Patrick K. Murray, MD (Academy)

*Cleveland, OH*

Mary M. Rodgers, PhD, PT (Congress)

*Baltimore, MD*

Elliot J. Roth, MD (Congress)

*Chicago, IL*

Dale C. Strasser, MD (Academy)

*Atlanta, GA*

Denise G. Tate, PhD, ABPP (Congress)

*Ann Arbor, MI*

Robert A. Werner, MD (Academy)

*Ann Arbor, MI*

### Editorial Board

<table>
<thead>
<tr>
<th>Associate Board</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mark S. Kaplan, MD (Academy)</td>
</tr>
<tr>
<td><em>Boston, MA</em></td>
</tr>
<tr>
<td>Christina M. Marciniak, MD (Academy)</td>
</tr>
<tr>
<td><em>Chicago, IL</em></td>
</tr>
<tr>
<td>Julie D. Moreland, MSc (Congress)</td>
</tr>
<tr>
<td><em>Hamilton, ON</em></td>
</tr>
<tr>
<td>Michael W. O’Dell, MD (Congress)</td>
</tr>
<tr>
<td><em>New York, NY</em></td>
</tr>
<tr>
<td>Jeffrey B. Palmer, MD (Congress)</td>
</tr>
<tr>
<td><em>Baltimore, MD</em></td>
</tr>
<tr>
<td>Thomas D. Rizzo Jr, MD (Academy CME Editor)</td>
</tr>
<tr>
<td><em>Jacksonville, FL</em></td>
</tr>
<tr>
<td>Joan C. Rogers, PhD, OTR/L (Congress)</td>
</tr>
<tr>
<td><em>Pittsburgh, PA</em></td>
</tr>
<tr>
<td>Bruce Shapiro, MD (Academy)</td>
</tr>
<tr>
<td><em>Baltimore, MD</em></td>
</tr>
<tr>
<td>Robert W. Teasell, MD, FRCPC (Academy)</td>
</tr>
<tr>
<td><em>London, ON</em></td>
</tr>
<tr>
<td>Richard D. Zorowitz, MD (Academy)</td>
</tr>
<tr>
<td><em>Baltimore, MD</em></td>
</tr>
</tbody>
</table>

*Editorial Office:* Suite 2510, 330 North Wabash Avenue, Chicago, Illinois 60611-7617. Phone: (312) 464-9550. Fax: (312) 464-9554. E-mail: ArchivesMail@aapmr.org Internet Home Page: www.archives-pmr.org
Benjamin Franklin: The First Physiatrist?

Steve M. Gnatz, MD, MHA


Benjamin Franklin would have been an excellent physiatrist. In this short address, I explore the qualities that Franklin possessed in the medical realm. Several ways in which we can emulate some of his traits to become better physiatrists are proposed.

Key Words: History of medicine; Physical medicine; Rehabilitation.

© 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

I recently read a marvelous book entitled Doctor Franklin’s Medicine1 by Stanley Finger. If I knew much at all about Franklin before reading this book, it was in his more traditional role as a Founding Father of the United States and, perhaps, his famous experiments with electricity. While reading this book, however, I became more interested in Franklin the scientist, researcher, thinker, and even physician. Perhaps it is fitting on this 300th anniversary of his birth that we contemplate some of his accomplishments and how they relate to us.

Franklin did not have formal medical training, but neither did many of the “medical” men of colonial America. In the 1770s, only a small minority of the people who were practicing medicine even had a college degree. And, of course, the sophistication of medical practice was much more nascent. There were patent medicines that usually contained some plant-derived active ingredient such as opium or digitals. Most of these had a “secret” formula, so to allow the inventor to profit from selling a concoction that only he/she could produce. As a relatively young man, Franklin was already financially well off because of his successful publishing business, so one might argue that he did not need to make a lot of money from his discoveries. There is evidence that Franklin shared much of his knowledge because of his altruistic nature.

Franklin was a prolific writer on many topics; luckily we have access to a great body of his writings through his letters and even his Poor Richard’s Almanac. Many of his witticisms are, of course, legendary; “Early to bed, early to rise, makes a man healthy, wealthy, and wise.” In essence, though, Franklin was a pragmatist and a keen observer of the world around him. He made several discoveries that he shared freely because of his altruistic nature.

Franklin did not have formal medical training, but neither did many of the “medical” men of colonial America. In the 1770s, only a small minority of the people who were practicing medicine even had a college degree. And, of course, the sophistication of medical practice was much more nascent. There were patent medicines that usually contained some plant-derived active ingredient such as opium or digitals. Most of these had a “secret” formula, so to allow the inventor to profit from selling a concoction that only he/she could produce. As a relatively young man, Franklin was already financially well off because of his successful publishing business, so one might argue that he did not need to make a lot of money from his discoveries. There is evidence that Franklin shared much of his knowledge because of his altruistic nature.

Franklin was a prolific writer on many topics; luckily we have access to a great body of his writings through his letters and even his Poor Richard’s Almanac. Many of his witticisms are, of course, legendary; “Early to bed, early to rise, makes a man healthy, wealthy, and wise.” In essence, though, Franklin was a pragmatist and a keen observer of the world around him. He made several discoveries that he shared freely because of his altruistic nature.

General health and fitness figured prominently in Franklin’s thinking and writing. He was a strong proponent of daily physical activity, especially swimming and horseback riding, as a way to promote and maintain health. He even taught some of his colleagues to swim, in an era when most people believed that voluntarily immersing one’s self into a body of water was too painful for him to walk. Franklin did not invent electricity, but his experiments yielded many results that are still in use. He invented the lightning rod, which was initially intended to provide protection from lightning for ships at sea, but subsequently was used ubiquitously on land as well.

Franklin’s experiments with electricity are particularly interesting when it comes to his attempts to cure “palsies” or patients with an inability to move a limb. He (among others) noted that electric current, when applied to a body part, resulted in involuntary movement. One might consider this the foundational physiologic observation that led to our electrodiagnostic techniques. He then proceeded to conduct experiments to see whether the return of voluntary function would follow if the paralyzed limb was stimulated repeatedly. Ultimately, he concluded that no prolonged benefit could be derived from his technique and he abandoned these trials. One might conclude that these were some of the earliest experiments with functional electric stimulation. Along the way, he discovered that there was a significant benefit from strong electric stimulation for patients suffering from “melancholia,” thus predating by about 200 years the routine medical use of electroconvulsive therapy for depression.

How can we apply this information to what we do today? Franklin was a natural scientist and had a keen interest in discovering how natural physical forces worked. He tried to apply this knowledge to help people deal with their disabilities and ailments. We know more about the physical world and forces now, and we can still apply that knowledge to better the human condition. Franklin was not afraid to abandon a practice if it demonstrated no useful benefit. And, once he made a discovery that improved the human condition, he shared it freely with the world. These timeless principles remain essential to our field of physical medicine and rehabilitation.

From the Department of Orthopaedics and Rehabilitation, Loyola University Medical Center, Maywood, IL.

Presented as the Presidential Address to the American Academy of Physical Medicine and Rehabilitation, November 9, 2006, Honolulu, HI.

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the author(s) or upon any organization with which the author(s) is/are associated.

Correspondence to Steve M. Gnatz, MD, MHA, Dept of Orthopaedics and Rehabilitation, Loyola University Medical Center, 2160 S 1st Ave, Maywood, IL 60153, e-mail: sg@lumc.edu.

0003-9993/07/8805-11396$32.00/0

doi:10.1016/j.apmr.2007.01.006


AAPM&R PRESIDENTIAL ADDRESS
Franklin’s belief that maintaining one’s body physically has significant health benefits was a concept that seems relevant even today. As the number of labor-saving devices has expanded, we have become progressively more sedentary as a population. Giving our attention to physical fitness as a way to maintain and preserve our patients’ health, as well as our own, is a concept that must remain central to our holistic physiatric practice.

Among Franklin’s own maladies later in his life were gout and an ongoing problem with bladder stones. He generally treated the gout through rest and adherence to a Spartan diet, avoiding red meats and alcohol, which seemed to aggravate his symptoms. Another of Franklin’s many inventions was a urinary catheter that he had a local silversmith fabricate. While urinary catheters of various materials had been used since at least the days of the Roman Empire, his solution was particularly ingenious. He instructed the silversmith to fabricate the catheter from interlocking small rings that would bend to conform to the anatomy of the urethra but would have the stiffness required for insertion. He sent the original version to his older brother (who also had trouble with a bladder stone) and provided explicit written instructions for its use. In his pragmatic way, Franklin contemplated a problem that led to pain and loss of function and designed a solution to the problem. In an era of puritan moralities, he did not shy away from addressing taboo or embarrassing issues.

Our patients have disabling conditions that may not be as evident as a hemiparesis or a missing limb. Sometimes these conditions involve basic physiologic function, such as the bowel, bladder, or sexual functions that can be uncomfortable or embarrassing for them to discuss—even in a doctor’s office. It is important that we recognize this and empower our patients to discuss openly these issues by creating an environment that is professional and supportive. We must anticipate the problems that people may be experiencing based on the underlying physiology of their disease or disability.

Another glimpse into Franklin’s character is afforded by his handling of an inquiry into the scientific veracity of Franz Mesmer’s techniques, now known as mesmerism. Mesmer’s theory, in a nutshell, postulated that there existed a force in nature he called “animal magnetism,” which he and his disciples could control to cure various maladies. In 1784 Franklin was asked to head a French commission that would investigate whether these techniques had a scientific basis. Franklin took an ingenious approach to the task. Rather than trying to disprove that Mesmer’s technique cured people, he set about to find evidence that such a force did exist. He tried to measure the force with all the instruments at his disposal and found that there was no identifiable measurement for the proposed force. Therefore, he concluded that the tools at his disposal could not substantiate such a force. Realizing that not all forces might be measured using the instruments of his day, he went a step farther and set up a test-retest paradigm. He demonstrated with this method that the force was not reproducible. The commission’s report resulted in Mesmer being discredited and he left France. Of course, Mesmer’s “force” subsequently became the foundation of what we now consider hypnotherapy. We have some well-accepted practices in physical medicine and rehabilitation today that are not easily proven to be effective, especially those practices that may prevent further deterioration or disability. So lack of evidence of efficacy of a treatment does not necessarily equal lack of efficacy. We still do not understand fully how the human body and mind work.

Being willing and able to challenge the scientific veracity of new, or even established, treatments can be a lonely and thankless task. Having the courage to do the research, however, is critical to the success of our medical specialty.

What were Franklin’s qualities that lead me to believe he would have been a great physiatrist? First, he was concerned about his fellow man and the maladies that caused pain and decreased function. In our current parlance, he was evidence-based and used the experimental techniques at his disposal to test how things worked. He was inventive. He came up with pragmatic solutions to compensate for some of the disabling conditions he (and others) experienced. He believed that maintenance of bodily function through exercise was key to a healthy life. He was so invested in freedom from regulatory injustice that he was an active supporter of America’s War of Independence. Last, and perhaps most important, he was altruistic and shared his knowledge so that others would not suffer.

Clearly, the ideal physiatrist is someone who thinks about the impairments and disabilities that afflict him/herself, the people he/she serves, and mankind. The ideal physiatrist is a person of action who does not shy away from controversy, is not easily deterred because an idea is unpopular or unconventional, and wants to personally view the evidence. Where there is no evidence, physiatrists should use pragmatic experience in their practices and rely on scientific methods to evaluate theories. They must subject their findings to peer review and publish their results, thus sharing their knowledge with the world. They must not abide regulatory tyranny and they must actively advocate for the needs of people with disabling conditions.

Perhaps we cannot really call Benjamin Franklin the “first physiatrist,” but if we emulate his example, the field of physical medicine and rehabilitation, and all of mankind, will certainly benefit.

Reference

I discuss novel dynamics in brain injury medicine that will shape the field of physical medicine and rehabilitation over the next several years. I review the lessons from previous clinical trials and discuss how rapid biotechnologic changes will influence the lives of people with disabilities. This lecture focuses on prior paradigms and addresses lessons learned, novel strategies for reinvention (including person-specific therapies), conventional therapy programs, biomaterials and devices, cellular-based therapies, and potential therapeutic interventions.

Key Words: Brain injuries; Rehabilitation; Technology, medical.
© 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

THE SPECIALTY OF PHYSICAL medicine and rehabilitation (PM&R) has had its share of visionary leaders—for example, Walter J. Zeiter, MD—who were bold enough to change the field. We must revisit our past in order to develop strategies for the future of our specialty. In this lecture, I will use the clinical care of, and research into, traumatic brain injury (TBI) as an exemplum of how we can reorganize the way we think about how we practice our particular specialty.

TBI is tiered into primary and secondary injuries. Primary injury occurs as a direct consequence of the injury (ie, subdural hematoma, contusion) while secondary injury exists as a cascade of processes that worsen after the initial injury. It is this series of interacting events that creates a complex and multifactorial disease process. These secondary injury components are a possible target for neuroprotective therapies.

This cascade of events has an impact on several pathways that influence both acute injury and recovery processes. Numerous acute care neuroprotective trials have been completed, yet many had been halted and others have failed to show efficacy for the primary outcome measure.1 The question remains, Why?

Preclinical studies using animal models have demonstrated the efficacy of various neuroprotective therapies. These animal studies have typically been performed with male, genetically pure, animals and have had with relatively proximal outcome measures such as lesion volumes. When these therapies have been used in clinical trials, they have failed to yield strong clinical dividends,2 and the results have been disappointing.3 Thus, we are now at a point at which no standard neuroprotective therapy has been developed.

Various agents have shown promise, but concerns about goals, trial design, or the measures used have slowed their successful development. Polyethylene glycol superoxide dismutase is a free-radical scavenger with a modest side-effect profile. A phase II trial of this agent by Muizelaar et al4 appeared to yield positive results but the definitive trial resulted in a negative finding for the primary measure.5 The investigators noted a 7.9% better outcome at 3 months postinjury. Perhaps this study asked for a differential outcome that was too robust and hence was too underpowered to show a small difference between groups. A small difference in outcome, however, may have meaning for the patients we serve and their families.

Several N-methyl-D-aspartate (NMDA) receptor antagonists have been tested in the treatment of acute head injury. These agents target a specific pathway and seek to ameliorate a portion of the excitotoxic process that follows TBI. No NMDA antagonist medication has yet been found efficacious in a terminal clinical trial for the primary outcome measure,2 and prior trials were stopped prematurely. For years, clinicians and researchers have focused on the role of corticosteroids in the treatment of trauma. The Corticosteroid Randomization After Significant Head injury (CRASH) study6 is the largest randomized trial of steroids for human TBI ever undertaken. The impact of intravenous steroids was evaluated in subjects with TBI who had Glasgow Coma Scale scores of 14 or lower. The study found no reduction in mortality, and in fact, reported an increase in the number of deaths at 2 weeks postinjury in the corticosteroid treatment group. Of note, mortality—even in the placebo groups—was higher than previous reports and the results of other recent steroid-based studies, such as with tirlazid mesylate (a 21-amino steroid), showed no significant benefits. The tirlazid mesylate study did have a randomization flaw because of a mismatch between groups, with subjects with more severe injuries being randomized to the treatment group. A recent phase III study7 of a proposed neuroprotective cannabinoid found that it had no significant beneficial effects.

Moderate hypothermia treatment for severe TBI has been discussed for almost 6 decades. While preliminary data have suggested positive outcomes, and indeed several studies have shown such outcomes with cardiopulmonary arrest patients, a large multisite controlled trial failed to demonstrate efficacy in TBI patients.8 The initial study from the University of Pittsburgh did, however, demonstrate efficacy.9 Why did these differences occur and did a patient subgroup benefit? The trial by Clifton et al9 brings into focus the difficulty in standardizing
multicenter trials because there was tremendous intersite variability in the acute care management and re-warming strategies; these variations may have played a significant role in the disparate findings. A subgroup analysis noted a potential benefit from hypothermia treatment among subjects who were 45 years of age or younger. The study also suggested that subjects who were already hypothermic at the time of admission also benefited from the treatment. Presently, a trial sponsored by the National Institutes of Health—the National Acute Brain Injury Study hyperthermia—II (NABIS—II)—is evaluating the role of moderate hypothermia treatment in a subgroup of subjects with severe TBI.

The use of practice guidelines appears to have an affect on the mortality rate and outcomes of people with TBI, thus making it even more challenging to detect differences. Both decreasing mortality and the decreasing rate of poor outcomes resulting from following the Brain Trauma Foundation guidelines for the treatment of acute severe head injury may be creating a benefit, but it may also be making more challenging the detection of differences resulting from innovative therapies.

In a review of TBI clinical trials, Narayan et al11 advocated for strong preclinical models and a wide window of opportunity. They also advocated for an understanding of population differences; I believe that this is a key component of the future of acute and postacute care. Careful consideration should be given to understanding genetic and biologic factors involved in the actual injury and recovery processes. Researchers should pay more attention to the adequacy of study design and/or be less preoccupied with obtaining a major effect (“hitting a home run”) than in achieving incremental improvement (the proverbial “single”).

Thus, can any drug that targets a singular mechanism be globally effective? Some researchers and clinicians have advocated using “sloppy drugs” (those that target multiple mechanisms), while others have called for incremental use of a medication cocktail to enhance neuroprotection and limit secondary injury. Of interest is a recent phase II trial by Wright et al12 of a relatively sloppy drug with multiple mechanisms of action. They studied progesterone and found it to be safe, noting a trend for better outcomes.

Postacute studies have verified many of the same concerns and they have been limited in their scope. We have learned that there is no effective treatment for the long-term prophylaxis of posttraumatic seizures. Few studies have optimally evaluated in a standardized fashion therapy treatment regimens for interventions. Postacute pharmacologic studies,15 have been criticized because of their small sample sizes and limited outcome measures. Several agents have shown promise, however. In a double-blind randomized controlled trial with 34 TBI patients who received methylphenidate, Whyte et al16 noted improved visual processing speed, but whether this relates to improved long-term outcomes is as yet unclear.

CONVENTIONAL CARE

We still do not have a clear understanding of the role of conventional therapy. Because some data suggest that too early an intervention may damage an organism, issues regarding the timing of therapeutic intervention and its appropriate withdrawal are key to the future of rehabilitation care. We also need to define the proper dosing for various therapeutic interventions and how some of the interventions can be augmented or inhibited by strategies that we now use.14

THE FUTURE: THE NEED TO REINVENT—RESTORATIVE MEDICINE AND THE BIOLOGY OF RECOVERY

Clearly, in brain injury medicine, and perhaps even in the entire field of PM&R, we must consider our future options. Using a scientific approach, we can lead the way in developing new biotechnologies. If we understand the mechanisms of recovery, we can create novel ways to enhance the recovery process. By thinking “outside the box” and challenging prior paradigms we can help design new ways to care for our patients. Examples exist in fields such as neurosurgery and otolaryngology that have reinvented themselves by designing new methods or innovations to solve long-standing problems. I believe such thinking will be critical as PM&R seeks to move forward in both musculoskeletal and rehabilitative arenas. A focus on the biology of recovery will have profound implications for our field.13 We have the opportunity to review the entire postacute continuum and its biologic links, including wellness and aging prevention.

THE NEXT GENERATION OF PHYSIATRY: THERAPEUTIC OPTIONS

Perhaps most important to the field of PM&R is the training of the next generation of physiatrists. Physicians entering physiatry today will find the practice landscape changed by biologic therapies and discoveries such as no previous generations ever experienced. Thus, we are entering an era of functional biologic physiatry. Understanding one’s own biologic factors and their role in the recovery process will carry important implications for the field of PM&R. Through an understanding of the more sophisticated and refined biologic behavior patterns, we can integrate our patients’ psychosocial and biologic profiles into our therapeutic regimes in a more seamless manner. It is likely that not every therapy will work for every person. Thus, understanding specific genetic and biologic factors that influence the recovery process is critical. A more refined appreciation of such factors will permit us to design more effective acute and postacute therapies. Work at the University of Pittsburgh has shown sex-based differences for biomarkers of oxidative stress injury after TBI.18 Genetic factors may affect recovery and it will be important to examine the relationship between genetic polymorphism and therapy (or an enriched environment).19 Using either novel imaging techniques or biochemical markers of injury to further define the extent and type of injury will be key in the future. There is the potential of developing a set of markers that can help us design and refine our patients’ therapy, throughout the sphere of physiatric practice.

Bioscaffolds are unique structural tools that provide a framework to enhance tissue healing. These novel substrates use active biologic tissues and permit these biologic structures to grow within a cocoon of a polymer-like network. Bioscaffold therapy has been used to enhance wound healing and has even been used to heal skeletal and heart muscle.20 Such therapy offers great promise for those with central or peripheral neurologic disorders and may have an important role in musculoskeletal care.

Neurotrophic factors can enhance the growth of neuronal tissue and enhance arborization and include factors such as brain-derived neurotrophic factor and nerve growth factors; these factors, and others, appear to have a significant role in synaptogenesis, plasticity, and cell survival. They are also critical in the process of neuron differentiation.21 There is clearly activity-dependent release of these neurotrophins and these factors play a role in protein synthesis in dendrites, thus en-

hancing outgrowth. A significant portion of restorative science will begin to focus on the proper delivery, method of delivery, and timing of delivery of these agents.

Cellular-based therapies are in our future, yet they remain controversial. Stem cells are pluripotential cells that differentiate into numerous specific cellular subtypes, including nerve, bone, or other tissues.22 The mechanism of stem cell facilitation and its efficacy as a potential tool in musculoskeletal and neurologic disease is not clear. Questions remain as to whether these cells act via direct replacement and proliferation or as a trophic support for the region, enhancing the milieu and encouraging endogenous processes. Stem cell therapy may also help us create a therapeutic environment, making surrounding tissue easier to manipulate and enhancing regeneration. Research using animal models has suggested that there is an attenuation of behavioral deficits and motor deficits after stem cell transplantation. Where, when, and how they repair lost neuronal connections and neural networks remains unclear, however. In addition, we must further clarify how these cells interact with activity, experience, and pharmacologic enhancement.

In a 2005 published study,23 a human cell implantation trial for those with stroke was completed. The researchers employed human N2 tera cells derived from teratocarcinoma to perform a human cellular implantation study. The study cells were implanted in stroke volunteers with residual motoric deficits. Several major findings were striking. The neurons were found at the graft site and they appeared permanently postmitotic. Although functional improvements in a small population study did not appear to be statistically significant, there was a trend for visual processing and executive function enhancement via a mechanism that is as yet unclear.

TECHNOLOGY FOR FUTURE CARE

Medicine is replete with novel companies employing biologic implantable devices for almost every extremity and disorder. The exact long-term benefit of many of these devices is not clear; however, such devices will become more common. Cortical stimulation has been advocated as a means to enhance recovery for those with focal motor weakness, as well as those with dystonia, motor control disorders, and of course Parkinson’s disease. Vagus nerve stimulation appears to diminish both epilepsy and resistant depression. The interaction among stimulatory therapy, classical therapy, and these devices is another venue for future research. Transcranial magnetic stimulation (TMS) allows us to direct magnetic current into the brain to create both an evaluative and therapeutic environment.24 TMS technology may help us map and understand further the parameters of recovery. It may also be employed as a therapeutic intervention for those with motor disorders and, perhaps, with neurocognitive disorders.

Nanotechnology will be important in the future of rehabilitation care; these engineered materials carry a functional size of less than 100 nanometers. Not only is the technology at the smallest functional level of organization but also some aspect of the material or device can be manipulated or controlled via physical or chemical signaling means. The technology is small enough to allow us to study cellular communication and signaling, and to reorganize the interaction between devices and human being. Nanotechnology will help us evaluate how neurons or muscles respond to chemical and physical forces and how subcellular stimuli work within a neuron. They are a potentially important route for drug testing as well as an evaluation tool for the optimal therapeutic environment and neuroprotection. Nanotechnology may also help us refine elements of drug transport, enhancing blood-brain barrier crossing, and opening up new venues for therapy delivering medications.25 By being employed for long-term implantation devices, nanoparticles may help enhance cellular integration and limit the chronic immune response. Nanobots are nanometer-sized robots that can be programmed to deliver on-site cellular therapy, and manipulation of muscle groups. These agents may also play a role in neural recovery and neural regeneration by limiting secondary toxins and serving as scavengers of unwanted materials postinjury.

We believe that assistive technology carries an important and underappreciated role for people with disability, especially those with brain injury. It is important to define a more comprehensive understanding of the technologic needs, and the potential benefits of these technologies to people with TBI and other disabilities. Our personalized joystick allows people with motor control disorders to control a powered wheelchair because the joystick does not permit marked deviation from a planned course.26,27 This system permits improved wheelchair mobility for those with the most severe motor disturbance. Critical to the development of cognitive assistive technology is an ability to tailor the technologic needs to the person being served. Such individualized care will allow people to function at the highest possible level.

Presently, most robot technology development has focused on motor control and developing. One of our local projects employs robots in virtual reality to provide a rehabilitation-by-distortion environment.28 This potential therapeutic concept uses virtual reality and perceptual-based gapping phenomena to enhance motor control for those with long-term significant motor deficit. Thus by fooling the system we may be able to increase motor capacity. In addition the role of so-called helper robots will be central to the future care of the elders and those with disabilities. Neuroprosthetics may be injectable microstimulators working at the neuromuscular level. They have the potential with radiofrequency control to enhance motor movement, to correct foot drop, or even to prevent skin lesions; potential therapeutic benefit may also involve central motor control with predictive equations allowing for robotic control of prosthetic devices.29 The key to this future is understanding sensory feedback in order to optimize central and peripheral control and functional utilization.30

QUESTIONING WHAT WE KNOW

In any further reinvention, we must understand whether what we do and believe is really true. Some of our strongest beliefs may need to evolve and paradigms of the past are often paradigms of failure. Because it appears that our practices may not always be optimal, we must question how we practice and look for basic and clinical evidence behind our beliefs. Some clinical examples may illustrate this idea. We had been taught to employ high-calorie diets for those post-trauma. What kind of early and late dietary regimes are optimal for those with severe TBI? What if the proportions of and type of calories that the patient should receive differ greatly from traditional convention? Some research suggests that we could be enhancing the inflammatory process by providing high-calorie diets laced with high carbohydrate.31 Can this impact on a patient’s outcome, length of stay, and recovery process? Our postacute pharmacologic belief systems have been predicated on a series of small studies and beliefs in an empirical set of dos and don’ts. It appears that at least in the animal literature pharmacologic agents, both in the acute and postacute stages can be linked to outcome after TBI. We have been told that haloperidol, at least in the postacute setting, has a negative impact on outcome. However, it appears that we need to understand “windows of opportunity” in a physiologic sense. Therapies that may have a negative role at one point in time may be neutral or positive at another period in time. Millbrandt et al32

evaluate haloperidol use among mechanically ventilated persons in an intensive care unit (ICU) setting. Millbrandt noted a lower hospital mortality among those treated with haloperidol early in their ICU course. While this study has limitations, it raises questions and suggests a possible therapeutic window when haloperidol may serve as a cytokine suppressant, and then another window, perhaps during rehabilitation, where it may have relatively negative effects on chronic recovery.

More recently, many physiatrists have turned to atypical antipsychotics for the treatment of agitation post TBI. As a field, we have done this without substantial evidence that the agents are safer or better than this chronic haloperidol. In a recent animal study, our group demonstrated that risperidone (Risperdal) acts like haloperidol, with both medications producing a potential long-term negative effect on the recovery model. Further study in animal models is needed to define the safety of other agents after TBI.

CONCLUSIONS
The future of PM&R is exciting because it is filled with the potential to further define the specific parameters of injury and to employ biotechnology in patient care. We need to return to our scientific roots and to begin to forge new thinking about the biology of recovery. We physiatrists need to be bold enough to reinvent our approach to patients and their care.

References
32. Millbrandt E, Kersten A, Kong L, et al. Haloperidol is associated with lower hospital mortality among those treated with haloperidol early in their ICU course. While this study has limitations, it raises questions and suggests a possible therapeutic window when haloperidol may serve as a cytokine suppressant, and then another window, perhaps during rehabilitation, where it may have relatively negative effects on chronic recovery.

More recently, many physiatrists have turned to atypical antipsychotics for the treatment of agitation post TBI. As a field, we have done this without substantial evidence that the agents are safer or better than this chronic haloperidol. In a recent animal study, our group demonstrated that risperidone (Risperdal) acts like haloperidol, with both medications producing a potential long-term negative effect on the recovery model. Further study in animal models is needed to define the safety of other agents after TBI.
Mirror Therapy Enhances Lower-Extremity Motor Recovery and Motor Functioning After Stroke: A Randomized Controlled Trial

Serap Sütbeyaz, MD, Gunes Yavuzer, MD, PhD, Nebahat Sezer, MD, F. Füsun Koseoglu, MD


Objective: To evaluate the effects of mirror therapy, using motor imagery training, on lower-extremity motor recovery and motor functioning of patients with subacute stroke.

Design: Randomized, controlled, assessor-blinded, 4-week trial, with follow-up at 6 months.

Setting: Rehabilitation education and research hospital.

Participants: A total of 40 inpatients with stroke (mean age, 63.5y), all within 12 months poststroke and without volitional ankle dorsiflexion.

Interventions: Thirty minutes per day of the mirror therapy program, consisting of nonparetic ankle dorsiflexion movements or sham therapy, in addition to a conventional stroke rehabilitation program, 5 days a week, 2 to 5 hours a day, for 4 weeks.

Main Outcome Measures: The Brunnstrom stages of motor recovery, spasticity assessed by the Modified Ashworth Scale (MAS), walking ability (Functional Ambulation Categories [FAC]), and motor functioning (motor items of the FIM instrument).

Results: The mean change score and 95% confidence interval (CI) of the Brunnstrom stages (mean, 1.7; 95% CI, 1.2–2.1; vs mean, 0.8; 95% CI, 0.5–1.2; \( P = .002 \)), as well as the FIM motor score (mean, 21.4; 95% CI, 18.2–24.7; vs mean, 12.5; 95% CI, 9.6–14.8; \( P = .001 \)) showed significantly more improvement at follow-up in the mirror group compared with the control group. Neither MAS (mean, 0.8; 95% CI, 0.4–1.2; vs mean, 0.3; 95% CI, 0.1–0.7; \( P = .102 \)) nor FAC (mean, 1.7; 95% CI, 1.2–2.1; vs mean, 1.5; 95% CI, 1.1–1.9; \( P = .610 \)) showed a significant difference between the groups.

Conclusions: Mirror therapy combined with a conventional stroke rehabilitation program enhances lower-extremity motor recovery and motor functioning in subacute stroke patients.

Key Words: Cerebrovascular accident; Feedback; Imagery; Motor skills; Rehabilitation.

© 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

STROKE IS THE LEADING CAUSE of serious long-term disability in adults. More than 60% of stroke survivors suffer from persistent neurologic deficits that impair activities of daily living.1 Lower-extremity motor function after a stroke is often impaired, causing restrictions in functional mobility.2,3 Traditionally, physical therapy for patients with hemiparesis in the weeks after their stroke consists of exercise therapy based on neuromuscular re-education, as well as on the practice of pre-walking functional tasks such as transfer activities, weight shifts in sitting or standing, and the maintenance of unassisted stance.4 It has been shown that functional organization of the motor system, including the primary motor cortex, can be modulated by both ipsilateral limb movement and passive observation of movement of the contralateral limb.5,7 Mirror therapy is a relatively new therapeutic intervention that focuses on moving the unimpaired limb. It was first introduced by Ramachandran and Roger-Ramachandran8 to treat phantom pain after amputation. Patients reported that they could move and relax the often-cramped phantom limb and experienced pain relief after mirror treatment. In a randomized crossover design study with chronic stroke patients, Altschuler et al9 reported that range of motion and speed and accuracy of arm movement were improved with mirror therapy more than without. Stevens and Stoykov10,11 also reported that stroke patients who trained with mirror therapy for 3 to 4 weeks had increased Fugl-Meyer Assessment scores, active range of motion, movement speed, and hand dexterity. Similarly, Sathian et al12 found that 2 weeks of intense mirror therapy in chronic stroke patients resulted in a significant recovery of grip strength and hand movement of the paretic arm. This therapy has been used to treat phantom limb pain in amputee patients,2 and in stroke patients with complex regional pain syndrome type I,13 peripheral nerve injury,14 brachial plexus avulsion,15 and the paretic hand.9,12 Mirror therapy in stroke patients involves performing movements of the unimpaired limb while watching its mirror reflection superimposed over the (unseen) impaired limb, thus creating a visual illusion of enhanced movement capability of the impaired limb.10

Functional brain imaging studies of healthy subjects suggest that excitability of the primary motor cortex ipsilateral to a unilateral hand movement is facilitated by viewing a mirror reflection of the moving hand.16 Reorganization of motor functions immediately around the stroke site (ipsilesional) is likely to be important in motor recovery after stroke, and a contribution of other brain areas in the affected hemisphere is also possible.17 Activation when a subject is doing motor tasks can also occur in the bilateral inferior parietal area, the supplementary motor area, and in the premotor cortex.18,19 Furthermore, Luft et al20 demonstrated that central adaptations occur in networks controlling the paretic as well as the nonparetic lower limb after stroke. Actions generated using motor imagery adhere to the same movement rules and constraints that physical movements follow, and the neural network involved in motor imagery and motor execution overlap, primarily in the premotor and parietal areas, basal ganglia, and cerebellum.9,10,16
In this study, we hypothesized that congruent visual feedback and motor imagery from the moving nonparetic lower extremity, as provided by a mirror, would help restore the integrity of cortical processing and thereby restore function in the affected lower extremity. We designed this randomized, controlled, assessor-blinded trial to evaluate the effects of mirror therapy using motor imagery training on lower-extremity motor recovery in patients with subacute stroke.

METHODS

Participants
We identified potential participants from an inpatient stroke rehabilitation ward. Two physiatrists (SS, NS) assessed the subjects to determine their eligibility and to obtain their written informed consent. The trial included 40 inpatients (23 men, 17 women) with hemiparesis after stroke (mean age, 63.4y; mean time since stroke, 3.7mo). Stroke was defined as an acute event of cerebrovascular origin causing focal or global neurologic dysfunction lasting more than 24 hours, diagnosed by a neurologist, and confirmed by computed tomography or magnetic resonance imaging. Patients were required to meet the following criteria for inclusion in the study: (1) first episode of unilateral stroke with hemiparesis during the previous 12 months, (2) a score between 1 and 3 (inclusive) on the Brunnstrom stages of motor recovery of the lower extremity, (3) no severe cognitive disorders that would interfere with the study’s purpose, and (4) ambulatory before stroke. The Ankara University Ethics Committee approved the protocol and all patients provided their written informed consent.

Sample Size
The required sample size was determined by using the pooled estimate of within-group standard deviations (SDs) of 4.9, obtained from the pilot data (n=10). Power calculations indicated that a sample of 40 subjects would provide an 80% (β=.20) chance of detecting a 20% (α=.05) difference in improvement between the groups.

Design
We used an assessor-blinded, randomized controlled design. The same investigator (SS), who was blinded to the treatment assignment, performed all the assessments. After baseline measurements were obtained, the patients were randomly assigned to either the mirror group (n=20) or the control group (n=20), using computer-generated random numbers (fig 1). Blocks were numbered, after which we used a random-number generator program to select numbers that established the sequence in which blocks were allocated to one or the other group. A physician who was blinded to the research protocol and was not otherwise involved in the trial conducted the random-number program.

Intervention
Both the mirror group and the placebo group participated in a conventional stroke rehabilitation program, 5 days a week, 2 to 5 hours a day, for 4 weeks. The conventional program is patient-specific and consists of neurodevelopmental facilitation techniques, physical therapy, occupational therapy, and speech therapy (if needed). The mirror group received an additional 30 minutes a day of a mirror therapy program consisting of nonparetic ankle dorsiflexion movements. Subjects were in a semi-seating position on a bed, while the mirror board (40×70cm) was positioned between the legs perpendicular to the subject’s midline, with the nonparetic leg facing the reflective surface. Subjects observed the reflection of the nonparetic leg while flexing and extending the ankle at a self-selected speed under supervision but without additional verbal feedback. The placebo group performed the same exercise for the same duration, but the nonreflecting side of the mirror was used.

Outcome Measures
Outcome was measured in terms of motor recovery (Brunnstrom stages), spasticity (Modified Ashworth Scale [MAS]), walking ability (Functional Ambulation Categories [FAC]), and motor functioning (motor items of FIM instrument). The measures were taken before treatment, at 1 month post-treatment, and at 6 months (follow-up).

Lower-Extremity Motor Recovery
We assessed lower-extremity motor recovery with the Brunnstrom stages. The 6 grades of those stages for the lower extremity are: (1) flaccidity, (2) synergy development (minimal voluntary movements), (3) voluntary synergistic movement (combined hip flexion, knee flexion, and ankle dorsiflexion, both sitting and standing), (4) some movements deviating from synergy (knee flexion exceeding 90° and ankle dorsiflexion with the heel on the floor in the sitting position), (5) independence from basic synergies (isolated knee flexion with the hip extended and isolated ankle dorsiflexion with the knee extended in the standing position), and (6) isolated joint movements (hip abduction in the standing position and knee rotation with inversion and eversion of the ankle in the sitting position). We used the Brunnstrom stages because they reflect the underlying motor control based on clinical assessment of movement quality.

Spasticity
We used the MAS to grade the spasticity of the ankle flexor and extensor muscles. The MAS is a 5-point ordinal rating scale with good interrater reliability and is designed to measure muscle tone. Higher MAS scores indicate worse spasticity.
Table 1: Characteristics of the 2 Study Groups

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mirror (n=20)</th>
<th>Control (n=20)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)</td>
<td>62.7±9.7</td>
<td>64.7±7.7</td>
<td>.36*</td>
</tr>
<tr>
<td>Sex (women/men)</td>
<td>10/10</td>
<td>7/13</td>
<td>.52*</td>
</tr>
<tr>
<td>Type of injury (ischemic/hemorrhagic)</td>
<td>16/4</td>
<td>17/3</td>
<td>.50*</td>
</tr>
<tr>
<td>Paretic side (right/left)</td>
<td>6/14</td>
<td>7/13</td>
<td>.50*</td>
</tr>
<tr>
<td>Time since stroke (mo)</td>
<td>3.5±1.3</td>
<td>3.9±1.9</td>
<td>.44*</td>
</tr>
<tr>
<td>Mean Brunnstrom stages</td>
<td>2.4±0.7</td>
<td>2.5±1.0</td>
<td>.73*</td>
</tr>
<tr>
<td>Median Brunnstrom stages</td>
<td>3.0</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>Mean MAS</td>
<td>2.6±0.5</td>
<td>2.3±0.7</td>
<td>.13*</td>
</tr>
<tr>
<td>Median MAS</td>
<td>3.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Mean FAC</td>
<td>1.9±0.5</td>
<td>2.0±0.7</td>
<td>.79*</td>
</tr>
<tr>
<td>Median FAC</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>FIM motor score</td>
<td>48.3±5.5</td>
<td>50.2±11.6</td>
<td>.51*</td>
</tr>
</tbody>
</table>

NOTE. Values are mean ± SD, median, or number of patients. *Student t test.

Walking Ability
Walking ability was assessed with the FAC and the motor items of the FIM. The FAC is a reliable and valid assessment tool whose 6 categories provide information on the level of physical support patients need to ambulate safely both indoors and outdoors.²⁴ Our subjects were permitted to use walking devices (eg, canes) during the measurements.

Motor Functioning
The FIM is the functional status component of the Uniform Data System for Medical Rehabilitation.²⁵ It is widely used in rehabilitation centers and has properties useful to stroke investigators. It has 18 items that measure independent performance in self-care, sphincter control, transfers, locomotion, communication, and social cognition. FIM scores range from 1 to 7: a FIM item score of 7 is categorized as “complete independence,” while a score of 1 indicated “complete dependence” (performs <25% of task). Scores below 6 mean that a subject requires supervision or assistance from another person.²⁶ We used the motor items of the FIM; the total score ranges from 13 (lowest) to 91 (highest). The reliability and validity of the Turkish version of the FIM has been well documented.²⁷

Table 2: Motor Recovery, Spasticity, Walking Ability, and Motor Functioning Scores of Patients at Pretreatment, Post-Treatment, and Follow-Up

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Group</th>
<th>Pretreatment</th>
<th>Post-Treatment</th>
<th>Follow-Up</th>
<th>∆ (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brunnstrom stages</td>
<td>Mirror</td>
<td>2.4±0.7</td>
<td>3.5±0.8</td>
<td>4.2±0.8</td>
<td>1.7 (1.2–2.1)</td>
<td>.002</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>2.5±1.0</td>
<td>3.0±0.7</td>
<td>3.4±0.8</td>
<td>0.8 (0.5–1.2)</td>
<td>.102</td>
</tr>
<tr>
<td>MAS</td>
<td>Mirror</td>
<td>2.6±0.5</td>
<td>2.3±0.5</td>
<td>1.8±0.7</td>
<td>0.8 (0.4–1.2)</td>
<td>.610</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>2.3±0.7</td>
<td>2.2±0.7</td>
<td>1.9±0.7</td>
<td>0.3 (0.1–0.7)</td>
<td>.001</td>
</tr>
<tr>
<td>FAC</td>
<td>Mirror</td>
<td>1.9±0.5</td>
<td>2.8±0.6</td>
<td>3.6±0.9</td>
<td>1.7 (1.2–2.1)</td>
<td>.002</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>2.0±0.7</td>
<td>2.9±0.7</td>
<td>3.5±0.9</td>
<td>1.5 (1.1–1.9)</td>
<td>.557</td>
</tr>
<tr>
<td>FIM motor</td>
<td>Mirror</td>
<td>48.3±5.5</td>
<td>65.9±4.8</td>
<td>69.9±5.9</td>
<td>21.4 (18.2–24.7)</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>50.2±11.6</td>
<td>61.7±14.6</td>
<td>62.9±12.8</td>
<td>12.5 (9.8–14.8)</td>
<td>.001</td>
</tr>
</tbody>
</table>

NOTE. Values are mean ± SD. P values were obtained using analysis of variance for repeated measures. Abbreviation: ∆, mean change at follow-up from baseline.

Statistical Analysis
We analyzed the data using SPSS® for Windows. Groups were compared at baseline using the t test for independent samples for the continuous variables, and the chi-square test for categorical data. All outcome variables were normally distributed; for that reason we chose analyses of variance with repeated measures to test our hypothesis, with a between-subject factor at 2 levels (the 2 groups) and a within-subject factor at 3 levels (the time: pretreatment, post-treatment, follow-up). The interaction of group and time determined the efficacy of the mirror therapy on the outcome measures.

RESULTS
Demographic and clinical characteristics of the 40 participants, as well as baseline comparisons of the groups, are presented in table 1. None of the patients missed more than 2 scheduled sessions. Three patients from the mirror group and 4 patients from the control group could not attend the follow-up clinic for final evaluation because of socioeconomic reasons. We did not observe any adverse events. Baseline comparisons revealed that age, gender, injury characteristics, time since stroke, Brunnstrom stages, MAS, FIM motor, and FAC scores did not differ between the groups (P>.05). All assessed outcome parameters improved significantly in both groups after the treatment and continued to improve at follow-up (table 2). The mean change score and 95% confidence interval (CI) of the Brunnstrom stages (mean, 1.7; 95% CI, 1.2–2.1; vs mean, 0.8; 95% CI, 0.5–1.2; P=.002), as well as the FIM motor score (mean, 21.4; 95% CI, 18.2–24.7; vs mean, 12.5; 95% CI, 9.6–14.8; P=.001), showed significantly more improvement at follow-up in the mirror group than in the control group. Neither MAS (mean, 0.8; 95% CI, 0.4–1.2; vs mean, 0.3; 95% CI, 0.1–0.7; P=.102) nor FAC (mean, 1.7; 95% CI, 1.2–2.1; vs mean, 1.5; 95% CI, 1.1–1.9; P=.610) showed significant differences between the groups.

DISCUSSION
This study reveals that in our group of stroke patients, mirror therapy combined with a conventional rehabilitation program provides additional long-term benefits in terms of lower-extremity motor recovery and motor functioning. Several recent studies on paretic upper extremity, although undersized and not sufficiently controlled, have indicated that mirror therapy may be a promising tool with which to promote motor...
recovery, mobility, muscle strength, dexterity and functionality after stroke. To our knowledge, ours is the first study to investigate the effects of mirror therapy on the paretic lower extremity.

It is well known that an increased inflow of signals from sensory modalities via various ways can enhance plasticity of the brain. Sensory processes (including vision, audition, proprioception, touch, and pressure) can mediate feedback information that is available as a result of movement. Verbal cueing and coaxing by therapists, visual and auditory feedback from electromyography, forceplate (balance and weight shift training), computer screen (virtual reality and web-based telerehabilitation), and kinematic feedback from an electrogoniometer are the most common examples used for stroke rehabilitation. In this study, we used mirror therapy to give visual feedback to the patients to enhance lower-extremity motor recovery. Studies have shown that mirror illusions have measurable effects on brain activity. Altschuler et al hypothesized that mirror therapy provides visual input of a normal movement of the affected arm in stroke patients, which may compensate for a decreased or absent proprioceptive input. Stevens and Stoykov defined mirror therapy as a form of visually guided motor imagery, which is the mental performance of a movement without overt execution of that movement. Extensive clinical, neurophysiologic, and neuroimaging evidence demonstrates that motor imagery involves the same neural networks as motor execution. Another possible mechanism is the involvement of the mirror neuron system. Mirror neurons are bimodal visuomotor neurons that are active during action observation, mental stimulation (imagination), and action execution. For example, it has been shown that passive observation of an action facilitates M1 excitability of the muscles used in that specific action. Mirror neurons are now generally understood to underlie the learning of new skills by visual inspection of the skill.

In our study, all subjects performed voluntary ankle dorsiflexion movements on their nonparetic side during the therapy sessions. We selected ankle dorsiflexion because it represents selective motor control in the lower extremity after stroke. Ankle movement training is known to facilitate brain reorganization, and the angle paradigm may serve as an ongoing physiologic assay of the optimal type, duration, and intensity of rehabilitative gait training. In the lower extremity, voluntary ankle dorsiflexion is a way of indicating the achievement of selective motor control. Once voluntary movement is achieved (Brunnstrom motor recovery stage II recovery or beyond), synergistic patterns are then modified to selective (out-of-synergy) patterns.

In this study, in contrast to motor recovery and motor functioning, the between-group difference was not significant for walking ability and spasticity. Spasticity has a complex pathophysiologic mechanism that enhanced visual feedback may not be sufficient to influence or control. Walking is also a complex performance and requires for normal gait include (but are not limited to) muscle strength, coordination, balance, endurance, etc. For motor learning to be successful, the desired motor task must be practiced in a pattern that is as close to normal as possible and practiced intensively. Our intervention did not provide for practice of a gait pattern that is close to normal. It is believed that repeated task-specific protocols induce brain reorganization that facilitates functional improvements. Cognitive involvement, functional specificity, and progressive complexity of the tasks being trained are the key variables of motor training and cortical reorganization. We did not increase the complexity level of the training in this study. Another explanation for the lack of effect on walking ability may be that the 4 weeks (20 sessions, a total of 10h) in which the mirror therapy was applied may have been too short to produce significant benefits for such a complex activity. Kwakkel reported that a minimum of at least 16 hours augmentation was necessary to determine the required amount of practice needed to affect function. Because there are very few studies on the use of mirror therapy in patients with stroke, there is no widely accepted agreement on the duration, timing, and application of such a program.

Our results confirm that visual feedback via a mirror is a simple, inexpensive and, most importantly, a patient-specific treatment. We believe that incorporating mirror therapy into the conventional rehabilitation program at an early stage of treatment, and applying it for a long period of time (perhaps continuing the therapy at home after discharge), may be beneficial in improving the effects and outcome on lower-extremity motor recovery and function. The limitations of this study are the relatively small study population and the fact that we did not use imaging techniques (eg, functional magnetic resonance imaging, positron emission tomography) that might have demonstrated brain reorganization after therapy.

CONCLUSIONS

Mirror therapy combined with a conventional rehabilitation program enhanced lower-extremity motor recovery and functioning in our subacute stroke inpatients.

References


Supplier

a. Version 11.5; SPSS Inc, 233 S Wacker Dr, 11th Fl, Chicago, IL 60606.
How Gender Impacts Career Development and Leadership in Rehabilitation Medicine: A Report From the AAPM&R Research Committee

Amy K. Wagner, MD, Jacinta McElligott, MD, Leighton Chan, MD, MPH, Eugene P. Wagner II, PhD, Neil A. Segal, MD, Lynn H. Gerber, MD


Objective: To examine the role that gender plays in meeting the medical academic mission by assessing career development, leadership, and research productivity among rehabilitation researchers.

Design: Prospective, cross-sectional cohort study.

Setting: National survey.

Participants: Three hundred sixty rehabilitation professionals linked to the American Academy of Physical Medicine and Rehabilitation, Association of Academic Physiatrists, and/or the American Congress of Rehabilitation Medicine.

Intervention: Online or paper survey.

Main Outcome Measures: Research skills, resources and productivity, salary, leadership, and academic advancement.

Results: Results suggested that women rated themselves as being less skilled and having fewer resources for research compared with their male counterparts. Additionally, significantly fewer women applied for grant funding and had a lower publication rate compared with men. A proportionally larger number of women remained at lower academic ranks than men, and fewer women achieved senior academic ranks or positions of leadership. Even after adjusting for potential confounding factors, female sex remained a significant variable associated with lower salaries and lower manuscript production. Unlike men, female respondents tended to believe that being a woman was a negative factor with respect to academic advancement, leadership opportunities, salary, and resources.

Conclusions: Female rehabilitation researchers were less developed professionally than their male counterparts and saw themselves as disadvantaged. These findings have potential implications for attracting women into rehabilitation research and the rehabilitation research community’s efforts to sustain its academic mission, to improve research capacity, and to meet the needs of the 52 million people in the United States with disabilities.

Key Words: Academic medical centers; Career mobility; Gender; Leadership; Rehabilitation; Research.

© 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

ACADEMIC MEDICAL CENTERS have a triple mission of clinical care, education, and research. Departments of physical medicine and rehabilitation (PM&R) and rehabilitation sciences within academic medical centers play a leading role in advancing the field of rehabilitation science and improving the health of vulnerable populations such as people with disabilities. In the current health care and research funding environment, academic medical centers continue to face unprecedented challenges in sustaining this triple mission.1,2 To sustain and achieve advances and leadership in medical science, the talents, perspectives, and skills of a diverse and vibrant faculty are required.

The Association of Academic Medical Centers (AAMC) has highlighted marked gender disparities for U.S. medical school faculty, in career choice, promotion, salaries, and leadership in academic medical centers.3 The percentage of women entering medical school in the United States grew from 5.1% in 1960 to 50% by 2003.4 Despite this rapid growth, several studies5-6 have shown that women in many disciplines progressed through academic ranks more slowly than men, were promoted less, published less, and had lower salaries than men of comparable rank. It is particularly concerning that women have not been attaining important leadership roles.7 In a 2003–2004 AAMC report, there were 108,000 U.S. medical school faculty of which 30% were women. Twenty-seven percent of associate professors were women, whereas only 12% of full professors and 10% of medical school deans were women.8 A survey of emergency physicians at 105 emergency medicine residency training programs in the United States identified that women were less likely to hold major leadership positions, spent a greater percentage of time in clinical and teaching activities, published less in peer-reviewed journals, and were less likely to achieve senior academic rank.9 Wright et al10 identified significant gender differences in faculty salaries, ranks, tracks, leadership positions, resources, and perceptions in 1 U.S. medical school. This study identified that women earned 11% less than men, on average, after controlling for rank, years in rank, track, degree, specialty, and administrative positions. These differences in academic advancement and salary could not be attributed to differences in productivity or family commitments.3 These data underscore the need to understand the key environmental drivers underlying an increasingly recognized gender gap between men and women.
with research productivity, career advancement, leadership, and salary.

Despite the recognition of this gap between men and women in academic medicine, gaining an understanding of their root causes is complex. Vance and Larson, in a review of the health care and business literature, showed that research on leadership in health care and business has been primarily descriptive to date. In addition, these authors identified that very limited research on leadership and health care outcomes exists. However, Bickel et al cite a number of workplace issues that could lead to a cumulative disadvantage for women including (1) difficulty finding and obtaining effective career advancing mentorship, (2) inflexible tenure and advancement policies that may compromise desired family-work balance at a time when women may want to start families, (3) inherently less value placed (by both men and women) on work women do, (4) increased risk of burnout for women, (5) less control over clinical work flow, (6) more difficulties with physician-allied health professional relationships, (7) less cultural latitude to display assertive behavior in the workplace, and (8) lower compensation for the same work.

Compared with other clinical specialties, there is a greater percentage of women in PM&R (41% compared with average of 30%). In addition, PM&R has higher rates of female associate and full professors compared with family practice and obstetrics and gynecology even though these specialties have a similar proportion of female faculty. However, little is known about the role of women as leaders and researchers in the rehabilitation community.

Importantly, the rehabilitation community has increasingly recognized a need to identify ways to increase research capacity in a manner that advances the scientific field and better addresses the needs of its growing consumer population. Exploration of how gender influences career choice, career advancement, research, and leadership potential of women in the rehabilitation sciences is essential to advance the field and also to enhance rehabilitation research capacity. Therefore, the purpose of this study was to characterize the impact of gender on a number of personal and professional indices of productive research and successful careers in the field of rehabilitation medicine. Additionally, perceptions of diversity within the rehabilitation community were examined. Solutions for optimizing gender diversity and fully integrating women in the academic rehabilitation community are discussed in the context of these findings.

**METHODS**

**Survey Development and Sampling**

The survey development, sampling procedures, and response rates for this analysis have been previously described. Briefly, an invitation to complete the survey was sent to professionals in the field of rehabilitation. To maximize the response rate, respondents were sent an introductory letter and frequent e-mail reminders. In addition, both online and paper surveys were made available. A total of 212 questions were included in the survey, and question formats included multiple choice, Likert rating scales, fill in the blank, and open-ended questions. For this analysis, demographic information, data regarding home life, training, academic rank, and salary were included. Additionally, information regarding research environment and resources, research productivity, personal research skills, personal research resources, and leadership roles was reported.

**Study Population**

Members of the American Academy of Physical Medicine and Rehabilitation (AAPM&R), the Association of Academic Physiatrists (AAP), and the American Congress of Rehabilitation Medicine (ACRM) were contacted for this survey. AAPM&R members were contacted if they had documented involvement in 1 of 20 AAPM&R special interest groups. Additionally, people known to be involved in publicly funded rehabilitation research were contacted for participation. There were a total of 360 surveys collected from over 100 institutions and used for analysis. Of those that responded, 70% held a medical (MD) or osteopathic (DO) degree, 61% held both an MD and a doctoral degree (PhD), 17% held a PhD, and 69% held other degrees. Because a major goal of this report was to understand how gender influences careers, leadership, and rehabilitation research capacity, analyses were performed only on those that indicated involvement in research within the past 5 years. As such, 271 respondents were included for the majority of the analysis (107 women, 164 men).

**Statistical Analysis**

Descriptive statistics, including means, medians, percent- ages, and standard error of the means, are reported. Means of each group are reported with medians in parentheses, except where noted. The number in each group used for each analysis is shown in the corresponding tables. Mann-Whitney nonparametric analysis was used to compare differences between the 2 groups for all continuous and Likert (ordinal)-scaled data. Chi-square analysis was used for categorical data. Yates correction for continuity is reported for ×2 tables.

Logistic regression analyses were performed to examine factors affecting academic rank, tenure status, and grant applications. For the purposes of regression modeling, rank was dichotomized into junior (instructor, assistant professor) versus senior (associate professor, professor). Linear regression was used to evaluate factors affecting salary and publication record. Univariate analysis was used in each case to identify significant variables (P < .05) to include for multivariate modeling. In addition to gender, representative factors reflecting professional experience (eg, years out from training), resources (eg, start up funding), academic productivity (eg, grant dollars, number of grants, publication record), and leadership roles (eg, institutional and national) were included as independent variables in the models. A backward stepwise approach for single elimination of independent variables was used to create multivariate models. The least significant variable was eliminated at each step until all remaining variables each had a significant P value of .05 or less. To assess the additive effect of gender on the outcome, the variable gender was forced back into the model if it did not meet significance with initial multivariate model construction. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated for the logistic models, and R 2 correlation values were computed for the linear models. All data analyses were conducted by using SPSS software, and P less than .05 was considered statistically significant.

**RESULTS**

**Population Demographics**

Respondents who reported engaging in research activities in the last 5 years were used for this analysis. Male respondents were slightly older than female respondents. The mean age of male respondents was 47.3 years, and the mean age for female respondents was 45.1 years (P = .07). There were no significant differences in racial distribution for men and women. Seventy-
nine percent of the women and 81.3% of the men were white. Eighty-one percent of respondents held an MD or DO degree, and 18.7% held a degree other than an MD or DO. Additionally, there were no significant differences between sex groups for practice location for both physicians and nonphysicians. The majority of physicians worked in an academic or academic-affiliated institution (women, 72.4%; men, 76.8%). The remainder reported working in a Veterans Affairs (VA) setting (women, 6.6%; men, 6.7%), private practice (women, 14.5%; men, 9.7%), or other venues. Seventy-one percent of female nonphysicians worked in an academic setting, and 56% of the men in this group worked in academics. Significantly more men reported holding both an MD and PhD compared with women (9.8% vs 0.9%, P = .001). Women had significantly less years of experience since graduating from residency or graduate training compared with men (12.4y vs 15.2y, P = .02), as well as fewer years of postdoctoral research fellowship training than men (.28y vs .66y, P = .02). Female respondents reported spending approximately 22.2% of their time engaging in funded or internally sponsored research activities, which was similar to men (26.1%) (P = .30). Additionally, more women reported working part time for at least some portion of their career (36.8% vs 13.5%, P < .001).

Home Responsibilities

Significantly more male respondents had children compared with female respondents (82.8% vs 66.9%, P = .003). However, of those reporting having children, significantly fewer men were the primary or shared caregiver for their children (49.6% for men vs 75% for women, P < .001). Additionally, 42.7% men reported that their spouse is the primary caregiver, whereas only 10.3% of women reported their spouse as being the primary caregiver (P < .001). Moreover, 41.7% of women reported that they are the primary person responsible for household chores, whereas only 9.4% of men reported being the primary person responsible for household (P < .001). Interestingly, for those who were married, the number of hours that female respondents reported that their spouses worked is significantly more than what male respondents reported for their spouses (42.6h/wk vs 23.0h/wk, P < .001). More women than men changed jobs because of their spouses (28.3% vs 10.6%, P < .001).

Academic Rank and Tenure

Although there were no significant differences in the number of years spent at their current rank (women, 5.3y vs men, 6.4y), men generally held higher ranks than women (P < .001). Thirty percent of men reported being at the full professor level, whereas 8.3% of women reported being at this level. These findings are consistent with 2004 AAMC data on rank for academic rehabilitation medicine departments. Although the percentage at the associate professor level was similar between sex groups (women, 22.6% vs men, 27.7%), women were more likely to hold assistant professor (women, 48.8% vs men, 36.2%) or instructor (women, 7.1% vs men, 2.3%) ranks. Interestingly, there were significant differences noted in the proportions of each group reporting their future goal for academic rank (P < .001). Eighty-one percent of men reported their goal for academic rank being full professor, whereas only 56.3% of women reported this as their goal. Although men and women similarly reported that their institution offered a tenure stream, significantly fewer female academicians reported that their positions were inside the tenure stream (25.6% vs 42.9%, P = .02). Additionally, female academicians self-reported that they worked fewer hours than their male counterparts (54.3h vs 59.8h, P < .001).

Academic Productivity

Women reported being less productive with their publication record compared with men. Women reported publishing an average of 2.66 publications as a first or senior author in a rehabilitation journal over the past 5 years, whereas men reported 4.28 publications (P = .03). Additionally, women reported having only published an average of 1.09 publications over the last 5 years as a first or senior author in a nonrehabilitation journal, whereas male respondents reported having published an average 2.97 publication in nonrehabilitation journals in the last 5 years (P < .001). Total number of publications authored or coauthored over the course of their career was also significantly less for women compared with men (12.3 vs 26.9, P < .001).

Fewer women also reported that they had applied for a research grant (58.9% vs 76.3%, P < .001). However, for those respondents who had applied for a grant, no sex differences were noted with the number of federal grants currently held as a principal investigator (women, 2.26 vs men, 2.27) or coinvestigator (women, 1.90 vs men, 2.11), the amount of federal grant funding (women, $1.56 million vs men, $2.60 million), or percentage of salary that was grant funded (women, 52.3% vs men, 38.5%). There was a trend toward fewer years of federal grant funding for women compared with men (8.5y vs 10.5y, P = .06), possibly because of the shorter number of years in their postgraduate career compared with men.

Research Skills and Resources

Although not statistically significant, women reported having less startup funding associated with their position compared with men ($26,769 vs $47,013). The majority of both groups (women, 75.7%; men, 71.6%) received no startup funding with their current position. Table 1 describes respondents’ opinion on the adequacy of current position resources. For many categories, both men and women believed the resources accessible to them in their current positions were less than adequate. However, women almost uniformly reported significantly worse scores than men for adequacy of these resources. Table 2 displays how men and women rated themselves for a variety of personal research skills. Here women rated themselves as significantly less prepared for understanding and implementing research design and methods, being a mentor, and being able to publish or present their work. Additionally women believed that they had less basic science and translational research skills and were less qualified to lead a research team. There was also a trend (P = .06) for women to report less skill with statistical methods and interpretation.

Salary and Compensation

Annual salary, including incentive pay, is reported for the group of respondents who were MDs or DOs and practiced full-time. Table 3 reports salary by practice location for both men and women. Table 4 shows salary by rank for those who reported being in an academic setting. Regardless of practice setting or academic rank, women reported smaller salaries compared with their male counterparts.

Leadership

Table 5 describes leadership roles for men and women at the institutional level as well as within national societies. In each case, women lagged significantly behind men in terms of leadership positions held. Although similar proportions of both men and women reported having a mentor, women received fewer research awards, participated in and led fewer institutional and national society committees, and held fewer direc-
GENDER AND CAREER DEVELOPMENT IN REHABILITATION, Wagner

Table 1: Respondent Opinion With Respect to Current Position Resources*

<table>
<thead>
<tr>
<th>Resource</th>
<th>Women Mean (Median)</th>
<th>Men Mean (Median)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office space</td>
<td>2.78 (3)</td>
<td>2.18 (2)</td>
<td>.000</td>
</tr>
<tr>
<td>Laboratory space</td>
<td>3.98 (5)</td>
<td>3.52 (4)</td>
<td>.004</td>
</tr>
<tr>
<td>Secretarial support</td>
<td>3.42 (3.5)</td>
<td>3.01 (3)</td>
<td>.006</td>
</tr>
<tr>
<td>Research assistant support</td>
<td>3.85 (4)</td>
<td>3.51 (4)</td>
<td>.051</td>
</tr>
<tr>
<td>Access to graduate student</td>
<td>4.02 (4)</td>
<td>3.57 (4)</td>
<td>.001</td>
</tr>
<tr>
<td>Access to resident support</td>
<td>3.92 (4)</td>
<td>3.38 (3)</td>
<td>.000</td>
</tr>
<tr>
<td>Departmental startup funding</td>
<td>4.09 (4)</td>
<td>3.69 (4)</td>
<td>.009</td>
</tr>
<tr>
<td>Departmental subsidized</td>
<td>4.03 (4)</td>
<td>3.56 (4)</td>
<td>.002</td>
</tr>
<tr>
<td>Department honors</td>
<td>3.82 (4)</td>
<td>3.44 (4)</td>
<td>.051</td>
</tr>
<tr>
<td>Access to statistical</td>
<td>3.16 (3)</td>
<td>2.64 (2)</td>
<td>.000</td>
</tr>
<tr>
<td>Access to PhD collaborators</td>
<td>3.01 (3)</td>
<td>2.42 (2)</td>
<td>.001</td>
</tr>
<tr>
<td>Access to MD collaborators</td>
<td>2.93 (3)</td>
<td>2.62 (3)</td>
<td>.029</td>
</tr>
<tr>
<td>Access to interdisciplinary</td>
<td>2.80 (3)</td>
<td>2.46 (2)</td>
<td>.011</td>
</tr>
<tr>
<td>Access to interdepartmental</td>
<td>2.84 (3)</td>
<td>2.52 (2)</td>
<td>.027</td>
</tr>
<tr>
<td>Access to professional</td>
<td>3.33 (3)</td>
<td>2.94 (3)</td>
<td>.010</td>
</tr>
<tr>
<td>development mentor(s)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE. Values are mean (median).
*Scaling for question: excellent, 1; good, 2; adequate, 3; inadequate, 4; not available, 5.

Table 2: Respondent Opinion With Respect to Personal Research Skills*

<table>
<thead>
<tr>
<th>Research Skill</th>
<th>Women Mean (Median)</th>
<th>Men Mean (Median)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research design and methods</td>
<td>2.53 (3)</td>
<td>2.18 (2)</td>
<td>.002</td>
</tr>
<tr>
<td>Statistical approach and interpretation</td>
<td>3.09 (3)</td>
<td>2.86 (3)</td>
<td>.063</td>
</tr>
<tr>
<td>Performing quantitative statistics</td>
<td>3.34 (4)</td>
<td>3.21 (3)</td>
<td>.261</td>
</tr>
<tr>
<td>Grant-writing skills</td>
<td>2.97 (3)</td>
<td>2.79 (3)</td>
<td>.223</td>
</tr>
<tr>
<td>Grant funding resources</td>
<td>3.25 (4)</td>
<td>3.05 (3)</td>
<td>.163</td>
</tr>
<tr>
<td>Research mentorship</td>
<td>3.08 (3)</td>
<td>2.70 (3)</td>
<td>.007</td>
</tr>
<tr>
<td>Research publication skills</td>
<td>2.87 (3)</td>
<td>2.37 (2)</td>
<td>.000</td>
</tr>
<tr>
<td>Research presentation skills</td>
<td>2.57 (2)</td>
<td>2.04 (2)</td>
<td>.000</td>
</tr>
<tr>
<td>Basic science techniques</td>
<td>3.88 (4)</td>
<td>3.55 (4)</td>
<td>.036</td>
</tr>
<tr>
<td>Translational research skills</td>
<td>3.54 (4)</td>
<td>3.14 (3)</td>
<td>.004</td>
</tr>
<tr>
<td>Research team management</td>
<td>3.12 (3)</td>
<td>2.70 (3)</td>
<td>.004</td>
</tr>
</tbody>
</table>

NOTE. Values are mean (median).
*Scaling for question: excellent, 1; good, 2; adequate, 3; inadequate, 4; not available, 5.

Table 3: Current Total Salary Including Incentive Pay*

<table>
<thead>
<tr>
<th>Practice Setting</th>
<th>Women Mean ($)</th>
<th>Men Mean ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Private practice</td>
<td>170,083</td>
<td>258,614</td>
</tr>
<tr>
<td>Academic</td>
<td>132,608</td>
<td>182,325</td>
</tr>
<tr>
<td>Academic affiliated</td>
<td>168,884</td>
<td>201,417</td>
</tr>
<tr>
<td>VA</td>
<td>151,143</td>
<td>164,375</td>
</tr>
<tr>
<td>Other</td>
<td>163,333</td>
<td>186,600</td>
</tr>
<tr>
<td>Total</td>
<td>151,380</td>
<td>201,003</td>
</tr>
</tbody>
</table>

*All full-time MD and DO respondents (2004) used in analysis (N=360).

Table 4: Current Total Salary Including Incentive Pay for Academicians*

<table>
<thead>
<tr>
<th>Academic Rank</th>
<th>Women Mean ($)</th>
<th>Men Mean ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructor</td>
<td>103,333</td>
<td>130,000</td>
</tr>
<tr>
<td>Assistant</td>
<td>134,767</td>
<td>154,618</td>
</tr>
<tr>
<td>Associate</td>
<td>150,534</td>
<td>193,545</td>
</tr>
<tr>
<td>Full</td>
<td>176,000</td>
<td>226,333</td>
</tr>
<tr>
<td>Other</td>
<td>125,667</td>
<td>180,000</td>
</tr>
<tr>
<td>Total</td>
<td>132,608</td>
<td>182,325</td>
</tr>
</tbody>
</table>

*All full-time MD and DO respondents (2004) in academics used in analysis (n=214).

Diversity Perceptions

Table 6 describes self-reported perceptions that men and women had about career development, access to mentorship, and leadership and promotion opportunities for women in the field of rehabilitation medicine. As a group, men reported that female sex was a neutral factor when considering each of the categories listed. In contrast, women reported that female sex was almost uniformly a negative factor with each category.

Regression Analysis

In univariate analysis, years out of training, publication record, federal funding, institutional and national leadership, and gender had a positive significant association with the probability of attaining a senior rank. Multivariate regression modeling (table 7) suggested that the factors most influential on rank were years out of school, first or senior author publications over the last 5 years, institutional leadership positions, and federal funding over the last 5 years. After adjusting for these factors, gender did not have a significant additive effect on the probability of achieving a senior rank. For tenure status, univariate analysis suggested that publication record, number of years out from training, gender, grant funding, and institutional leadership all were positively associated with an increased probability of attaining tenure. In multivariate analysis, career publications and institutional leadership positions remained significantly associated with tenure status. After adjusting for these factors, gender did not have a significant additive effect on the probability of achieving tenure (table 8). In both models, many of the factors positively impacting rank and tenure status were also negatively associated with gender in descriptive analysis.
Gender was most strongly associated with salary in univariate analysis. Other significant variables included publication record, years out of training, leadership positions, rank, and tenure status. In multivariate analysis, female sex remained the most significant negative predictor of higher salary. More national and institutional leadership positions as well as more hours worked and years out from training were positively associated with higher salaries (table 9).

Univariate regression of factors associated with leadership positions (defined as chairmanships, directorships, committee chairs at an institutional or national level) was also explored. Once again, grant funding, publication record, years out from training, gender, and hours worked were positively associated with more leadership positions. In multivariate analysis, more hours worked, more years out from training, and higher current grant dollars remained positively associated with more leadership positions. There was also a trend for total publication record to be significantly associated with more leadership positions (P = .06). Gender, however, was not significantly associated with leadership positions after adjusting for hours worked, years out from training, current grant dollars, and total publication record (table 10). Because grant funding and publication record significantly influenced the endpoints of rank, tenure, leadership, and both were significantly associated with gender, we sought to further evaluate the factors affecting total publication record and probability of writing a grant.

Univariate analysis showed that more years out from training, male sex, and more postgraduate research training were significantly associated with higher publication productivity. In multivariate analysis, years out from training, male sex, and more years of research training remained positively associated with a better publication record (table 11). Univariate factors associated with an increased probability of writing a grant include more hours worked, years out from training, gender, research training, and more startup funding. In multivariate analysis, years out from training, research training, and hours worked remained statistically significant. After adjusting for these factors, gender did not have a significant additive effect on the probability of writing a grant (table 12).

### DISCUSSION

To our knowledge, this was the first study to assess the role of gender in research participation and productivity, as well as career development and leadership in the rehabilitation community. The unadjusted results of this survey suggested that there are many significant differences between male and female rehabilitation professionals in regard to important factors that lead to a successful academic career including development of research skills, allocation of resources, salary, external funding, publications, and preparation for research. Because some of the outcomes we assessed, such as salary and rank, were affected by many factors, we performed multivariate analyses in an attempt to control for these variables. After controlling for a variety of confounding factors, we found that gender remained an independent and significant factor in salary and manuscript production but not rank, tenure, and leadership positions. However, given the interrelationships among gender and factors ultimately affecting rank, tenure, and leadership, gender gaps within these domains must still be considered. Additionally, we

#### Table 5: Leadership Roles Held by Respondent

<table>
<thead>
<tr>
<th>Leadership Position</th>
<th>Women</th>
<th>Men</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Has received national award(s) for research (%)</td>
<td>16.8</td>
<td>32.9</td>
<td>.005</td>
</tr>
<tr>
<td>Has served on any type of federal scientific review board (eg, NIH NIDRR, CDC, VA) (%)</td>
<td>27.9</td>
<td>43.5</td>
<td>.015</td>
</tr>
<tr>
<td>No. of leadership positions held in last 5y</td>
<td>0.70</td>
<td>1.28</td>
<td>.000</td>
</tr>
<tr>
<td>No. of institutional or departmental committees ever served over last 5y</td>
<td>4.00</td>
<td>6.01</td>
<td>.003</td>
</tr>
<tr>
<td>No. of institutional or departmental committee ever chaired over last 5y</td>
<td>0.90</td>
<td>2.52</td>
<td>.000</td>
</tr>
<tr>
<td>No. of national committees served over last 5y</td>
<td>1.55</td>
<td>2.64</td>
<td>.009</td>
</tr>
<tr>
<td>Number of national committees chaired over last 5y (eg, AAP, AAPM&amp;R, ACRM)</td>
<td>0.38</td>
<td>1.18</td>
<td>.007</td>
</tr>
<tr>
<td>No. of national organizations in which you have been an officer in the last 5y</td>
<td>0.40</td>
<td>0.78</td>
<td>.007</td>
</tr>
<tr>
<td>No. of journals in which you are currently a reviewer</td>
<td>1.42</td>
<td>2.64</td>
<td>.000</td>
</tr>
<tr>
<td>No. of journals in which you are currently an editor</td>
<td>0.15</td>
<td>0.49</td>
<td>.001</td>
</tr>
</tbody>
</table>

Abbreviations: CDC, Centers for Disease Control and Prevention; NIDRR, National Institute on Disability and Rehabilitation Research; NIH, National Institutes of Health.

#### Table 6: Perceptions About Gender Diversity*

<table>
<thead>
<tr>
<th>Being a Woman in the Rehabilitation Professional Community as a Factor When Considering</th>
<th>Women</th>
<th>Men</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic rank</td>
<td>3.30</td>
<td>2.99</td>
<td>.000</td>
</tr>
<tr>
<td>Ability to get promoted</td>
<td>3.41</td>
<td>2.97</td>
<td>.000</td>
</tr>
<tr>
<td>Tenure status</td>
<td>3.37</td>
<td>2.97</td>
<td>.000</td>
</tr>
<tr>
<td>Starting salary</td>
<td>3.65</td>
<td>3.16</td>
<td>.000</td>
</tr>
<tr>
<td>Salary increases</td>
<td>3.52</td>
<td>3.12</td>
<td>.000</td>
</tr>
<tr>
<td>Degree of protected research time</td>
<td>3.25</td>
<td>3.01</td>
<td>.000</td>
</tr>
<tr>
<td>Degree of patient and clinical responsibilities</td>
<td>3.03</td>
<td>2.93</td>
<td>.041</td>
</tr>
<tr>
<td>Degree of teaching responsibilities</td>
<td>2.99</td>
<td>2.92</td>
<td>.113</td>
</tr>
<tr>
<td>Departmental research support</td>
<td>3.19</td>
<td>2.95</td>
<td>.000</td>
</tr>
<tr>
<td>Ability to find good mentorship</td>
<td>3.31</td>
<td>2.97</td>
<td>.000</td>
</tr>
<tr>
<td>Departmental leadership and status</td>
<td>3.30</td>
<td>2.99</td>
<td>.000</td>
</tr>
</tbody>
</table>

NOTE. Values are mean (median).

*Scaling for question: very positive factor, 1; positive factor, 2; neutral, 3; negative factor, 4; very negative factor, 5.

#### Table 7: Results for Backward Stepwise Multivariate Regression Analyses for Factors Affecting Academic Rank (junior, senior) (n=197)

<table>
<thead>
<tr>
<th>Variable</th>
<th>P</th>
<th>OR</th>
<th>95% CI for OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years out of school</td>
<td>.000</td>
<td>1.232</td>
<td>1.143–1.329</td>
</tr>
<tr>
<td>Publications past 5y</td>
<td>.001</td>
<td>1.101</td>
<td>1.039–1.166</td>
</tr>
<tr>
<td>Institutional leadership positions past 5y</td>
<td>.007</td>
<td>1.284</td>
<td>1.071–1.539</td>
</tr>
<tr>
<td>Total years federal funding</td>
<td>.015</td>
<td>1.146</td>
<td>1.026–1.279</td>
</tr>
</tbody>
</table>
found that there were many personal life characteristics that differed between sex groups.

Our results confirmed the findings of others and extended them to include the rehabilitation community interested in pursuing an academic mission.\textsuperscript{5,6,8,10,13-15} For instance, Tesch et al\textsuperscript{15} has reported that, among a cross-section of medical school faculty, women physicians were promoted more slowly than men. Bickel et al\textsuperscript{10} reported that women lagged behind men in academic rank across both medical and surgical specialties. One previous study\textsuperscript{13} evaluating cardiothoracic surgeons reported slower academic advancement, lower salaries, less publication productivity, and gender bias for women during the course of their careers. Similar to our study, personal life characteristics differed between sex groups, with men having more children and less household and childcare responsibilities.\textsuperscript{13} Similar results were also noted in a population of general surgeons surveyed.\textsuperscript{14} It is interesting that, similar to our study, Schroen et al\textsuperscript{14} reported that women lagged behind men in academic rank, tenure status, and (importantly) career aspirations to achieve high academic ranks and position of leadership.

Female physiatrists conducting rehabilitation research within the last 5 years fared substantially less well than men financially. They were paid less at all faculty ranks. Additionally, although gender disparities in compensation appeared to occur regardless of practice setting, they were least noticeable in the VA setting. In our study, gender remained a significant predictor of compensation among academics, even after adjusting for other independent variables (such as productivity, seniority, leadership) associated with gender. To make appropriate gender comparisons, salary analysis was limited to full-time physicians because they were the largest subgroup of respondents. Although the survey results did not identify conclusive reasons for this finding, sex discrimination or cultural or environmental biases regarding the value and contributions of female physiatrists to the academic mission cannot be dismissed. Additionally, it is well documented that women do not negotiate as frequently and as successfully as men for resources, salary, and promotion. The reasons for this phenomenon are complex.\textsuperscript{16}

Sex differences in salary are pervasive across all job classes. In fact, the U.S. Census Bureau reported that, even when adjusting for differences in work patterns and other key variables, women still earned approximately 80% of what men did across all job classes and categories in the year 2000.\textsuperscript{17} Sex differences with physician compensation were more pronounced, with female physicians earning approximately 63% of what male physicians earned across all medical and surgical specialties.\textsuperscript{17} In our cohort, our unadjusted results suggested that female physiatrists earned approximately 75% of what male physiatrists did. Despite increased awareness regarding equitable compensation and the fact that women comprise such a large segment of the population in medicine, this disparity still remains. As such, a high priority should be placed on understanding the issue and eliminating wage discrimination.

Another key area in which women lagged behind men in the rehabilitation community was in the area of leadership. Women in rehabilitation did not attain as many leadership positions within their university or nationally. Explanations for this were not clear from the survey data. It could be that women were not offered these jobs or perhaps turned them down more often than men. Although similar numbers of women and men reported having a professional development mentor, satisfaction with access to good mentors and self-rated mentorship skills lagged for women. These factors could adversely affect the ability of women to develop the professional networks necessary for involvement with leadership positions. Regardless, women seemed less well integrated into the interdepartmental committee structures and had fewer hospital-wide responsibilities. In multivariate analysis, gender did not remain a significant factor after controlling for other variables. However, because the primary factors in multivariate analysis affecting the acquisition of leadership positions were related to academic productivity, an area in which women lagged behind, gender still appeared to be integrally related to leadership in rehabilitation medicine. Because administrative and committee activities are often necessary for tenure and promotion, the lack of participation in these activities by women likely has a significant impact on both career and financial advancement. As women engage in scholarly activities and participate in administrative duties important to rehabilitation departments, the value of these activities need to be adequately recognized to result in appropriate academic advancement, promotion, and salary increases.\textsuperscript{7}

Women published less and obtained considerably less public or private funding than men. These findings were also consistent with other studies examining the careers of women in academic medicine.\textsuperscript{5,8,13,15,18} For the group of women who chose to write grant applications, the level of funding was

### Table 8: Results for Backward Stepwise Multivariate Regression Analyses for Tenure (n=203)

<table>
<thead>
<tr>
<th>Variable</th>
<th>$P$</th>
<th>OR</th>
<th>95% CI for OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total publications</td>
<td>.000</td>
<td>1.041</td>
<td>1.025–1.057</td>
</tr>
<tr>
<td>Institutional leadership positions past 5y</td>
<td>.003</td>
<td>1.244</td>
<td>1.077–1.437</td>
</tr>
</tbody>
</table>

### Table 9: Results for Backward Stepwise Multivariate Regression Analyses for Salary*

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\beta$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>.262</td>
<td>.000</td>
</tr>
<tr>
<td>Institutional leadership positions past 5y</td>
<td>.245</td>
<td>.000</td>
</tr>
<tr>
<td>Hours worked</td>
<td>.183</td>
<td>.006</td>
</tr>
<tr>
<td>National leadership positions past 5y</td>
<td>.153</td>
<td>.016</td>
</tr>
<tr>
<td>Years out of school</td>
<td>.145</td>
<td>.029</td>
</tr>
</tbody>
</table>

* $n=191$, $R^2=.345$.

### Table 10: Results for Backward Stepwise Multivariate Regression Analyses for Leadership*

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\beta$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours worked</td>
<td>.240</td>
<td>.000</td>
</tr>
<tr>
<td>Years out of school</td>
<td>.223</td>
<td>.002</td>
</tr>
<tr>
<td>Current grant dollars</td>
<td>.200</td>
<td>.004</td>
</tr>
<tr>
<td>Total publications</td>
<td>.152</td>
<td>.059</td>
</tr>
</tbody>
</table>

* $n=211$, $R^2=.235$.

### Table 11: Results for Backward Stepwise Multivariate Regression Analyses for Factors Affecting Total Publications*

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\beta$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years out of school</td>
<td>.441</td>
<td>.000</td>
</tr>
<tr>
<td>Gender</td>
<td>.151</td>
<td>.005</td>
</tr>
<tr>
<td>Postgraduate training</td>
<td>.130</td>
<td>.015</td>
</tr>
</tbody>
</table>

* $n=271$, $R^2=.270$. 

comparable to that of men who applied for grant funding. Reasons for less success within these domains could be both professional and personal. With respect to resources for research, there was not a single item in which women indicated that they were better supported than men. Additionally, female respondents reported being less capable to conduct research compared with male respondents. One study\textsuperscript{10} suggests that women in training programs felt less capable of conducting research compared with male counterparts participating in the same program. Importantly, women in our study were more likely than men to feel that gender played a role in the disparities described in the survey with respect to resources, environment, leadership opportunities, and advancement. Differences in perceptions (see Table 6) about the presence of a gender gap with resources and advancement were supported by the fact that, unlike their male counterparts, 75% of female respondents were not in a tenure-track position, had less startup funding, had inferior salaries, and lagged behind men in leadership positions and promotion through the academic ranks.

Our survey results regarding research productivity may have been influenced by the fact that more men than women reported having combined MD and PhD degrees and reported spending more time in fellowship training. However, there also may be potential sex differences in this type of self-reporting scheme that reflect confidence with research skill levels or a bias with interpretation of skill level within the context of the survey questions posed. Conversely, sex differences in research productivity may be because of substantially differing personal responsibilities or choices. Although about half of both male and female respondents said they shared childcare and other caregiver roles, women were more likely to take on the role of primary caregiver.

Table 12: Results for Backward Stepwise Multivariate Regression Analyses for Factors Affecting Writing a Grant (n=271)

<table>
<thead>
<tr>
<th>Variable</th>
<th>P</th>
<th>OR</th>
<th>95% CI for OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years out of school</td>
<td>.000</td>
<td>1.116</td>
<td>1.070–1.164</td>
</tr>
<tr>
<td>Hours worked</td>
<td>.000</td>
<td>1.038</td>
<td>1.025–1.051</td>
</tr>
<tr>
<td>Postgraduate training</td>
<td>.010</td>
<td>1.518</td>
<td>1.106–2.082</td>
</tr>
</tbody>
</table>

Although men and women seemed to be attracted to academic careers for similar reasons, women were less satisfied with both their educational preparation for research and their research environment. Our survey suggests that, similar to many other studies from other specialties, women in rehabilitation medicine did not achieve senior academic ranks or tenure as frequently as men.\textsuperscript{10,11,13-15} Importantly, women also were less likely to aspire to promotion to full professor. Slightly more than half the women (56%) in our cohort aspired to reach the professor level, whereas 81% of men wanted to attain this rank. Relative differences in reporting for this question may be related to lack of participation of women in activities required to achieve this rank. In the traditionally male-dominated arena of academic medicine, it is possible that women are not being encouraged to participate in activities commonly accepted as important for promotion or advancement. It is also possible that women lack female role models who can successfully mentor them. Additionally, there also may be gender differences in the personal and cultural value men and women place on their participation in these activities.\textsuperscript{10,20,21} Unfortunately, AAMC reported that the proportion of men to women at senior ranks has remained largely unchanged for the last 15 to 20 years.\textsuperscript{20} As such, a path of significant cultural change that leads to an improvement for how women in academic medicine advance through the ranks is both challenging and complex.

This survey did not provide conclusive evidence for why these sex differences exist. Future research should specifically address this question. Results could be because of different financial or cultural expectations, differing personal or professional goals, or outright sex discrimination. However, what is clear is that the field of rehabilitation research needs to increase its capacity.\textsuperscript{12} Because 50% of medical students are women, the future of any specialty, including PM&R, is intimately linked with its ability to develop and effectively use women.\textsuperscript{70} From this perspective, it is important that we recruit and retain a higher number of all rehabilitation researchers, including women. To do this, we will need to address gender barriers, particularly within the professional environment. Our findings may have wider implications for the medical profession as a whole, particularly given that rehabilitation medicine has a greater percentage of women than other clinical specialties.

Several other industry groups have been reexamining their work culture in order to retain and develop higher numbers of women.\textsuperscript{22} The potential advantages of fostering diversity and leadership have been shown in the corporate world. For example, IBM’s strategy in taking full advantage of a diverse market for talent was strongly associated with the generation of billions of dollars of new revenue.\textsuperscript{23} Morahan and Bickel\textsuperscript{22} have suggested that the medical community adopt some of the positive changes occurring in the business community by eliminating environmental factors that disadvantage women, rewarding team work and collaboration, and valuing diversity and work effectiveness. To promote the development and advancement of women, the U.S. Department of Health and Human Services included women’s leadership as a required component of the nationally funded Centers of Excellence in Women’s Health. Establishing a network of women working together has been aimed to reduce feelings of isolation and encourage role modeling as a strategy to train a cadre of female researchers.\textsuperscript{24}

Development of a mentorship program for residents, students, and entering faculty, as well as chairs and senior faculty may very well have a positive impact on the rehabilitation research enterprise. When implementing these mentorship programs, the specific needs of women within the current culture and environment and leadership methods to effectively use human resources and increase intellectual capital through faculty development should be emphasized. Disparities might also be addressed through reexamining academic procedures for promoting diversity when recruiting new faculty. Furthermore, qualified women should be placed in the tenure track. Salary and incentive structures should be equally applied for both men and women, with leadership checking regularly to ensure that sex-related compensation inequities do not occur. Developing a standard metric for promotion may lead to increased participation, compensation, and opportunity for women in academic research. Contributions to the overall academic enterprise, including teaching, mentoring, program development, and research, should be taken into account. Specific federally funded training programs targeting women in rehabilitation research are warranted in light of the current findings regarding women’s perceptions about their research skills. Additionally, increased participation in senior level career development forums such as the Executive Leadership in Academic Medicine program for women at AAMC\textsuperscript{25} and AAMC’s Early Career Women Faculty Professional Development (for which the Foundation for PM&R provides a scholarship)\textsuperscript{26} may be helpful to generate more women interested in and capable of pursuing leadership positions.

Previous reports\textsuperscript{10} have suggested that academic reward and disadvantage are largely created and reinforced at the depart-
ment level and that chairs play a key role in integrating and advancing both women and minorities. Recently, AAMC has reported that department chairs believed that the constraints of traditional gender roles, manifestations of sex discrimination in the medical environment, and lack of effective mentors are among the most prominent barriers for women in medicine. Among the intervention strategies discussed, mentorship networks, extending the tenure clock, and institutional mechanisms for addressing unprofessional conduct were considered effective strategies by the chairs interviewed.27 Although specific interventions for mentorship, improving the tenure process, and reporting misconduct may be helpful, major changes in the organizational culture in rehabilitation medicine are likely to be the most fruitful. Here, chairs may also take a leading role by being held personally accountable for encouraging diversity. They should be rewarded for creating a flexible and effective work environment and promoting academic advancement for women. In light of the survey findings presented, it is critically important that we in the rehabilitation community examine these cultural issues within our discipline and departments. We must actively engage our peers and our leaders in a meaningful dialog with the goal of promoting the entry and retention of women into academic and research positions. These explorations should be broad and include issues surrounding diversity, salaries, resource allocation, academic advancement, work effectiveness, job satisfaction, research capacity, and the academic rehabilitation mission.

Study Limitations

Our study had several limitations. The survey invitation was to all members of AAP, AAPM&R, and ACRM. However, many of the societies enlisted to participate in this survey were comprised of large numbers of private practitioners and people without a primary interest in research or academics. This sampling strategy, although inclusive, may have led to many people choosing not to respond to the survey. In addition, there was always the possibility of selection bias in our study. However, data suggested that faculty from over 100 institutions responded to the survey. AAMC reported that in 2004 there was approximately 1046 faculty in academic rehabilitation departments.11 Given that a minority of rehabilitation faculty actively engage in research, it is likely that the survey has captured a substantial percentage of those in the academic rehabilitation community conducting research. Because of the sampling strategy, the vast majority of our respondents were physicians. Thus, our results may not extend as well to those PhDs performing rehabilitation research. However, the majority of respondents did report participating in research over the last 5 years, suggesting that the survey captured a fairly large proportion of the small rehabilitation community interested in the academic and research mission. Additionally, some key data points like rank were consistent with AAMC data,4 suggesting that the respondents were a representative sample of academicians and researchers in rehabilitation. The respondents were a relatively young group and because of the cross-sectional nature of the study, we lack robust data on salary, promotion, resources, and environment for respondents over time.

CONCLUSIONS

Our results suggest that women in the academic rehabilitation community are less successful than their male counterparts and see themselves as disadvantaged. The issues raised by this study require aggressive actions in identifying root causes for gender disparities and creating effective action plans to correct compensation discrepancies and promote academic development and leadership for women within the rehabilitation community. Importantly, these findings have implications for the rehabilitation research community’s efforts to sustain its academic mission, to improve research capacity, and to meet the needs of the 52 million people in the United States with disabilities.

References


Supplier
a. Version 12.0; SPSS Inc, 233 S Wacker Dr, 11th Fl, Chicago, IL 60606.
Reduced Longitudinal Excursion of the Median Nerve in Carpal Tunnel Syndrome

Alan D. Hough, PhD, Ann P. Moore, PhD, Mark P. Jones, PhD

ABSTRACT. Hough AD, Moore AP, Jones MP. Reduced longitudinal excursion of the median nerve in carpal tunnel syndrome. Arch Phys Med Rehabil 2007;88:569-76.

Objective: To determine if longitudinal excursion of the median nerve is reduced in patients with carpal tunnel syndrome (CTS).

Design: Case-control study.

Setting: University human movement laboratory.

Participants: Nineteen patients with CTS (8 men, 11 women; mean age, 57±15y), and 37 healthy controls (8 men, 29 women; mean age, 48±10y).

Interventions: Not applicable.

Main Outcome Measures: Longitudinal excursion of the median nerve, and the ratio of nerve to flexor digitorum superficialis tendon excursion at the carpal tunnel evoked by finger extension. Measurements were taken using a validated Doppler ultrasound technique, and tests were conducted with the elbow positioned in extension and flexion.

Results: Mean longitudinal excursion of the median nerve was significantly greater in controls (11.2±2.8mm) than patients (8.3±2.6mm) with the elbow extended (P=0.013), but not with the elbow flexed (controls, 12.5±2.5mm; patients, 10.2±3.1mm; P=0.089). Mean nerve/tendon excursion ratios were significantly greater in controls (.32±.07) than patients (.23±.06), with the elbow extended (P<.001), and flexed (controls, .36±.06; patients, .28±.10; P=0.019). Discriminant analysis identified that 11 (58%) of the 19 patients and 3 (8%) of the 37 controls showed a nerve/tendon excursion ratio of .25 or less when tested with the elbow in extension.

Conclusions: Reduced longitudinal excursion of the median nerve at the carpal tunnel was identified in a substantial proportion of patients with CTS. Further studies are merited to determine if reduced median nerve excursion at the carpal tunnel is clinically relevant in CTS, and can be influenced by movement-based interventions.

Key Words: Carpal tunnel syndrome; Median nerve; Rehabilitation; Ultrasonography, Doppler.

© 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

CARPAL TUNNEL SYNDROME (CTS) is the most commonly reported peripheral nerve entrapment, with a prevalence rate in a general population estimated at about 3%.1-2 The pathophysiology is incompletely understood but aspects of mechanical injury within the carpal tunnel are considered most likely.3-5 Although the mechanical injury is primarily associated with compression of the median nerve as it passes through the carpal tunnel, it has been hypothesized by a number of sources that reduced or restricted longitudinal excursion of the median nerve contributes to the pathophysiologic process.6-11 Based on this hypothesis, nerve and tendon gliding exercises have been advocated for the conservative and postoperative management of CTS with the putative aim of restoring the normal gliding mechanisms of the nerve and tendons through the carpal tunnel.7,12 The effectiveness of gliding exercises remain inconclusive,13,14 and currently only 1 study provides indirect evidence that longitudinal excursion of the median nerve may be reduced in CTS patients.9 Indeed, during the course of the study reported here, Erel et al.15 found no significant reduction in the longitudinal excursion of the median nerve in CTS patients compared with asymptomatic controls, when measured by a speckle tracking ultrasound technique proximal to the carpal tunnel.

Ultrasound imaging can be used to visualize longitudinal excursion of peripheral nerve trunks in vivo,16 and the movement can be reliably quantified using a Doppler,17 or speckle tracking18 approach. The Doppler technique described in this study was subject to a series of validity studies using string-phantoms and porcine specimens. Using the results of these studies, which are the subject of another article currently in preparation, we estimated that the accuracy of the technique described in this study would be in the region of ±0.7mm (limits of agreement19) when appropriate corrections were made to take account of the effects of intrinsic spectral broadening,20 and out-of-plane (nonlongitudinal) movements.21

The purpose of this study was to use the Doppler technique to determine if there was direct evidence for reduced longitudinal excursion of the median nerve in patients with CTS. Measurements were taken of median nerve and flexor digitorum superficialis (FDS) tendon excursion at the carpal tunnel evoked by active finger extension. The FDS tendon values were used to calculate a nerve/tendon excursion ratio as described by Szabo et al.11 on the basis that this value may control for intersubject anthropometric variations and provide a more sensitive method for detecting reduced nerve excursion. As far as we are aware, this is the first time that longitudinal excursion of the median nerve and the nerve/tendon ratio has been measured directly at the carpal tunnel in vivo.

METHODS

Experimental Hypothesis

Patients with CTS would show reduced longitudinal excursion of the median nerve at the carpal tunnel during finger extension (either absolute or relative to FDS tendon excursion) compared with asymptomatic controls.
Study Design and Equipment

The study design was a comparative case-control involving 1 group of participants with idiopathic CTS (n=19), and a control group of healthy volunteers (n=37). The sample size calculation was made with reference to data from previous in vivo and cadaveric studies,9-11,17,18,22 and selected to be sufficient to detect at least a 15% between-group difference with an α level of .05 and a power of 0.8.

We used a Sononace 6000C ultrasound system8 with a wideband linear-array transducer center frequency of 7.5MHz (B-Mode); 5.13MHz (D-Mode) for all image acquisition. An upper-limb support jig was custom designed for the study, consisting of wooden struts attached to the left and right frame-sides of a standard hydraulic plinth, and 1 adjustable upper-limb support that could attach to the left or right struts depending on the limb to be tested. The jig allowed adjustment for forearm length, carrying angle, and positioning of the elbow in flexion or extension (fig 1).

Participants

We recruited 21 patients with a clinical diagnosis of idiopathic CTS of at least 1 month in duration from local orthopedic clinicians and general practitioners. Thirty-seven asymptomatic volunteers aged 18 years or over were recruited from staff and students at the local university. Prospective volunteers with the following conditions were excluded from participation in either group: (1) history of injury or symptoms attributed to the neck or upper limbs during the previous 6 months, which had required treatment (excluding CTS for the patient group), (2) history of major trauma to the neck or limb to be tested, (3) systemic neuropathy, and (4) pregnancy. Ethics approval was obtained from the local research ethics committee and the university ethics committee. All participants gave written informed consent prior to undertaking the ultrasound test procedure.

Prospective patient group volunteers were diagnosed at either an orthopedic or hand clinic consultation (n=18) or by their general practitioner (n=3). Ten of the patient group reported undergoing electrodiagnosis (all reported as positive for CTS). Prior to undergoing the ultrasound test procedure, all patient group volunteers completed a Hand Symptom Diagram (HSD)23 and a Hand Symptom Severity and Functional Status Questionnaire (HSSFSQ).24 The Phalen test25 was conducted on the affected hand (or most severely affected hand in bilateral cases) by the principal researcher.

Ultrasound System Settings and Test Procedures

We adjusted the ultrasound system settings according to basic tissue Doppler principles26,27 and previous experience with this specific ultrasound system, including: (1) removing the wall (clutter) filters, (2) setting Doppler power and gain low (20% and 5dB, respectively), and (3) optimizing the scale of the Doppler velocity-time display for low-velocity measurements. General Doppler mode settings were as follows: frequency of 5.13MHz, electronic beam steering on (15°), and sample volume size set to 1mm. The Doppler angle was set to 0° for all tests regardless of beam-to-target angle (correction for actual Doppler angle was made offline later). All the above settings were held constant throughout the ultrasound scanning procedures.

The support jig was positioned on the side to be tested, which for patient group participants was the affected side, or the most severely affected side in bilateral cases. The participant lay supine on the plinth with the limb to be tested resting in the support jig such that the glenohumeral joint was at 45° of abduction and the forearm supinated (see fig 1). The elbow was initially positioned in either 90° of flexion or full extension dependent on test sequence (see below). The shoulder girdle was relaxed in a neutral position, and the head supported on a pillow in 0° rotation and side flexion. The length of the forearm component of the jig was adjusted such that the junction of the forearm support and hand plate lay between the proximal and distal wrist creases and held the wrist in 30° of extension. The hand of the participant was strapped to the hand plate using a self-adhesive (Velcro) strap across the metacarpals. The contralateral upper limb was relaxed by the side with the palm down and the participant was asked to maintain this position during all tests.

Test Protocols

The participant was asked to make a fist, ensuring that the fingers were fully flexed at the metacarpophalangeal and interphalangeal joints and the thumb positioned alongside the index finger. Nerve and tendon excursion was evoked by asking the participant to make a fist and then extend the fingers and thumb to the maximum allowed by the hand plate (figs 2A, 2C). The following 4 measurement protocols were conducted in the order shown for the first participant and then reversed for alternate participants to control for any sequencing effects of elbow position: (1) longitudinal excursion of the median nerve (elbow flexed); (2) longitudinal excursion of the flexor digitorum superficialis tendons (elbow flexed); (3) longitudinal excursion of the median nerve (elbow extended); and (4) longitudinal excursion of the flexor digitorum superficialis tendons (elbow extended).

Acquisition of Ultrasound Images

**Longitudinal excursion of the median nerve.** We placed the transducer longitudinally at the wrist such that a clear image of the median nerve and flexor tendons was acquired in this plane. Duplex Doppler mode was then selected (B/D mode) and the on-screen Doppler sample volume indicator adjusted to lie within the median nerve at approximately the level of the lunate-capitate intercarpal joint. The participant was asked to perform the test movement while the operator ensured that the sample volume was accurately located in the median nerve and a clear Doppler waveform was produced. After any necessary adjustments of the transducer or sample volume location, the test movement was performed and the resultant image containing the Doppler waveform was saved. The test was repeated on a further 2 occasions, so that 3 Doppler waveforms representing nerve excursion were saved (fig 2D).

Fig 1. Position of upper limb for test procedures (elbow flexed).
Longitudinal Excursion Measurements

The Doppler angle (angle subtended by the center of the Doppler beam and the target nerve or tendon) was determined for each test image by taking the mean of 3 measurements made with the angle calipers. The velocity-time integral (VTI) of the Doppler waveform (representing nerve or tendon excursion) was obtained by setting the lower threshold of ImageJ to 10, isolating the waveform, and using the wand tool to automatically calculate the VTI in pixels (appendix 1).

The maximum brightness of the Doppler waveform pixels was also determined for each image using the Min & Max Gray Value measurement option in ImageJ. We included this measurement to allow control for any effects on Doppler recorded excursion of intersubject variability in the strength of the received Doppler signal (due to variations in tissue echogenicity), which we suspected might be present (based on the findings of our validity studies).

Transverse Plane Displacement of the Median Nerve

We determined the axial location of the median nerve from the B/B-mode ultrasound images using the mean of measurements made from the skin surface to the deep and superficial borders at the central part of the nerve. Axial displacement was calculated as the difference in depth location of the nerve between the split-screen images. Lateral location of the median nerve was determined using the mean of measurements made from the left edge of each image on the split-screen to the lateral and medial borders of the nerve. Lateral displacement was calculated as the difference in lateral location between the split-screen images.

The VTI, axial, and lateral displacement in pixels for each image were input into an Excel spreadsheet and converted to millimeters according to the display scale.

Adjustment for Intrinsic Spectral Broadening and Transverse Plane Movements

The longitudinal excursion values were initially adjusted according to the measured Doppler angle, minus 22° beam-edge correction factor to take account of intrinsic spectral broadening,28,29 and adjusted for any concurrent transverse plane displacement in accordance with devised geometric formulas (appendixes 2–4).

Statistical Analysis

Descriptive and inferential statistics were calculated using Excel® and SPSS®, respectively.

RESULTS

Excluded Data Sets

Ultrasound test data on 2 of the 21 patients were excluded from subsequent analysis. One patient was unable to complete the test protocol satisfactorily due to dystonia affecting the wrist and hand, and 1 patient’s data was excluded because the HSD had strongly suggested ulnar nerve entrapment rather than CTS.

| Table 1: Summary Statistics of Group Age, Height, Weight, and BMI |
|------------------|------------------|------------------|
|                  | CTS Patient Group (n=19) | Control Group (n=37) |
| Characteristics  | Mean ± SD | Range | Mean ± SD | Range |
| Age (y)          | 57.4±14.7 | 35.0—86.0 | 48.0±10.0 | 21.0—64.0 |
| Height (m)       | 1.67±0.14 | 1.45—1.93 | 1.67±0.08 | 1.54—1.88 |
| Weight (kg)      | 75.9±14.8 | 53.0—102.0 | 69.8±11.1 | 51.0—102.0 |
| BMI (kg/m²)      | 27.0±3.8  | 22.1—35.6  | 25.0±3.8  | 18.7—39.8  |

Abbreviation: SD, standard deviation.
Intrarater Reliability

Intrarater reliability was tested using repeat measurements on 5 control group participants on 3 separate occasions. Intra-class correlation coefficient model 2,1 (ICC2,1)30,31 and standard error (SE) of measurement31,32 indicated a satisfactory level of reproducibility for the 4 nerve and nerve/tendon ratio excursion measurements as shown in table 2.

Transverse Plane Displacement of the Median Nerve

Mean axial and lateral displacement of the median nerve was 0.35±0.3mm, and 1.75±1.3mm, respectively. The direction of axial displacement, when it occurred, was from superficial to deep. Lateral displacement was predominantly toward the radial aspect of the limb (n=48).

Longitudinal Excursion of the Median Nerve and FDS Tendons

A summary of the nerve and tendon excursion data, adjusted for intrinsic spectral broadening and concurrent transverse plane movement is shown in table 3. The Kolmogorov-Smirnov test indicated that all excursion data sets were sufficiently normally distributed to enable parametric data analysis to be conducted (P range, .55–.98).33

Effect of Elbow Position

The mean nerve excursion values were significantly reduced for tests with the elbow positioned in extension compared with flexion (t test, P<.001), although some participants in both groups showed slightly greater nerve movements with the elbow extended (n=12). There was no significant effect of elbow position on FDS tendon movement.

Effect of Doppler Signal Strength on Excursion Measurements

A significant positive correlation was found between maximum brightness values of the Doppler waveform (produced by variations in Doppler signal strength) and nerve excursion measurements (r range, .48–.5, P<.001). The association was anticipated from provisional findings in validity studies, and the result confirmed the need to include maximum brightness value of the Doppler waveform as a covariate in subsequent analysis of nerve excursion data.

Between-Group Comparison of Nerve and Nerve/Tendon Ratio Excursion

The 4 sets of nerve and nerve/tendon ratio excursion data were analyzed individually using univariate analysis of variance, with age, height, BMI, and maximum brightness value of Doppler waveform input as possible covariates.33 The patient group showed significantly reduced excursion for 3 of the 4 measures; median nerve excursion with the elbow extended (P=.013), and both nerve/tendon ratios (elbow flexed, P=.019; elbow extended, P<.001). The patient group also showed a mean reduction in median nerve excursion with the elbow in flexion but this did not reach statistical significance (P=.089). With the exception of maximum brightness of the Doppler waveform (P=.002), none of the covariates (age, height, weight, BMI) showed a significant effect for any of the test measurements.

Discriminant analysis using the nerve/tendon excursion ratio data for the test with the elbow extended placed 11 (58%) of the 19 patients and 3 (8%) of the 37 control subjects in the lower nerve/tendon ratio category with a cutoff point of .25 (table 4).

Nerve/Tendon Ratio Association With Demographic and/or Diagnostic Characteristics

In the patient group, membership of the lower nerve/tendon ratio excursion group was significantly associated with dominance of the hand tested (Fisher exact test, P=.013). No other demographic or diagnostics characteristics were found to be significantly associated with nerve/tendon ratio group membership (table 5). The 3 control group subjects categorized as “low” nerve/tendon excursion ratio were women and had their dominant hand tested.

### Table 2: Intrarater Reliability of Nerve and Nerve/Tendon Ratio Excursion (n=5, occasions=3)

<table>
<thead>
<tr>
<th>Test Measurement</th>
<th>ICC2,1</th>
<th>SE of Measurement (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>95% CI</td>
</tr>
<tr>
<td>Median nerve (elbow flex)</td>
<td>.95</td>
<td>0.77–0.99</td>
</tr>
<tr>
<td>Median nerve (elbow ext)</td>
<td>.89</td>
<td>0.58–0.99</td>
</tr>
<tr>
<td>Nerve/tendon ratio (elbow flex)</td>
<td>.93</td>
<td>0.70–0.99</td>
</tr>
<tr>
<td>Nerve/tendon ratio (elbow ext)</td>
<td>.98</td>
<td>0.91–1.00</td>
</tr>
</tbody>
</table>

Abbreviation: CI, confidence interval.

### Table 3: Longitudinal Excursion of the Median Nerve, FDS Tendons, and Nerve/Tendon Ratio at the Carpal Tunnel During Finger and Thumb Extension

<table>
<thead>
<tr>
<th>Test Measurement</th>
<th>Patient Group (mm) (n=19)</th>
<th>Control Group (mm) (n=37)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ± SD</td>
<td>Range</td>
</tr>
<tr>
<td>Median nerve (elbow flex)</td>
<td>10.2±3.1</td>
<td>4.9–15.8</td>
</tr>
<tr>
<td>Median nerve (elbow ext)</td>
<td>8.3±2.6</td>
<td>4.2–13.7</td>
</tr>
<tr>
<td>FDS tendon (elbow flex)</td>
<td>37.1±6.5</td>
<td>23.9–49.6</td>
</tr>
<tr>
<td>FDS tendon (elbow ext)</td>
<td>37.3±7.1</td>
<td>22.1–50.9</td>
</tr>
<tr>
<td>Nerve/tendon ratio (elbow flex)</td>
<td>0.28±0.10</td>
<td>0.12–0.53</td>
</tr>
<tr>
<td>Nerve/tendon ratio (elbow ext)</td>
<td>0.23±0.06</td>
<td>0.10–0.34</td>
</tr>
</tbody>
</table>
The findings of this study provide the first direct evidence that reduced longitudinal excursion of the median nerve may be present in a substantial proportion of patients with CTS. Although the results are consistent with the indirect measurements reported by Valls-Solé et al., they appear to conflict with the results of the ultrasound study by Erel et al. Methodologic differences are the most likely explanation for this apparent discrepancy. Erel employed a smaller nerve excursion evoking protocol (metacarpophalangeal flexion), and nerve excursion was measured 5 to 15 mm proximal to the carpal tunnel. A particular distinguishing feature of the present study, however, is the inclusion of tendon excursion measurements, and the calculation of the nerve/tendon excursion ratio.

**Nerve/Tendon Excursion Ratio**

The magnitude of nerve or tendon excursion evoked by a joint movement is influenced by the distance of the nerve from the axis of the joint. Thus, for example, during wrist extension flexor tendons traversing the anterior aspect of the wrist are required to accommodate a relatively greater length change the further from the axis of movement they lie, and this biomechanical consideration may partly explain the positive association of CTS with a high wrist index. It is hypothesized that flexor tendons traversing the carpal tunnel in subjects with a high wrist index are subject to greater excursion during everyday activity, and are consequently at greater risk of developing tenosynovitis, which may predispose to CTS. Equally, it can be argued that greater nerve excursion occurring during everyday activities may predispose to CTS via a direct mechanical effect on the nerve or its specialized gliding tissue. In this regard, it is interesting to note that the mean FDS excursion was found to be greater in the patient group than the controls (see table 3), although this difference did not reach statistical significance.

If the magnitude of nerve excursion during everyday movements is contributing to CTS then it is possible that mean nerve excursion magnitudes in CTS patients may be above average before the onset of their condition. Therefore, attempts to detect any relative reduction of median nerve excursion in CTS compared with asymptomatic controls could be confounded. It was principally in order to apply some control for this possible anthropometric influence that we included nerve/tendon ratios as an outcome measure of this study. In this respect, it is of particular note that both nerve/tendon ratios were found to show significant between-group differences, but that one of the absolute nerve excursion measures (test with elbow flexed) showed no significant difference.

**Effect of Proximal Loading of the Median Nerve Tract**

The mean effect of positioning the elbow in extension was to reduce nerve excursion significantly compared with the same test with the elbow in flexion. On an individual basis, the effect was inconsistent, with some subjects in both patient and control groups showing minimal differences or a small increase in movement with the elbow extended. Previous studies have also identified apparent intersubject variability of effect of proximal loading on subsequent distal excursion, which we hypothesize may be related to individual variations in neural extensibility (e.g., the relative “slack” in the system). Scrutiny of significance level for between-group differences in nerve excursion and nerve/tendon ratio, however, suggest that proximal loading greatly increases the sensitivity of detecting reduced nerve excursion.

**DISCUSSION**

The magnitude of nerve or tendon excursion evoked by a joint movement is influenced by the distance of the nerve from the axis of the joint. Thus, for example, during wrist extension flexor tendons traversing the anterior aspect of the wrist are required to accommodate a relatively greater length change the further from the axis of movement they lie, and this biomechanical consideration may partly explain the positive association of CTS with a high wrist index. It is hypothesized that flexor tendons traversing the carpal tunnel in subjects with a high wrist index are subject to greater excursion during everyday activity, and are consequently at greater risk of developing tenosynovitis, which may predispose to CTS. Equally, it can be argued that greater nerve excursion occurring during everyday activities may predispose to CTS via a direct mechanical effect on the nerve or its specialized gliding tissue. In this regard, it is interesting to note that the mean FDS excursion was found to be greater in the patient group than the controls (see table 3), although this difference did not reach statistical significance.

If the magnitude of nerve excursion during everyday movements is contributing to CTS then it is possible that mean nerve excursion magnitudes in CTS patients may be above average before the onset of their condition. Therefore, attempts to detect any relative reduction of median nerve excursion in CTS compared with asymptomatic controls could be confounded. It was principally in order to apply some control for this possible anthropometric influence that we included nerve/tendon ratios as an outcome measure of this study. In this respect, it is of particular note that both nerve/tendon ratios were found to show significant between-group differences, but that one of the absolute nerve excursion measures (test with elbow flexed) showed no significant difference.

**Effect of Proximal Loading of the Median Nerve Tract**

The mean effect of positioning the elbow in extension was to reduce nerve excursion significantly compared with the same test with the elbow in flexion. On an individual basis, the effect was inconsistent, with some subjects in both patient and control groups showing minimal differences or a small increase in movement with the elbow extended. Previous studies have also identified apparent intersubject variability of effect of proximal loading on subsequent distal excursion, which we hypothesize may be related to individual variations in neural extensibility (e.g., the relative “slack” in the system). Scrutiny of significance level for between-group differences in nerve excursion and nerve/tendon ratio, however, suggest that proximal loading greatly increases the sensitivity of detecting reduced nerve excursion.

**Discriminant Analysis**

Although the discriminant analysis highlighted the fact that a substantially greater number of CTS patients (58%) exhibited “low” nerve/tendon excursion ratios compared with the control group (8%), it also indicates that if reduced nerve excursion does contribute to the pathophysiology of CTS it only does so in some cases. This suggests that studies investigating possible subtypes of CTS (e.g., compressive or uniaxial stretch related) may be a useful area of future work. The present study was not designed to identify possible subtypes of CTS, but the finding that a significantly greater number of subjects classified in the “low” nerve/tendon excursion ratio group had their dominant hand tested (affected hand, or most affected hand in bilateral cases) compared with patients classified as “high” nerve/tendon excursion ratio is interesting. Bay et al. reported that the nerve/tendon excursion ratio was unaffected by simulated carpal tunnel compression in cadaveric specimens, and concluded that if this ratio was indeed reduced in CTS patients that it was more likely to be related to disruption of the specialized gliding tissue surrounding the nerve at this location. If one assumes that the dominant hand is likely to be more active during everyday activities then the findings of our study lend support to these conclusions, and may also explain why Tuzuner et al. found no significant increase in median nerve excursion after endoscopic carpal tunnel decompression.

**Table 4: Discriminant Analysis Using the Nerve/Tendon Excursion Ratio (elbow extended)**

<table>
<thead>
<tr>
<th>Group</th>
<th>Nerve/Tendon Excursion Ratio</th>
<th>$\leq$ .25</th>
<th>$&gt;$ .25</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (n)</td>
<td>34</td>
<td>3</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Patient (n)</td>
<td>8</td>
<td>11</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Control (%)</td>
<td>81.9</td>
<td>8.1</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Patient (%)</td>
<td>42.1</td>
<td>57.9</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

**Table 5: Summary of Demographic and Diagnostic Characteristics of CTS Group by Nerve/Tendon Excursion Ratio**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Nerve/Tendon Ratio Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ratio $\geq$ .25 (n=8)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>3</td>
</tr>
<tr>
<td>Female</td>
<td>5</td>
</tr>
<tr>
<td>Dominance of tested hand</td>
<td></td>
</tr>
<tr>
<td>Dominant</td>
<td>2</td>
</tr>
<tr>
<td>Nondominant</td>
<td>6</td>
</tr>
<tr>
<td>Phalen sign</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>7</td>
</tr>
<tr>
<td>Negative</td>
<td>1</td>
</tr>
<tr>
<td>Electrodiagnosis reported as positive</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>5</td>
</tr>
<tr>
<td>No</td>
<td>4</td>
</tr>
<tr>
<td>Mean symptom severity (HSSFSQ)</td>
<td>3.1 $\pm$ 0.8</td>
</tr>
<tr>
<td>Mean functional status score (HSSFSQ)</td>
<td>2.7 $\pm$ 1.3</td>
</tr>
<tr>
<td>Mean age (y)</td>
<td>59 $\pm$ 15</td>
</tr>
<tr>
<td>Mean height (m)</td>
<td>1.67 $\pm$ 0.15</td>
</tr>
<tr>
<td>Mean weight (kg)</td>
<td>72 $\pm$ 14</td>
</tr>
<tr>
<td>Mean BMI (kg/m²)</td>
<td>26.7 $\pm$ 1.7</td>
</tr>
<tr>
<td>Mean duration of symptoms (mo)</td>
<td>16 $\pm$ 17</td>
</tr>
</tbody>
</table>

NOTE. Values are n or mean $\pm$ SD.
Study Limitations and Implications for Future Research

Doppler calculations of velocity and/or excursion require an accurate knowledge of the angle subtended by the Doppler beam and the target tissue.21 Because of the focused nature of the Doppler beam, calculating the Doppler angle with reference to the on-screen indicator only represents the Doppler angle at the center of the beam and does not take account of the maximum and minimum angles present at the beam edges. Therefore, excursion values calculated according to the method described here result in an overestimated value if this factor (intrinsic spectral broadening) is not taken into account.20,28,40,41 Furthermore, the effective edge of the Doppler beam is not a fixed quantity but is influenced by transducer beam geometry, ultrasound system settings (eg, Doppler gain and dynamic range), and the reflective characteristics of the target tissue (as evidenced in this study by the correlation found between maximum brightness on the Doppler waveform and excursion values). The ImageJ threshold level selected for excursion measurement employed in this study also influences the excursion calculated. A higher threshold effectively eliminates some of the weaker Doppler signals from the display, thereby indirectly affecting the actual Doppler gain and dynamic range, and the reflective characteristics of the target tissue (as evidenced in this study by the correlation found between maximum brightness on the Doppler waveform and excursion values). The ImageJ threshold level selected for excursion measurement employed in this study was based on string-phantom, porcine, and in vivo validity studies.9,11,44 We believe that the most likely explanation for this is because some wrist extension was occurring during the test movement due to incomplete fixation of the hand by the self-adhesive strap, and this would have led to higher evoked excursion.44 The mean nerve/tendon excursion ratios recorded for the control group in this study were consistent with those found previously in cadaveric studies.10,11

All ultrasound test protocols and image measurements were conducted by a single operator who was the principal researcher on the project. The researcher was blinded to group membership during the offline measurement of the saved images, but the possibility that experimenter bias was present during image acquisition (which was not blinded) cannot be completely ruled out although every effort was made to acquire Doppler waveforms in a consistent and unbiased manner.

Despite these limitations, the results of the study suggest that exploration of a causative role of reduced median nerve excursion in CTS, and the influence of mechanical interventions such as gliding exercises on nerve excursion, is warranted. Nerve excursion measurements in addition to conventional outcome measures would be key components of such study, as would diagnostic profiling of the patients, including electrodagnosis, compression-based tests (eg, Phalen), and tests with stretch-based components (eg, tethered median nerve stress test35,46 or upper-limb neurodynamic tests47,48). The profiling element could help identify CTS subtype characteristics that could lead to optimized case-by-case management of this condition.

CONCLUSIONS

This study has found that the mean ratio of median nerve and FDS tendon excursion at the wrist, evoked by finger movement, was significantly less in patients with CTS compared with asymptomatic controls. The results of this study indicate that reduced longitudinal excursion of the median nerve is present in a substantial proportion of CTS patients. The possibility that reduced excursion of the median nerve contributes to the pathophysiology in some cases, possibly a subtype, of CTS cannot be excluded, and this hypothesis merits further investigation.

APPENDIX 1: MEASUREMENT OF THE VTI

(A) Doppler waveform from a test image. (B) Using ImageJ, the lower display threshold has been set at 10 (upper at maximum), waveform has been isolated by cutting the baseline, and the area automatically measured using the wand tool.
APPENDIX 2: BEAM-EDGE CORRECTION METHOD

\[ \theta' \] is the Doppler angle taken from the center of the beam to the target (T). The effective minimum Doppler angle is \( \theta_L \) and the beam-edge correction factor is \( \eta \) (determined as 22° for the system and measurement set-up described in this study).

APPENDIX 3: CALCULATING THE EFFECT OF LATERAL MOVEMENT OF THE TARGET OCCURRING CONCURRENTLY WITH LONGITUDINAL MOVEMENT

\[ A = \sqrt{(C^2 - B^2)} \]

The target is assumed to move longitudinally in the direction A (T to Tₐ). If this movement is accompanied by lateral movement in the direction B, the actual vector of movement is given by C (T to Tₑ). Because the Doppler measurement method calculates the distance C, the true longitudinal movement (A) is given by \( \sqrt{(C^2 - B^2)} \). This correction assumes that concurrent lateral movement occurs in a consistent linear manner throughout the acquisition of the Doppler waveform.

APPENDIX 4: SCHEMATIC OF THE EFFECT ON DOPPLER VELOCIMETRY WHEN AXIAL MOVEMENT ACCOMPANIES LONGITUDINAL MOVEMENT OF THE TARGET

The assigned Doppler angle (\( \theta' \)) and assumed longitudinal velocity vector (\( e' \)) are incorrect if target movement is also axial (x). The actual Doppler angle and velocity vector with respect to the Doppler beam (d) are shown as \( \theta \) and \( e \) respectively. The angle \( \beta \) can be found by the equation:

\[ \beta = \arctan \left( \frac{f_0 \times \sin \alpha \cos \theta'}{\left( 770 f_0 s \right) - f_0 \times \sin \alpha \sin \theta'} \right) \]

and the true Doppler angle (\( \theta \)) = \( \theta' - \beta \).

References


The Role of the Back Rx Exercise Program in Diskogenic Low Back Pain: A Prospective Randomized Trial

Vijay B. Vad, MD, Atul L. Bhat, MD, Yasir Tarabichi


Objective: To determine the efficacy of the Back Rx program in patients with diskogenic low back pain (LBP).

Design: Prospective, randomized study.

Setting: Outpatient setting of a major university teaching hospital.

Participants: Subjects with LBP greater than leg pain for at least 3 months duration and no evidence of disk pathology. Fifty of 87 eligible patients consented and were randomized into age- and sex-matched groups.

Interventions: Group I participated in the Back Rx program for 15 minutes a day, 3 times a week. All patients, from both groups, received celecoxib (200mg) and hydrocodone (5mg) for 15 minutes before bedtime.

Main Outcome Measures: Roland-Morris Disability Questionnaire score, numeric pain rating score, patient satisfaction score, measured forward flexion, use of celecoxib, hydrocodone, and acetaminophen, time off work, and rate of symptom recurrence.

Results: At minimal 12-month follow-up, 70% of group I reported over 50% pain reduction with good or better patient satisfaction, compared with 33% in group II (P<.001). Average daily hydrocodone and acetaminophen use and time off work were less for group I (all, P<.05). Recurrence of symptoms at the end of the year was less for group I (P=.001).

Conclusions: Back Rx exercises, combined with use of a lumbar cryobrace and oral medications, yielded superior therapeutic results than with use of medications and cryobrace alone. Also significant was the reduced rate of recurrence in these patients.

Key Words: Exercise; Intervertebral disk; Low back pain; Rehabilitation.

LOW BACK PAIN (LBP) usually is considered to be a self-limiting condition that tends to improve over time.1-4 It also is among the leading causes of disability. A large variety of therapeutic interventions are available for the treatment of patients with LBP. The effectiveness of most of these interventions has not been shown beyond doubt, however. Consequently, the therapeutic management of these patients varies widely. Exercise is one therapy that is frequently prescribed for patients with LBP.5-23 It encompasses a wide array of interventions ranging from general physical fitness or aerobic exercise, flexibility, and stretching exercises, to strength training. Despite its frequent application, exercise therapy has not been shown to be more efficacious than other treatment modalities, especially in patients with acute LBP. In 1991, Koes et al8 sought to address the efficacy of exercise in LBP with a systematic review of 16 randomized controlled trials, most of which were considered to be of poor methodologic quality. No conclusions regarding the efficacy of exercise therapy compared with other conservative treatments could be drawn from this review and little evidence was found in favor of a specific type of exercise. In 1996, Faas19 published his own review of the matter in which he conducted a Medline search for randomized trials concerning exercise therapy in patients with back pain published from 1991 to 1995. Faas concluded that in acute back pain, exercise therapy is ineffective, whereas in subacute back pain, exercises with a graded activity program deserved attention, and in the case of chronic back pain, intense exercises may be promising. These equivocal findings prompted another systematic review of the literature by van Tulder et al24 to assess the effectiveness of exercise therapy for LBP with regard to pain intensity, functional status, overall improvement, and return to work. Their conclusions did not indicate that specific exercises are effective for the treatment of acute LBP, but rather that exercise may help patients with chronic LBP accelerate their return to normal daily activities and work. In subacute LBP populations, some evidence suggests that a graded-activity program improves absenteeism outcomes in occupational settings, although evidence for other types of exercise is unclear.24,25

It has been postulated that the degenerative process of the intervertebral disk evolves through 3 stages, namely, dysfunction, instability, and stabilization.25-30 Though each has distinct clinical and radiologic findings, any stage may coexist independently of another at any point along the entire lumbar axial skeleton. The initial or dysfunction stage is characterized by circumferential and radial tears within the annulus and synovitis of the zygapophyseal joints and typically presents in a younger age group. Diskogenic LBP constitutes a subgroup within the broad category of patients with LBP and these patients usually fall within the stage of dysfunction more than the stage of instability.

Successful treatment of subjects with the so called “diskogenic LBP” depends on making the specific diagnosis and merging of the biochemical and biomechanic etiologic constructs—again, for which there is no criterion standard treatment modality.
Back Rx is specifically designed as a lumbar stabilization program that restores flexibility, strength, and endurance while eliminating positions such as sitting and forward flexion that may increase intradiskal pressure and in turn lead to diskogenic LBP.

The purpose of the present study is to assess the efficacy of the Back Rx exercise program when coupled with oral medication and a back brace, as compared with the use of medication and a lumbar brace alone in patients with axial, diskogenic LBP. The hypothesis the authors set out to test was that the Back Rx exercise regimen can decrease pain as well as reduce recurrence of pain in patients with subacute or chronic diskogenic LBP.

METHODS

Approval for this study was obtained from the institutional review board of the Hospital for Special Surgery. Inclusion criteria included: symptoms of LBP greater than leg pain of at least a 3-month duration, exacerbation of pain with sitting and alleviation with walking, and magnetic resonance imaging (MRI) documented evidence of disk pathology (eg, disk protrusion or extrusion on a T2-weighted sagittal image without any central and/or foraminal stenosis or degeneration of the facet joints). Patients were excluded if they had a recent history of trauma, prior history of lumbar spinal surgery, or had undergone any recent spinal interventional procedures. Similarly, subjects with pending legal claims or worker’s compensation claims were excluded. Of the 87 patients assessed, 65 met the aforementioned criteria, and 50 consented to be enrolled in this prospective study. Fifteen of the 65 who met the inclusion criteria were unwilling to commit themselves to a regular home-exercise program and/or to come in for compliance monitoring.

We then randomized the patients into 2 groups matched for age and sex. Subjects from group I (n=25) participated in the Back Rx program for at least 15 minutes a day, 3 times a week. Exercises were done based on a real-time Back Rx DVD handed to patients and subjects were given a calendar to mark the days so as to monitor self-compliance. Further compliance was monitored by the principal author at timely intervals of 3 weeks, 6 weeks, 3 months, 6 months, and at the end of 12 months. This was undertaken by means of a face-to-face interview of at least 20 minutes duration. The overall compliance rate was 91%.

Patients from both groups used a lumbar cryobrace for 15 minutes before bedtime daily. Medications permitted in both groups included up to 200mg of celecoxib per day, as well as 5mg of hydrocodone with 500mg of acetaminophen for breakthrough pain as needed.

The Back Rx program progresses through series A, B, and C, all of which develop flexibility, strength, and endurance with elements of physical therapy and rehabilitation, yoga, and Pilates. The yoga- and Pilates-based elements in the program were modified to exclude exercises that may easily traumatize a weak back. Positions that increase intradiskal pressures by forcing patients to sit and bend forward, for instance, were either modified or ruled out. Patients in group I underwent 6 months of series A exercise regimen, followed by at least 6 months of series B.

Series A emphasizes isometric muscle work derived mostly from physical therapy. Series B builds on series A by including more dynamic muscle movements, as well as more yoga-based exercises that intensify the isometric loading of the core muscles of the back. Other targeted areas in both series include: chest, shoulder, abdominal, thigh, and full hip musculature.

We monitored patients for a minimum of 12 months. Outcome measures included Roland-Morris Disability Questionnaire (RMDQ) score, numeric pain rating score, patient satisfaction score, and the finger-to-floor distance during forward flexion with knees extended. Time taken off from work, medication usage, and the recurrence of symptoms between both groups were monitored as well. A successful outcome was defined as greater than 50% pain reduction with good or better patient satisfaction.

RESULTS

Two patients from group I and 4 from group II received spinal epidural injections during the study duration, and were subsequently excluded from the final data analysis. After 1 year, group I (n=23) was composed of 11 men and 12 women, and group II (n=21) had 10 men and 11 women (fig 1). The average age was 31.4 years for group I and 30.9 for group II. Patients from both groups were younger and almost all had sedentary jobs with excessive sitting, which may explain early onset of diskogenic pain. The subjects had isolated diskogenic etiology without any associated disk degeneration or lumbar facet arthritis. None had a history of trauma to the lumbar spine. Comparisons of RMDQ scores, pain scores, forward flexion, and patient satisfaction at different time periods are shown in tables 1 through 4, respectively. The Wilcoxon signed-rank test was used for statistical analysis. At the minimal 12-month follow-up period (range, 12–15mo), 70% of the patients in group I reported a successful outcome, as compared with only 33% in group II (P=.001). During this 12-month duration, 48% of the subjects in group II reported a recurrence of acute symptoms lasting for more than 3 days, as compared with only 17% in group I (P=.001). The overall average daily use of hydrocodone with acetaminophen and time off work for group I were statistically less (all, P<.05) when compared with...

Fig 1. CONSORT flow diagram of subjects through the trial.
group II (tables 5, 6). There was no statistically significant difference between the groups for the average usage of celecoxib (table 7). This portion of the intervention (medication usage) is treated as a dependent measure in the statistical analysis as it was left at the discretion of the subjects themselves.

### DISCUSSION

LBP is a multifactorial disorder with many possible etiologies. Its lifetime prevalence ranges from 65% to 85%, despite all efforts expanded into its prevention, treatment, and rehabilitation. Though the majority of patients with acute LBP improve over time, a few continue to experience symptoms that can lead to absenteeism from work, extra expenses, and disability. Published findings indicate that this condition tends to relapse, with 28% to 75% of patients experiencing multiple episodes with persistent pain. Dishogenic etiology has been implied in 26% to 39% of patients with LBP. In the past 15 years, the treatment options for axial dishogenic LBP have almost reversed, progressing from lumbar fusion for eliminating motion to an artificial disk replacement aimed at maintaining maximal intersegmental flexibility to physiologic loads. The precariousness of clinical treatment preferences with respect to LBP is obvious from the limited quantity and quality of scientific evidence, for example, the scarcity of conclusive studies addressing the efficacy of exercise. The Back Rx program was structured in a way that allows patients to gradually and comfortably develop flexibility, strength, and endurance. As already mentioned, many Pilates- and yoga-based elements in the program that may easily traumatize a weak back were modified accordingly. The program starts off in an accessible manner: patients conduct the first portion of series A lying flat on their backs, allowing them to minimize pressures on potentially deconditioned and injured muscles and disks. Exercises in the first series consist of isometric work derived predominantly from physical therapy, allowing patients to lay a foundation of core muscle flexibility and prepare their body for increased strength and endurance training. When ready, patients may progress to the slightly more challenging yet potentially more therapeutic series B. Series B consists of more yoga-based work that intensifies the isometric loading of the core muscles, as well as more dynamic muscle work that builds strength through concentric, eccentric, and plyometric contractions. Throughout all 3 series in the program, patients are asked to focus on their breathing—an essential aspect of both yoga and Pilates. By doing so, patients can pace themselves appropriately as well as potentially modulate their pain to make the exercises both easier to do and more therapeutic.

The overall efficacy of the Back Rx program was evident in the finding that, at 1 year, 70% of the group that had participated in the therapy reported a successful outcome, as compared with only 33% in the other group ($P = .001$). Furthermore, the average time off from work and daily medication usage were both significantly lower in the patients that participated in the exercise program (all, $P = .005$).

Perhaps the most noteworthy outcome in our trial, however, is the statistically lower rate of recurrence of acute symptoms in the group that enrolled in the Back Rx program ($P = .001$). Only 17% of these patients experienced a recurrence of symptoms, compared with 48% in our second group and 28% to 75% previously reported in the literature. The aforementioned are all welcomed outcomes with regard to a health condition that has taken a significant toll on the health care system, with disability from LBP rising exponentially over the past 5 decades.
It should be noted, however, that patients in the program experienced mildly increased symptoms for the first 3 weeks after initiation of the *Back Rx* exercises (see table 2). Though this effect is transient, it is imperative that patients be warned that exercise therapies such as this one may slightly increase their discomfort before potentially having a long-term therapeutic effect.

Now to discuss the important aspect of reliability and validity issue surrounding the diagnosis of diskogenic pain. The ideal tool for the diagnosis of diskogenic LBP pain should have clear applications, produce valid and reproducible results, and be free of complications. It must be sensitive with a low false-positive rate and specific with a low false-negative rate. Some have found MRI to be as good as diskography and even preferable because of its noninvasive nature, whereas proponents of lumbar diskography contend that pain provocation by intradiskal injection is the only method that can determine which disk is responsible for a patient’s symptoms. This group also maintains that the diskography image can show lesions not revealed by other methods.5,8,32,33

**Study Limitations**

One of the limitations of this study was that provocative lumbar diskography was not used to confirm the diagnosis of diskogenic LBP. Lumbar diskography serves the single purpose of identifying the painful disk and is a physiologic evaluation consisting of a volumetric, manometric, radiographic, and pain-provocative challenge. Throughout the literature, lumbar diskography has been found to be a useful diagnostic tool but at the same time has been criticized for its shortcomings. Patients with no history of lumbar pain who had undergone posterior iliac crest bone graft harvesting for nonlumbar procedures have often experienced a concordant painful sensation during lumbar diskography. Thus the ability of a patient to separate spinal from nonspinal sources of pain on diskography is questioned, and a response of concordant pain on diskography may be less meaningful than often assumed. The diskogram is a tool and does have certain clear limitations.8,31,37

Because the procedure assesses a subjective complaint of pain, it may be subject to false-positive responses. Furthermore, diskography by itself is painful, because it is ideally performed without any sedation, to optimize the patient response. The limitations of diskography are its invasive nature, moderate radiation exposure during fluoroscopy, and the potential complications including the remote risk for disk-space infection. For these reasons, it was not included as a mandatory inclusion criterion for this study and we chose to presume the diagnosis of diskogenic LBP based on symptoms (back pain greater than leg pain), physical examination (forward flexion then extension recreated usual symptoms), and MRI.

**CONCLUSIONS**

These preliminary results suggest that a well-designed exercise program combined with use of a back cryobrace and oral medications may yield superior results for patients with axial diskogenic LBP when compared with oral medications and back cryobrace alone supporting the hypothesis set forth. Such a program, when done routinely with monitoring of compliance, may lessen chances of recurrence of acute LBP episodes, medication use, and time off work. A large-scale multicenter controlled trial should be undertaken for the further evaluation of our findings.

**References**

Association of Mobility Limitations With Health Care Satisfaction and Use of Preventive Care: A Survey of Medicare Beneficiaries

Jeanne M. Hoffman, PhD, Anne Shumway-Cook, PhD, PT, Kathryn M. Yorkston, PhD, Marcia A. Ciol, PhD, Brian J. Dudgeon, PhD, Leighton Chan, MD, MPH


Objective: To examine the association between satisfaction with health care, the use of preventive health care, and mobility limitation.

Design: Cross-sectional analysis of survey data.

Setting: Community.

Participants: A total of 12,769 people, age greater than 65, who participated in the 2001 Medicare Current Beneficiary Survey.

Interventions: Not applicable.

Main Outcome Measures: Self-report of mobility limitation, satisfaction, and use of preventive health care (immunizations, cancer screening). Sampling weights were used in all analyses, including logistic regression for survey data, to calculate estimates for a Medicare population of 31 million.

Results: After controlling for sociodemographic characteristics, Medicare beneficiaries with mobility limitations were significantly more dissatisfied with their health care compared with beneficiaries without mobility limitations. Receipt of preventive care did not differ for those with and without mobility limitation on some preventive services.

Conclusions: Mobility limitation is highly associated with dissatisfaction with health care among older adult beneficiaries. Although Medicare beneficiaries may receive similar rates of preventive care, those with mobility limitation may have more difficulty accessing services and be more dissatisfied with their health care in general.

Key Words: Disabled persons; Medicare; Patient satisfaction; Rehabilitation.

© 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

STUDIES HAVE SUGGESTED that physical disability is a risk factor for dissatisfaction with health care and access to preventive services, such as mammograms and Papanicolaou (Pap) smears for women. Although people with limitations in activities of daily living (ADLs) may access specialty care to address their specific disabling condition, access to preventive care may be found to be limited. Iezzoni et al examined access to care for people who participated in the National Health Interview Survey (ages ≥18 y) and found that those people with disability were less likely to receive preventive services. Chan et al found that Medicare beneficiaries with ADL limitations were at risk for not receiving mammograms and Pap smears, but that ADL limitations did not impact access to immunization. These researchers postulate that reduced access may be due to a variety of factors including providers who focus on the disabling condition to the exclusion of preventive services as well as transportation and accessibility issues. In addition, some health care providers may believe that such people do not merit these procedures.

Satisfaction with health care has also been studied in the estimated 54 million Americans with disability. This issue is important because dissatisfaction with care has also been associated with noncompliance and poor follow-up. Dissatisfaction has been assessed by questioning respondents’ level of satisfaction with a variety of health care issues including overall quality, access, and out-of-pocket costs. Results suggest that, for Medicare beneficiaries, dissatisfaction increased as the level of ADL disability or functional limitations increased. Ease of getting to doctors, follow-up care, and access to specialists were the areas of most dissatisfaction for disabled older adults compared with those without ADL disability. In those with communication disable, results were similar except for less dissatisfaction with ease of getting to doctors.

One key factor that impacts access to health care is mobility, which has not been examined for its impact on dissatisfaction and access to preventive care. Mobility, defined as the ability to move independently around the environment, is an essential part of personal and instrumental ADLs. Mobility limitation significantly restricts participation and can lead to social isolation, anxiety, and depression in older adults, and is also the strongest predictor of self-perceived disability. National estimates suggest that functional mobility limitation is a significant problem for many older adults, and is associated with some modifiable characteristics. Among older adults, a decline in mobility often precedes the onset of disability in activities of daily life and may be the most important limitation in ADLs, contributing to dissatisfaction and lack of access to preventive care. Restrictions in the ability to walk may place older adults at risk for less than optimal health care due to reduced access to needed services. This in turn may adversely impact patient satisfaction with health care.

The purpose of this study was to estimate the extent to which mobility limitation is associated with satisfaction with care and...
access to preventive services. If mobility limitation is similar to other types of disability, including ADLs and communication disability, it would suggest that presence of any type of disability is an important factor that needs to be considered by all health care providers to increase both access to preventive care and satisfaction with care. In addition, because mobility disability is usually the first activity limitation that a patient confronts, a strong relationship between mobility and satisfaction would suggest that a patient’s satisfaction is affected soon after they become disabled. Finally, mobility limitation is often a reversible problem. If it is related to satisfaction and access, perhaps addressing a patient’s mobility directly through treatment, adaptation or through the use of equipment, may lead to improvements in health care and quality of life.

METHODS

Sampling Frame
Our analyses examined 12,769 Medicare beneficiaries (≥65y) sampled in the 2001 Medicare Current Beneficiaries Survey (MCBS). The MCBS is a longitudinal survey of the Medicare population publicly available from the Centers for Medicare and Medicaid Services through a data use agreement. It uses a complex, multistage, stratified sampling design to obtain a nationally representative sample of all Medicare beneficiaries. The United States is separated by counties into 107 primary sampling units, and then further divided into clusters by postal ZIP codes. Beneficiaries within each cluster are selected by systematic random sampling within age groups.

Beneficiaries participating in the survey are interviewed in person and data are collected on a wide variety of items including use of health services, satisfaction with medical care, health status, and functioning, as well as demographic information such as income, education level, and living arrangements. Interviews with beneficiaries not living at home (eg, in skilled nursing facilities) were excluded.

Preventive Care
Survey respondents were asked a series of questions related to receipt of preventive care. The questions included whether a person had received a flu shot the previous winter and whether they had ever received a pneumonia shot. All women were asked if they had received a mammogram, and women under the age of 71 if they had received a Pap smear in the last year. Men were asked if they had received a digital rectal prostate exam and a blood test for prostate cancer.

Dissatisfaction
The MCBS includes 10 items assessing satisfaction with health care: overall quality, availability of services during off hours, ease of access, out-of-pocket costs, information provided, follow-up care, physician’s concern, receiving care at 1 location, availability of specialists, and ease of obtaining answers on the telephone. Respondents were asked to rate their satisfaction as: 1, very satisfied; 2, satisfied; 3, dissatisfied; or 4, very dissatisfied. For all analyses, coding was collapsed into: satisfied (1, 2) and dissatisfied (3, 4).

Mobility Limitation
We defined and validated mobility limitation through a process described previously. Briefly, beneficiaries were categorized by their responses to 4 questions: Do you have any difficulty walking? Do you have difficulty walking a quarter mile—2 to 3 blocks? Do you need equipment to walk? Do you need personal assistance to walk? A person reporting no difficulty walking was classified as having no mobility limitation. A person reporting difficulty walking, but requiring no equipment or personal help or had difficulty walking 2 to 3 blocks was classified as having mild mobility limitation. A person reporting difficulty walking and who used equipment, but had no need of personal assistance was classified as having moderate mobility limitation, and a person needing personal assistance was classified as severely limited. A person reporting that they “do not walk” was classified separately and those people are presented as a group. Although specific causes of walking difficulty could not be determined, all were self-reported as being health related.

Sociodemographic Characteristics
Sociodemographic characteristics for the population are presented in table 1. Race was categorized into those who reported being white or nonwhite. Marital status was dichotomized to compare those currently married with everyone else (single, divorced, widowed). Socioeconomic status was categorized into income above versus below or equal to $25,000. Education was categorized as having less than graduation from high school compared with high school graduates or further education. Respondents living alone were compared with those living with others. Rural or urban status was determined by the county in which a respondent lived. Urban locations were defined by metropolitan statistical areas, which categorize counties based on whether they contain a city with a population greater than 50,000. Self-report of general health (reporting fair/poor health vs good/excellent health), and whether a person was currently a smoker (vs not) were also included. The number of comorbidities was determined by a count of the number of clinical conditions a respondent endorsed. Body

---

**Table 1: Weighted Estimates of Proportion of Sociodemographic and Clinical Characteristics for 2001 Medicare Beneficiaries**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Proportion (95% Confidence Limits)</th>
<th>Estimated Population Size (millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobility limitation*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None (n=6421)</td>
<td>53.3 (52.0–54.5)</td>
<td>16.4</td>
</tr>
<tr>
<td>Mild (n=4124)</td>
<td>31.4 (30.3–32.4)</td>
<td>9.7</td>
</tr>
<tr>
<td>Moderate (n=1496)</td>
<td>10.5 (9.9–11.2)</td>
<td>3.3</td>
</tr>
<tr>
<td>Severe (n=519)</td>
<td>3.6 (3.2–4.0)</td>
<td>1.1</td>
</tr>
<tr>
<td>Does not walk (n=177)</td>
<td>1.2 (0.9–1.4)</td>
<td>0.4</td>
</tr>
<tr>
<td>Mean age (y)</td>
<td>75.3 (75.2–75.4)</td>
<td></td>
</tr>
<tr>
<td>Mean no. of comorbidities</td>
<td>2.5 (2.5–2.6)</td>
<td></td>
</tr>
<tr>
<td>Sex (% women)</td>
<td>57.8 (57.0–58.6)</td>
<td>17.8</td>
</tr>
<tr>
<td>Race (% nonwhite)</td>
<td>13.3 (12.2–14.4)</td>
<td>4.1</td>
</tr>
<tr>
<td>Marital status (% married)</td>
<td>55.4 (54.2–56.6)</td>
<td>17.1</td>
</tr>
<tr>
<td>SES (% ≤$25,000)</td>
<td>59.2 (57.8–60.6)</td>
<td>18.3</td>
</tr>
<tr>
<td>Education (% &lt;12y)</td>
<td>31.5 (30.1–32.9)</td>
<td>9.7</td>
</tr>
<tr>
<td>% living alone</td>
<td>31.7 (30.7–32.6)</td>
<td>9.8</td>
</tr>
<tr>
<td>General health (% fair/poor)</td>
<td>22.8 (21.8–23.7)</td>
<td>7.0</td>
</tr>
<tr>
<td>Current smoker (%)</td>
<td>11.0 (10.4–11.7)</td>
<td>3.4</td>
</tr>
<tr>
<td>Rural/urban status (% rural)</td>
<td>23.2 (20.6–25.8)</td>
<td>7.2</td>
</tr>
<tr>
<td>Participates in HMO (%)</td>
<td>18.7 (17.2–20.2)</td>
<td>5.8</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% underweight</td>
<td>5.3 (4.8–5.8)</td>
<td>1.6</td>
</tr>
<tr>
<td>% normal</td>
<td>74.8 (73.9–75.7)</td>
<td>23.1</td>
</tr>
<tr>
<td>% overweight</td>
<td>19.9 (19.0–20.7)</td>
<td>6.1</td>
</tr>
</tbody>
</table>

*The n value is the number in the survey.

Abbreviations: BMI, body mass index; HMO, health maintenance organization; SES, socioeconomic status.
mass index (BMI) was classified as overweight (BMI >30 kg/m²), underweight (BMI <18.5 kg/m²), or in the normative range (BMI range, 18–30 kg/m²). Participation in a health maintenance organization (HMO) was also included.

**Statistical Analysis**

The MCBS survey must be weighted to make inferences about the entire Medicare population. The probability of being included in the survey is not the same for all beneficiaries but is determined by a beneficiary’s address and age. Use of the sampling weights minimized the sampling error and makes the estimates for the population more accurate. Sampling weights were used to calculate estimates of means (for continuous variables) and proportions (for categorical variables) of the entire Medicare population by each mobility limitation category.

We studied the association between the 2 levels of dissatisfaction (satisfied, dissatisfied) or preventive care (yes, no) and mobility limitation using logistic regression. Because the level of dissatisfaction may be influenced by both person and environmental characteristics, the model included the following covariates: age, sex, race, marital status, socioeconomic (income and education) status, living arrangement, rural or urban status, self-perceived health, number of comorbid conditions, smoking status, BMI, and HMO participation. The results of logistic regression analysis are reported as odds ratios with no mobility limitation as the comparison group. All analyses were performed using SAS PROC SURVEYLOGISTIC, which takes into account the complex sampling design and weighting of the MCBS.

**RESULTS**

**Preventive Care and Dissatisfaction With Health Care by Level of Mobility Disability**

Table 2 shows the estimated use of preventive care for different levels of mobility limitation as well as the 95% confidence intervals. Results suggest that there is little variability in immunization rates. It is noteworthy, however, that more than 30% of respondents are not receiving immunizations regardless of mobility category. For women, the proportion of mammograms decreased as the level of mobility limitation increased, with those who do not walk reporting the lowest rates. For women under age 71, the proportion of Pap smears was below 50% for all categories. For men, the proportion of blood tests for prostate cancer and digital rectal prostate exams declined as mobility limitations increased; on average 30% of the

![Table 2: Weighted Estimate of Proportion of Preventative Care and Dissatisfaction With Health Care Reported by Respondents by Level of Mobility Limitation](image-url)

NOTE. Values are percent and 95% confidence interval (CI).
men in this sample did not receive blood tests for prostate cancer screening and 45% did not receive digital rectal prostate exams.

Table 2 also shows the estimated proportion of respondents reporting dissatisfaction at different levels of mobility limitation. Results suggest that mobility limitation is strongly associated with dissatisfaction. Dissatisfaction with overall quality of health care and ease of getting to the doctor increased as the level of mobility limitation increased, with those who did not walk reporting the most dissatisfaction. In addition, for people with mobility limitation, dissatisfaction increased with level of severity for getting needs met at 1 location and concern of doctors for overall health, with those who do not walk reporting less dissatisfaction. For those who do not walk, rates varied, but were consistently above those who had no mobility limitation. Dissatisfaction with out-of-pocket costs for health care was highest for all groups ranging from 12.0% to 20.2%.

**Adjusted Odds Ratios for Access to Preventive Care and Dissatisfaction With Health Care**

Table 3 presents adjusted odds ratios for preventive care for each level of mobility limitation compared with those having no mobility limitation. The model of association included age, sex, race, marital status, income, education, living arrangement, health status, smoking status, number of comorbidities, BMI, rural or urban status, and participation in an HMO. This table displays the odds that a person in a certain mobility limitation category received preventive care compared to a person without mobility limitation when the values of all other variables are the same for each person. Given this, the level of mobility impairment was not associated with receipt of flu vaccines or rectal prostate exam. In addition, respondents with mild mobility limitations had higher odds and those with severe mobility limitation had lower odds of receiving pneumonia vaccinations compared with those with no mobility limitation.

Women with any mobility limitation had significantly lower odds of receiving mammograms. For women under age 71 there was no association between mobility limitation and receipt of Pap smears. For men, those with mild and moderate mobility limitation had significantly lower odds of receiving

<table>
<thead>
<tr>
<th>Items</th>
<th>Significance Level of Mobility Limitation</th>
<th>Mobility Limitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preventive care items</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Received a flu shot in the past year.</td>
<td>.10</td>
<td>Mild 0.90 (0.85–1.05) Moderate 0.88 (0.67–1.16) Severe 0.72 (0.50–1.15) Does Not Walk 1.08 (0.93–1.27)</td>
</tr>
<tr>
<td>2. Received a pneumonia shot in the past year.</td>
<td>&lt;.001</td>
<td>Mild 1.13 (1.02–1.26) Moderate 0.76 (0.59–0.98) Severe 0.77 (0.49–1.21) Does Not Walk 1.08 (0.93–1.27)</td>
</tr>
<tr>
<td>3. Had a mammogram in the past year.</td>
<td>&lt;.001</td>
<td>Mild 0.83 (0.73–0.95) Moderate 0.38 (0.28–0.51) Severe 0.25 (0.14–0.43) Does Not Walk 0.76 (0.59–0.98)</td>
</tr>
<tr>
<td>4. Had a Pap smear in the past year (women &lt;71 only).</td>
<td>.06</td>
<td>Mild 0.71 (0.53–0.95) Moderate 0.59 (0.26–1.32) Severe 0.60 (0.22–1.64) Does Not Walk 0.59 (0.26–1.32)</td>
</tr>
<tr>
<td>5. Had a blood test for prostate cancer in the past year.</td>
<td>.02</td>
<td>Mild 0.80 (0.68–0.94) Moderate 0.69 (0.45–1.06) Severe 0.53 (0.28–1.01) Does Not Walk 0.69 (0.45–1.06)</td>
</tr>
<tr>
<td>6. Had a digital rectal prostate exam in the past year.</td>
<td>.33</td>
<td>Mild 0.92 (0.80–1.05) Moderate 0.87 (0.59–1.28) Severe 0.64 (0.36–1.13) Does Not Walk 0.76 (0.59–0.97)</td>
</tr>
<tr>
<td>Dissatisfaction items</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. The overall quality of the medical services you have received in the last year.</td>
<td>&lt;.001</td>
<td>Mild 1.44 (1.08–1.92) Moderate 2.56 (1.62–4.05) Severe 2.51 (1.45–4.35) Does Not Walk 2.56 (1.62–4.05)</td>
</tr>
<tr>
<td>2. The ease and convenience of getting to a doctor from where you live.</td>
<td>&lt;.001</td>
<td>Mild 1.23 (0.94–1.59) Moderate 2.30 (1.56–3.78) Severe 2.27 (1.36–3.82) Does Not Walk 2.15 (1.63–2.82)</td>
</tr>
<tr>
<td>3. Getting all your medical care needs taken care of at the same location.</td>
<td>&lt;.001</td>
<td>Mild 1.40 (1.10–1.77) Moderate 2.58 (1.75–3.81) Severe 1.73 (0.76–3.95) Does Not Walk 2.58 (1.75–3.81)</td>
</tr>
<tr>
<td>4. The concern of doctors for your overall health rather than just for an isolated symptom or disease.</td>
<td>&lt;.001</td>
<td>Mild 1.73 (1.38–2.18) Moderate 2.51 (1.73–3.65) Severe 1.28 (0.60–2.71) Does Not Walk 1.88 (1.44–2.46)</td>
</tr>
<tr>
<td>5. The availability of medical services at night and on weekends.</td>
<td>&lt;.001</td>
<td>Mild 1.49 (1.11–2.02) Moderate 2.34 (1.45–3.76) Severe 2.17 (1.14–4.10) Does Not Walk 2.05 (1.49–2.82)</td>
</tr>
<tr>
<td>6. The out-of-pocket cost you paid for medical services.</td>
<td>&lt;.001</td>
<td>Mild 1.18 (1.02–1.37) Moderate 1.62 (1.22–2.17) Severe 0.98 (0.62–1.55) Does Not Walk 1.62 (1.33–1.98)</td>
</tr>
<tr>
<td>7. The information given to you about what was wrong with you.</td>
<td>&lt;.001</td>
<td>Mild 1.75 (1.39–2.20) Moderate 2.70 (1.83–3.98) Severe 2.28 (1.32–3.92) Does Not Walk 2.38 (1.77–3.22)</td>
</tr>
<tr>
<td>8. The follow-up care you received after an initial treatment or operation.</td>
<td>&lt;.001</td>
<td>Mild 1.82 (1.31–2.52) Moderate 4.37 (2.62–7.29) Severe 2.46 (1.06–5.69) Does Not Walk 2.89 (2.01–4.16)</td>
</tr>
<tr>
<td>9. The availability of care by specialists when you feel you need it.</td>
<td>&lt;.001</td>
<td>Mild 1.45 (1.07–1.95) Moderate 2.28 (1.39–3.75) Severe 1.58 (0.58–4.28) Does Not Walk 1.48 (1.48–3.20)</td>
</tr>
<tr>
<td>10. The ease of obtaining answers to questions over the telephone about your treatment or prescriptions.</td>
<td>&lt;.001</td>
<td>Mild 1.43 (1.14–1.78) Moderate 2.29 (1.59–3.28) Severe 1.83 (1.06–3.16) Does Not Walk 1.99 (1.47–2.71)</td>
</tr>
</tbody>
</table>

NOTE: Values are odds ratio and 95% CI. Statistical significance of the mobility variable overall is given in the first column. Statistically significant odds ratios are shown in boldface. All analyses control for age, sex (where appropriate), ethnicity, marital status, education, socioeconomic status, living arrangement, rural status, perceived health, number of comorbid conditions, smoking status, BMI, and HMO participation.
leads to reduced access to preventive care. Mobility limitation as a factor in quality of life and is found in nearly 15 million older adults who did not walk (n = 177) were similar in reports of dissatisfaction to those without mobility limitation in out-of-pocket costs, concern of doctors with overall health, getting all medical care at the same location, and availability of care by specialists. This may be due to the small sample size of the group and to the variability in responses seen in the confidence intervals. These adults may also have compensated for any mobility limitation that interfered with their health care, however.

**DISCUSSION**

Results of the current study show that 46.7% of Medicare beneficiaries have some level of mobility limitation. Although increasing mobility limitation was not associated with immunization for flu or pneumonia, it is associated with undergoing fewer mammograms for women and having fewer blood tests for prostate cancer for men. In addition, mobility limitation was highly related to dissatisfaction with a wide range of health care domains, from quality of care, to information received, even after controlling for sociodemographic characteristics. People who were categorized as nonwalkers showed some similarities to those with no mobility limitation in dissatisfaction, which suggests that this group is different from those with other levels of mobility limitation. Perhaps this is due to the fact that these people may have compensated for their significant limitations through the use of wheelchairs. Although mobility limitation did not have a consistent relationship with use of preventive care, it is notable that up to 35% of older adults did not receive flu shots and up to 40% did not receive pneumonia vaccination. In addition, for women under age 71, receipt of Pap smears was below 50% even for those with no mobility limitation suggesting that this group is different from those with other levels of mobility limitation. Perhaps this is due to the small sample size of the group and to the variability in responses seen in the confidence intervals. These adults may also have compensated for any mobility limitation that interfered with their health care, however.

**Study Limitations**

This study has limitations. First, only community-dwelling Medicare beneficiaries, 65 years and older, were included and therefore we cannot generalize to persons with mobility limitation under age 65 or those living in skilled nursing facilities. In addition, we may have been unable to account for the influence of all other adults' expectations. People who report less satisfaction with health care may also report less satisfaction with their life in general. People with psychologic distress may be less satisfied with their care because they cannot get relief from their distress or because providers may mismanage them. We attempted to control for this potential bias by accounting for a person's perceived health. It is possible, however, that some of our findings were due to psychologic distress and general life dissatisfaction in the persons with mobility disabilities.

Future research is needed to examine the group of older adults who "did not walk" in order to determine whether they differ significantly from others with mobility limitation or have been able to compensate for their limitation through the use of wheelchairs or other assistive technology. Study of this group may provide ideas and suggestions for potential interventions that would improve the satisfaction with care and access to preventive services for those with less significant mobility limitation. At present, Medicare provides funding for wheelchairs only if they are required to function in the home setting. If improved accessibility leads to improved access to health care and satisfaction, review of this policy may be needed.

**CONCLUSIONS**

Our results suggest that mobility limitation is associated with less use of some preventive care and highly associated with dissatisfaction with health care. Mobility limitation is a major factor in quality of life and is found in nearly 15 million Medicare beneficiaries. Because mobility limitation has been found to be associated with several reversible patient characteristics: smoking, obesity, social isolation, poverty, and poor...
education, some of these associated variables warrant further evaluation. Development of interventions to improve mobility limitations, whether therapeutic, technologic, or policy changes, may also help maximize patient access to and satisfaction with health care.

References

Supplier
a. Version 9.1; SAS Institute Inc, 100 SAS Campus Dr, Cary, NC 27513.

Objective: To address a neglected research area: the attributes of rehabilitation patients associated with “thoughts of suing a physician” (S-MD).

Design: The S-MD statement “I am thinking about suing one of my doctors” was administered to 2264 people, along with the Battery for Health Improvement (BHI 2). Items predictive of S-MD were identified.

Setting: Acute physical therapy, work hardening programs, chronic pain programs, physician offices, and vocational rehabilitation programs.

Participants: Participants included 777 rehabilitation patients and 1487 nonpatient community-dwellers.

Interventions: Not applicable.

Main Outcome Measures: We used a multivariate analysis of variance to determine which of the 18 BHI 2 scales predicted the S-MD statement. Items from the scales found to be predictive, plus other variables, were then used in a chi-square analysis that compared people who wished to sue with those who did not. We then used a stepwise regression analysis with significant items from the prior analyses to build a model for predicting a potential S-MD patient.

Results: The highest percentage (11.5%) of patients affirming the S-MD statement were those involved in workers’ compensation and personal injury litigation, compared with only 1.9% of community-living subjects. Stepwise regression of BHI 2 variables produced a 13-variable model explaining 38.04% of the variance. A logistic regression of demographic variables (eg, education, ethnicity, litigiousness) explained 20% of the variance.

Conclusions: Anger (P<.001), mistrust (P<.001), a focus on compensation (P<.001), addiction (P<.001), severe childhood punishments (P<.001), having attended college (P<.001), and other patient variables were associated with thoughts of suing a physician.

Key Words: Malpractice; Rehabilitation.

Because of its rising costs in both human and financial terms, medical malpractice has become a major concern in the health care field. It is estimated that in addition to more than $5 billion dollars in annual malpractice premiums and billions of dollars in court costs, 1 “defensive medicine” procedures performed to protect against increasing litigation add as much as $97.5 billion annually to the cost of medical services. 2 Thus, research into the causes of malpractice litigation is indicated.

In analyzing the reason for a malpractice suit, it is helpful to separate its causes into 4 general areas: specific attributes of the injury (negligence vs none); provider (physician) attributes; the physician-patient relationship; and patient attributes. Research into specific attributes of the injury has indicated that negligence may not be a major factor in whether a lawsuit is initiated. Actual negligence appears to be poorly correlated with the incidence of lawsuits. 3 For example, clinical analysis of 100 medicolegal cases found negligence to be an issue in slightly more than half (56%). 4 In those lawsuits where no negligence was found (44%), reasons for filing the lawsuit were: inability to come to terms with the disease or its end results (21%); lack of understanding of the disease process (16%); and unreasonable medicolegal action (7%). 5 Such data indicate that patients are sometimes dissatisfied with their care for reasons other than that of alleged negligence.

Some evidence suggests that another reason for initiating a malpractice suit pertains to patient dissatisfaction with the physician-patient relationship. 6,7 Here, information from the risk management services division of St. Paul Fire and Marine Insurance Company indicated that of 100 hospitalized patients who could legitimately bring a malpractice action against a medical care provider for failure to act or for acting inappropriately, less than 10% did. 8 Similar studies found litigation rates of 16% 9 and 13%. 10 This finding may be explained by the strength of the patient-physician relationship 11,12 and physician-patient communication. 13,14

Research into provider (physician) attributes associated with malpractice suits has also been limited. It appears that the number of lawsuits incurred by a medical practitioner does not relate to the quality of medicine practiced. 15 One study, 16 however, found that a surgeon’s tone of voice may be related to his/her malpractice history. Similarly, another study 17 found that the amount of time spent with a patient, good communication skills, and use of humor were also associated with a practitioner not having a malpractice claim history.

There has been a paucity of research into patient attributes associated with the initiation of a malpractice suit. At present, it appears that women 18 and people who are more affluent and have a higher education level 19 are more likely to initiate malpractice suits. Nothing is known about the personalities of

From the Rosomoff Comprehensive Pain & Rehabilitation Center (Fishbain) and Department of Psychiatry (Fishbain, Lewis), University of Miami School of Medicine, Miami, FL; Department of Psychiatry, Veterans Administration Hospital, Miami, FL (Fishbain); Health Psychology Associates, Greeley, CO (Bruns); and Integrated Therapies, Lakewood, CO (Disorbio).

This study was conducted without external funding. The study, however, reanalyzed data from a previous study that was funded and supported by Pearson Assessments.

A commercial party having a direct financial interest in the results of the research supporting this article has conferred or will confer a financial benefit upon 1 or more of the authors. Bruns and Disorbio receive battery for Health Improvement 2 royalties.

Reprint requests to Daniel Bruns, PsyD, 1275 58th Ave, Ste C, Greeley, CO 80634, e-mail: daniel.bruns@healthpsych.com.

0003-9993/07/8805-1102$32.00/0

doi:10.1016/j.apmr.2007.02.007

the patients initiating lawsuits\(^\text{17}\) except that on the basis of clinical observation, some studies have concluded that patient anger was a factor.\(^\text{3,18,19}\) Thus, at present, we have little knowledge about patient attributes that are associated with the initiation of a malpractice suit. This is significant because others\(^\text{20}\) have postulated that because of these patient attributes, the initiation of some lawsuits may be secondary to reasons that the physician can “neither anticipate nor control.” These patients may have a “low threshold” for filing a lawsuit.\(^\text{21}\) If such patients can be identified, they might be treated with greater care.\(^\text{21}\)

This study addressed medical malpractice by investigating a group of patients who reported thoughts of suing a physician. We attempted to discern some patient attributes associated with these thoughts. To our knowledge, this is the first such study to be reported.

**METHODS**

The statement “I am thinking about suing one of my doctors” (S-MD), which was the focus of this study, was 1 of 600 questions and/or statements in the Battery for Health Improvement research version (BHI-R), and the Battery for Health Improvement 2 (BHI 2),\(^\text{22}\) which is a shorter version of the BHI-R. We administered the BHI-R to subjects in this study and scored the BHI 2 scales from the BHI-R.

The BHI 2 is a standardized test intended for use in the psychologic assessment of medical patients and is based on a biopsychosocial theory\(^\text{23}\) and has been integrated into clinical protocols.\(^\text{24}\) To establish its validity and reliability, the test was examined under a formal process that included development of theory-based items, review by a panel of expert judges, confirmatory factor analysis, comparison of test scales to criterion variables, and analyses of test-retest and internal reliability.\(^\text{22}\)

The test has also received favorable third-party reviews.\(^\text{25,26}\) A weakness of this recently published instrument is that it has not yet been used in any longitudinal studies.

The BHI 2 has 18 scales: 2 validity scales (self disclosure, defensiveness); 4 physical symptoms scales (somatic complaints, pain complaints, functional complaints, muscular bracing); 3 affective scales (depression, anxiety, hostility); 5 character scales (borderline, symptom dependency, chronic maladjustment, substance abuse, perseverance); and 4 psychosocial scales (family dysfunction, survivor of violence, doctor dissatisfaction, job dissatisfaction). We did not include the job dissatisfaction scale in the analyses in this study because many of the subjects were not in the workforce.\(^\text{22}\)

We administered the BHI-R to 777 rehabilitation patients who were being treated for pain or a physical injury; they were from 30 states in all 4 geographic regions of the United States. They were recruited by posters or flyers given to them by their health care providers in a variety of settings: acute physical therapy, work hardening programs, chronic pain programs, physician offices, and vocational rehabilitation settings. The patients were also drawn from various payer systems (Medicare, private insurance, workers’ compensation, auto insurance), and their diagnoses included a range of orthopedic injuries, headache and head injuries, fibromyalgia, and chronic regional pain syndrome. Any patient who wished to participate was accepted into the study. The only exclusion criteria were being less than 18 or more than 65 years old and being unable to read at the 6th grade level. Of the 777 patients, 527 were selected to approximate U.S. Census demographics for sex, age, ethnicity, and level of education.\(^\text{22}\)

Another 1487 community-living subjects from 16 states in all 4 geographic areas of the United States were also administered the BHI-R in order to develop the nonpatient BHI-R control group. These subjects, recruited through newspaper advertisements and posters, were recruited to match the demographics of race, education, age, and sex. No subject was excluded on the basis of past or present medical or psychologic diagnoses. A subset of the community sample (n = 725), representing the community norm group, was selected by matching the subjects to U.S. Census demographics for sex, age, ethnicity, and level of education. All community-dwelling subjects were asked if they had any serious medical conditions and those who reported having none constituted the “healthy” subset of the community sample.

The BHI-R was administered anonymously to all participants; they signed an informed consent stating that their information would be used for research purposes only and that no results or feedback would be given. Patients were informed that the information would not influence the course of their clinical care. The S-MD statement “I am thinking about suing one of my doctors” was scored on a Likert-scale format, with the responses being “strongly disagree,” “disagree,” “agree,” and “strongly agree” being assigned scores respectively of 1 through 4. This made it possible to assess not only the presence of the thought of suing a physician, but also the strength of the associated feelings.

**Data Analysis**

Data were analyzed using the SPSS software.\(^\text{8}\) Frequency and descriptive statistics were calculated to check all relevant characteristics of the data. Although the total number of patient and community subjects was 2264, the large number of variables (600, not including demographic data and other information) precluded the use of many statistical approaches. We addressed this difficulty by using a statistical means to identify promising groups of variables and then focusing on them. Additionally, there were different types of data, ranging from categorical data such as sex to continuous data such as standardized psychologic test scores. These differing types of data required different statistical approaches.

Before we conducted additional analyses, we randomly split the combined patient and community-dwelling subjects into a developmental phase sample (1811 subjects) and a cross-validation sample (453 subjects). The latter sample was used to assess validity and reliability of the S-MD regression equations. None of the cross-validation sample was used during the developmental phase.

In the preliminary step, we separated by patient and community-dwelling subsets the percentage of respondents who reported that they were thinking about suing one of their physicians (table 1). Next, we did a multivariate analysis of variance (MANOVA) to examine the relationship between all the 18 scales of the BHI 2 and the S-MD item. For this analysis, the S-MD variable was transformed to a dichotomy. Subjects were classified as having thoughts of suing their physician if they agreed or strongly agreed with the S-MD statement. The differences between the S-MD and non-S-MD groups on 14 of the 18 scales of the BHI 2 were highly significant (table 2).

Next, we used those BHI 2 scales from the MANOVA that were most closely associated with the S-MD item (P < .01) as independent variables in a stepwise regression equation (table 3) using the development group. Note that our purpose in this study was to make it possible to predict who will sue a physician, so in that sense, this is the primary dependent variable and all other variables are the independent ones. In the MANOVA, though, this distinction is reversed. This is because of an artifact of the particular statistical analysis and is not really indicative of the distinction between independent and dependent status because both types of analyses are attempts to
distinguish attributes of patients who want to sue a physician from the attributes of those who do not. To check for spurious findings, we did a stepwise regression on the cross-validation group and found similar results.

We then studied BHI 2 critical items in a 2×2 chi-square analysis using the development group to assess the relationship between those variables and the S-MD variable. In addition, because the previous analysis (see table 3) had demonstrated that the doctor dissatisfaction scale explained the greatest amount of variance (24.1%), the 10 individual items (questions from this scale) were also used individually in the chi-square analysis to see which aspects of this scale were the most predictive (table 4). The same type of chi-square analysis was performed on demographic and non-BHI 2 variables thought to potentially play a role in predicting the S-MD variable (table 5).

In the next analysis, using the development group, we used significant (P<.01) variables from the previous analyses (see tables 4, 5) as independent variables in a stepwise regression model to assess the predictability of the S-MD variable (table 6). For this analysis, we used the 4-point rating scale version of the S-MD variable.

In the final analysis (table 7), demographic variables found to be significantly related to S-MD were used as independent factors to assess their combined impact on the S-MD variable in a logistic regression framework. For this analysis, we used the dichotomous version of the S-MD variable. The demographic variables were age, education, sex, ethnicity, insurance status, and litigation status. Sequential logistic regression analysis was then performed to test the individual contribution of each predictor in the omnibus step chi-square and Nagelkerke R². Table 7 includes the regression coefficient, omnibus step chi-square and significance level, Nagelkerke R², Wald statistic, and odds ratio (OR).

RESULTS

Table 1 shows the percentage of the reference group samples who said that they were thinking about suing one of their treating physicians. The highest percentage (11.5%) was in the sample of patients who were workers’ compensation or injury litigation cases.

MANOVA Analyses of BHI 2 Scales and the S-MD Item

Table 2 shows the results of the MANOVA using the BHI 2 scales as the dependent variables and S-MD as the independent variable. To check for spurious findings, given that all scales were highly significant except for perseverance, we also conducted a stepwise regression on the cross-validation group and found similar results.

Stepwise Regression for Prior Significant BHI 2 Scales With S-MD as the Dependent Variable

We then conducted a stepwise regression with the significant variables from the prior MANOVA analysis as the independent variables. The final R² was .28 with a final model significant F ratio (F6,1804 = 118.27, P<.001), which included doctor dissatisfaction, substance abuse, survivor of violence, somatic complaints, borderline, and family dysfunction scales. Dissatisfaction with a doctor was the strongest predictor of thoughts of suing, based on the relative weights of the estimates. While statistically significant, the other 5 variables only accounted for an additional 4% of the variance in the S-MD variable (see table 3).

In the final analysis (table 7), demographic variables found to be significantly related to S-MD were used as independent factors to assess their combined impact on the S-MD variable in a logistic regression framework. For this analysis, we used the dichotomous version of the S-MD variable. The demographic variables were age, education, sex, ethnicity, insurance status, and litigation status. Sequential logistic regression analysis was then performed to test the individual contribution of each predictor in the omnibus step chi-square and Nagelkerke R². Table 7 includes the regression coefficient, omnibus step chi-square and significance level, Nagelkerke R², Wald statistic, and odds ratio (OR).

RESULTS

Table 1 shows the percentage of the reference group samples who said that they were thinking about suing one of their treating physicians. The highest percentage (11.5%) was in the sample of patients who were workers’ compensation or injury litigation cases.

MANOVA Analyses of BHI 2 Scales and the S-MD Item

Table 2 shows the results of the MANOVA using the BHI 2 scales as the dependent variables and S-MD as the independent variable. To check for spurious findings, given that all scales were highly significant except for perseverance, we also conducted a stepwise regression on the cross-validation group and found similar results.

Stepwise Regression for Prior Significant BHI 2 Scales With S-MD as the Dependent Variable

We then conducted a stepwise regression with the significant variables from the prior MANOVA analysis as the independent variables. The final R² was .28 with a final model significant F ratio (F6,1804 = 118.27, P<.001), which included doctor dissatisfaction, substance abuse, survivor of violence, somatic complaints, borderline, and family dysfunction scales. Dissatisfaction with a doctor was the strongest predictor of thoughts of suing, based on the relative weights of the estimates. While statistically significant, the other 5 variables only accounted for an additional 4% of the variance in the S-MD variable (see table 3).
Chi-Square Analyses of Clinical Variables and Doctor Dissatisfaction Items With the S-MD Variable

The items on the doctor dissatisfaction scale were subjected to further analysis to see which ones had the most predictive power. Additionally, we assessed BHI II critical items having to do with violent ideation, entitlement, and other important clinical variables. Table 4 shows the results of the chi-square tests of the doctor dissatisfaction items and BHI II critical items with S-MD. Table 5 presents the results of demographics and non-BHI II variables with the S-MD variable.

Stepwise Regression for Prior Significant Independent BHI II Items With S-MD as the Dependent Variable

We conducted a stepwise regression with the significant variables from the chi-square analyses as the independent variables. This statistic determined the combination of variables for predicting S-MD, with the final adjusted $R^2$ being .38 with a significant F ratio ($F_{13,1797} = 84.87$, $P < .001$), and included 13 different BHI II items. “Angry with MD,” “Forced to see MD don’t trust,” “Thoughts of killing people,” and “Frequent thoughts of suicide” were the strongest predictors of the wish to sue, based on the relative weights of the estimates and the amount of variance that each item explained. The first 2 items were taken from the doctor dissatisfaction scale of the BHI II and the last 2 were other BHI II items. Table 6 displays the final results of the model.

Using the 20% holdout sample for cross-validation, the final adjusted $R^2$ decreased slightly to .32. Using the Cronbach’s alpha, these items exhibited an internal reliability of .70 at cross-validation.

---

### Table 3: Final Model for Stepwise Regression Results for S-MD as the Dependent Variable With Significant BHI II Scales From Prior MANOVA Analysis as Independent Variables

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Order of Entry into Equation</th>
<th>Cumulative $R^2$</th>
<th>Regression Equation Coefficient</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-MD</td>
<td>Doctor dissatisfaction</td>
<td>1</td>
<td>.241</td>
<td>.022</td>
<td>249.3</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Substance abuse</td>
<td>2</td>
<td>.262</td>
<td>.007</td>
<td>19.8</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Survivor of violence</td>
<td>3</td>
<td>.269</td>
<td>.009</td>
<td>23.1</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Somatic complaints</td>
<td>4</td>
<td>.272</td>
<td>-.005</td>
<td>11.3</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Borderline</td>
<td>5</td>
<td>.278</td>
<td>.01</td>
<td>24.2</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Family dysfunction</td>
<td>6</td>
<td>.282</td>
<td>-.01</td>
<td>11.9</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

NOTE. $n=1811$, $F_{6,1804} = 118.27$, $P < .001$.

### Table 4: Significant Chi-Square Values of Doctor Dissatisfaction Items and Critical BHI II Variables With S-MD

<table>
<thead>
<tr>
<th>Comparison Variable</th>
<th>n (Both Variable and S-MD)</th>
<th>n (S-MD)</th>
<th>$\chi^2$ Statistic</th>
<th>$P$</th>
<th>OR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angry with MD*</td>
<td>22</td>
<td>60</td>
<td>50.02</td>
<td>&lt;.001</td>
<td>10.09</td>
<td>5.74–17.74</td>
</tr>
<tr>
<td>Harmful treatment*</td>
<td>9</td>
<td>60</td>
<td>11.15</td>
<td>&lt;.001</td>
<td>4.39</td>
<td>2.07–9.24</td>
</tr>
<tr>
<td>Forced to see MD don’t trust*</td>
<td>18</td>
<td>60</td>
<td>54.08</td>
<td>&lt;.001</td>
<td>15.54</td>
<td>8.33–29.00</td>
</tr>
<tr>
<td>Imagining symptoms*</td>
<td>13</td>
<td>60</td>
<td>2.18</td>
<td>.14</td>
<td>1.93</td>
<td>0.86–4.35</td>
</tr>
<tr>
<td>MDs only want money*</td>
<td>25</td>
<td>60</td>
<td>52.25</td>
<td>&lt;.001</td>
<td>9.29</td>
<td>5.39–16.02</td>
</tr>
<tr>
<td>MDs never help*</td>
<td>11</td>
<td>60</td>
<td>7.26</td>
<td>&lt;.001</td>
<td>2.80</td>
<td>1.42–5.51</td>
</tr>
<tr>
<td>Same old treatment*</td>
<td>28</td>
<td>60</td>
<td>9.20</td>
<td>&lt;.003</td>
<td>2.27</td>
<td>1.35–3.80</td>
</tr>
<tr>
<td>MDs don’t listen*</td>
<td>25</td>
<td>60</td>
<td>22.27</td>
<td>&lt;.001</td>
<td>3.87</td>
<td>2.29–6.57</td>
</tr>
<tr>
<td>Some MDs are idiots*</td>
<td>31</td>
<td>60</td>
<td>18.94</td>
<td>&lt;.001</td>
<td>3.22</td>
<td>1.92–5.41</td>
</tr>
<tr>
<td>Reason to mistrust MD*</td>
<td>30</td>
<td>60</td>
<td>17.84</td>
<td>&lt;.001</td>
<td>3.13</td>
<td>1.87–5.25</td>
</tr>
<tr>
<td>Hearing voices†</td>
<td>6</td>
<td>52</td>
<td>13.74</td>
<td>&lt;.001</td>
<td>8.73</td>
<td>3.40–22.41</td>
</tr>
<tr>
<td>Nothing seems real†</td>
<td>11</td>
<td>49</td>
<td>11.94</td>
<td>&lt;.001</td>
<td>3.96</td>
<td>1.99–7.86</td>
</tr>
<tr>
<td>Addiction to prescription meds†</td>
<td>8</td>
<td>60</td>
<td>11.57</td>
<td>&lt;.001</td>
<td>5.03</td>
<td>2.27–11.12</td>
</tr>
<tr>
<td>History of substance treatment†</td>
<td>11</td>
<td>60</td>
<td>8.01</td>
<td>.006</td>
<td>2.97</td>
<td>1.51–5.86</td>
</tr>
<tr>
<td>Fear of dying†</td>
<td>11</td>
<td>60</td>
<td>8.47</td>
<td>.004</td>
<td>3.08</td>
<td>1.56–6.08</td>
</tr>
<tr>
<td>More than one week in jail†</td>
<td>7</td>
<td>60</td>
<td>2.18</td>
<td>.14</td>
<td>1.93</td>
<td>0.86–4.35</td>
</tr>
<tr>
<td>Flashbacks†</td>
<td>16</td>
<td>60</td>
<td>10.66</td>
<td>&lt;.001</td>
<td>2.92</td>
<td>1.62–5.27</td>
</tr>
<tr>
<td>Severe childhood punishments†</td>
<td>26</td>
<td>60</td>
<td>23.71</td>
<td>&lt;.001</td>
<td>3.98</td>
<td>2.35–6.74</td>
</tr>
<tr>
<td>Thinking about killing†</td>
<td>13</td>
<td>60</td>
<td>17.71</td>
<td>&lt;.001</td>
<td>4.93</td>
<td>2.58–9.43</td>
</tr>
<tr>
<td>Feeling dangerous†</td>
<td>10</td>
<td>60</td>
<td>13.39</td>
<td>&lt;.001</td>
<td>4.73</td>
<td>2.31–9.72</td>
</tr>
<tr>
<td>Many violent thoughts†</td>
<td>12</td>
<td>60</td>
<td>8.74</td>
<td>.003</td>
<td>2.99</td>
<td>1.55–5.77</td>
</tr>
<tr>
<td>Thinking about killing MD†</td>
<td>12</td>
<td>60</td>
<td>28.07</td>
<td>&lt;.001</td>
<td>0.11</td>
<td>0.05–0.21</td>
</tr>
<tr>
<td>Suicidal ideation†</td>
<td>13</td>
<td>60</td>
<td>20.63</td>
<td>&lt;.001</td>
<td>5.78</td>
<td>3.00–11.11</td>
</tr>
<tr>
<td>History of suicide attempt†</td>
<td>16</td>
<td>60</td>
<td>13.64</td>
<td>.001</td>
<td>3.43</td>
<td>1.89–6.21</td>
</tr>
<tr>
<td>My concerns are more important than those of others†</td>
<td>19</td>
<td>60</td>
<td>12.86</td>
<td>.001</td>
<td>3.02</td>
<td>1.72–5.29</td>
</tr>
<tr>
<td>Expects special attention†</td>
<td>19</td>
<td>60</td>
<td>9.38</td>
<td>.002</td>
<td>2.53</td>
<td>1.45–4.43</td>
</tr>
<tr>
<td>Somebody owes me for pain and suffering†</td>
<td>24</td>
<td>60</td>
<td>39.02</td>
<td>&lt;.001</td>
<td>6.68</td>
<td>3.88–11.47</td>
</tr>
<tr>
<td>I should be paid for the rest of my life†</td>
<td>28</td>
<td>60</td>
<td>12.02</td>
<td>.001</td>
<td>2.56</td>
<td>1.52–4.30</td>
</tr>
</tbody>
</table>

NOTE. $n=1811$. Abbreviations: CI, confidence interval; OR, odds ratio. *Items from doctor dissatisfaction scale of the BHI II. †BHI II critical item.
Logistic Regression for Significant Variables From Prior Analyses as the Independent Variables With S-MD as the Dependent Variable

We conducted a logistic regression with demographic variables as the independents and S-MD as the dependent variable. The independent variables were: education, ethnicity, sex, and attorney involvement in health care matters. We did a sequential logistic regression analysis to test the individual contribution of each predictor in the fit of the model with the omnibus step chi-square and Nagelkerke $R^2$. Table 7 includes the regression coefficient, omnibus step chi-square and significance level, Nagelkerke $R^2$, Wald statistic, and OR. The overall chi-square for the analysis was significant ($\chi^2$ test $= 119.1$, $P < .001$) and the model classified 95% of the subjects correctly. All variables were significant predictors of the wish to sue according to the Wald test. The odds of reporting a desire to sue a physician were decreased by 35% for lower levels of education and by 78% if the subject did not have an attorney involved who was involved in health care issues. The probability of suing a physician was increased by a multiplicative factor of 3.63 if the subject was of a race other than white. Other variables were insignificant and were not retained in the final model.

DISCUSSION

This is the first study known to us that has addressed the question of which patient attributes are associated with a wish to initiate a malpractice lawsuit against his/her physician. First, we attempted to determine the prevalence among both rehabilitation patients and community-dwellers of a wish to sue a physician. That wish was present in a range from 1.9% to 2.8% in the community sample and from 4.3% to 11.5% in the patient sample. The highest prevalence was in the workers’ compensation and personal injury litigation subsample (11.5%) (see Table 1). The 2 subsets of our data may be of particular importance in regard to establishing prevalence for the wish

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Variable Entry Order</th>
<th>Percentage of S-MD Variance Accounted For ($R^2$)</th>
<th>Regression Coefficient ($b$)</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-MD Anger with MD*</td>
<td>1</td>
<td>.233</td>
<td>.173</td>
<td>61.09</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>Forced to see MD don’t trust*</td>
<td>2</td>
<td>.288</td>
<td>.133</td>
<td>26.86</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>Thoughts of killing people†</td>
<td>3</td>
<td>.328</td>
<td>.128</td>
<td>42.86</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>My job is being disabled and collecting what I am owed</td>
<td>4</td>
<td>.343</td>
<td>.075</td>
<td>8.21</td>
<td>.005</td>
<td></td>
</tr>
<tr>
<td>MD is getting rich off of my suffering</td>
<td>5</td>
<td>.351</td>
<td>.070</td>
<td>6.70</td>
<td>&lt;.01</td>
<td></td>
</tr>
<tr>
<td>MDs should change their schedules to fit mine</td>
<td>6</td>
<td>.356</td>
<td>.068</td>
<td>7.43</td>
<td>&lt;.01</td>
<td></td>
</tr>
<tr>
<td>Addiction to prescription drugs†</td>
<td>7</td>
<td>.361</td>
<td>.076</td>
<td>13.45</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>Frequent thoughts of suicide†</td>
<td>8</td>
<td>.365</td>
<td>.103</td>
<td>20.44</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Want the system to pay me money</td>
<td>9</td>
<td>.370</td>
<td>.044</td>
<td>10.76</td>
<td>.002</td>
<td></td>
</tr>
<tr>
<td>At times I wish I were dead†</td>
<td>10</td>
<td>.373</td>
<td>.047</td>
<td>9.50</td>
<td>&lt;.003</td>
<td></td>
</tr>
<tr>
<td>Somebody should pay me for my suffering</td>
<td>11</td>
<td>.376</td>
<td>.062</td>
<td>7.12</td>
<td>&lt;.01</td>
<td></td>
</tr>
<tr>
<td>Severe childhood punishment†*</td>
<td>12</td>
<td>.379</td>
<td>.041</td>
<td>8.01</td>
<td>&lt;.01</td>
<td></td>
</tr>
<tr>
<td>Flashbacks of painful memories†</td>
<td>13</td>
<td>.380</td>
<td>.035</td>
<td>5.56</td>
<td>.02</td>
<td></td>
</tr>
</tbody>
</table>

NOTE. $n = 1811$. $F_{13,1797} = 84.87$, $P < .001$.

*Items from doctor dissatisfaction scale of the BHI 2.
†Other BHI 2 item.
to fit mine (OR = H11005). Medications, hearing voices, and other symptoms. These data report suicidal or homicidal ideation, addiction to prescription medications, and other symptoms. Patients who report thoughts of suing were more likely to have feelings of entitlement, may have already involved an attorney in their health care.

Information From Continuous Variables

The data from the continuous variables (see tables 2, 3) suggest that the single strongest scale score was an elevated score on the BHI 2 doctor dissatisfaction scale. This scale was designed to identify patients who harbor a deep antipathy toward physicians. A high score on this scale suggests that the patient is very angry with 1 or more physicians and perceives physicians in general as uncaring, untrustworthy, incompetent or ineffective, and motivated only by money. This scale score was the strongest single predictor identified in this study, with the S-MD subjects receiving a mean standardized score on this scale that was a full standard deviation higher than that of their non-S-MD counterparts (see table 2). Although the thought of suing is probably influenced heavily by events such as negligence and poor outcome in the course of medical care, this scale by itself accounted for more than 28% of the variability in S-MD. Other scale variables identified were substance abuse vulnerabilities, a history of multiple episodes of violent and/or sexual abuse, family dysfunction, somatization, and borderline personality traits.

The final stepwise regression model (see table 3) explained 38% of the variance in S-MD, with the greatest percentage of the variance being explained by anger and cynical beliefs directed toward physicians (doctor dissatisfaction scale items). Other variables that characterized these patients in the stepwise regression were: being forced to see a physician who was not trusted (also a doctor dissatisfaction scale item), being violence prone, having a belief in entitlement, possible addiction, depressed or suicidal, childhood trauma, and possible posttraumatic stress disorder.

Overall, these results indicate that the patient who contemplates suing his/her physician has the following characteristics: He/she is nonwhite, has a higher educational level, is a workers’ compensation patient, and has an attorney. This patient has cynical beliefs about physicians in general and is angry with the physician who he/she may have been forced to see. He/she may also have significant psychiatric problems such as addiction, depression or suicidal ideation, violent ideation, posttraumatic stress disorder symptoms, and is from a dysfunctional family with a history of abuse in childhood. Finally, this patient also has feelings of entitlement and is focused on compensation. 

S-MD, Patient Attributes, and Causality

Because this was a correlational study, it was not possible to determine the causal relationship between S-MD and other variables. Alternate causal relationships for some of the identified variables are readily apparent and this creates a “chicken or the egg” type of question about “which came first.” For example, it is possible that angry feelings directed toward the physician who he/she may have been forced to see may be caused on being compensated for their medical difficulties, may have feelings of entitlement, and may have already involved an attorney in their health care.
patient toward suing and frivolous litigation. Even though anger is widely regarded as a predisposing factor, it can be argued that the thought of litigating may produce angry feelings, and that this accounts for the relationship of anger to S-MD. While this is possible, some evidence suggests that the variables in this study may be pre-existing.

In contrast to affective variables such as anger, which are changeable in specific contexts, some of the risk factors for S-MD we identified are “trait” type variables. Traits are durable attributes of a person and do not readily change. For example, race, level of education, and history of treatment for substance abuse are trait-type variables. Thus, even though correlation data cannot determine causality by itself, trait variables such as these exclude some causal interpretations and may suggest others.

For example, we found that anger at physicians was associated with wanting to sue. It is possible that pre-existing anger predisposes a patient to sue. An alternate causal scenario, though, is that a medical error occurs and produces both S-MD and anger. In contrast, we also found that a college education is associated with an increased frequency of filing a lawsuit. In this case, however, while a college education could incline a person to use the legal system, a medical error could not cause a college education. As a result, when trait variables are being considered, the nature of the variable may make it possible to logically rule out certain causal interpretations and in so doing, suggest another.

While some of the identified risk factors are state variables (e.g., anger, suicidal ideation), and others are clearly traits (e.g., sex, race, education, and scales to assess life style such as substance abuse and survivor of violence), some of the BHI 2 scales contain both state and trait items. Measures such as these are less changeable than affective states but are still subject to change in reaction to life events. For example, the doctor dissatisfaction scale assesses more than just anger; it attempts to assess a broad, underlying belief system about the medical profession. As such, it is probably less easily changed than an affective state because 1 type of trait is a deeply engrained belief system.

As noted in the Introduction, the issue of patient attributes associated with initiation of a lawsuit has not been extensively addressed in the literature. There is, however, direct and indirect support for some of our findings. The first of these is socioeconomic status. Here, the literature indicates that the poor do not sue more and that the affluent are more likely to file a lawsuit. We also found that patients with higher levels of education are more likely to sue.

The second finding, which has some support in the literature, was that of patient anger. Previously investigated this variable in the only experimental study done in this area of research. He exposed several subjects to sand in a university building’s basement. He then had an alleged “construction worker” tell the subjects that exposure to the sand would damage their lungs. He found that the subjects who perceived danger and had the personality characteristic of anger contemplated filing a lawsuit. Some clinical case study insight data have also found “the underlying cause of almost every lawsuit is patient anger regarding some aspect of the patient-doctor relationship.”

The third finding—race (nonwhite)—as a variable in predicting lawsuits may also have some indirect support in the literature based on clinical observation. It is well established that poor communication between patients and physicians can result in lawsuits. It has been proposed, however, that this is not a deficit in the physician’s interpersonal skill but reflects language and cultural differences between patients and physicians and, perhaps, unexpressed prejudice by the physician against the patient. This could then be a potential explanation for the finding about race (nonwhite).

The fourth finding, which may have some indirect support in the literature, is that of depression and suicidal tendencies. The literature indicates that patients sue when they are unable to come to terms with their disease or its end result. Because “not coming to terms” may lead to depression, such data may then be indirectly related to our finding.

The fifth finding, which is addressed but not supported in the literature, is that of sex differences. The literature claims that women file more lawsuits than do men, but we found that sex was not a factor in contemplating a lawsuit. Suing and contemplating suing are operationally different. As such, they may have different predictive variables, which could be the reason for this discrepancy.

Study Limitations

Our subjects included both community-dwellers and patients recruited from rehabilitation and pain treatment sites. Although ours was a diverse group of subjects from many sites, we are uncertain whether the data of patients in rehabilitation and pain settings can be generalized to patients in other medical settings. Some variables, such as patient anger, are consistent with those found in the literature, while most other variables are general in nature and not exclusive to the rehabilitation and pain settings. At this point in time, however, we know of no studies that have explored differences in the dynamics of litigiousness across medical settings; further research into this area is indicated.

As we noted in our Introduction, in alleged malpractice situations approximately 10% of patients proceed with lawsuits. Our data pertain to a patient’s thoughts of suing his/her physician, which is not the same as actually filing a lawsuit. Therefore, it is not known how helpful our data would be in identifying patients who will sue their physicians. Such data can only be gathered from a prospective study. Nevertheless, because it is likely that lawsuits would originate in a population that contemplates suing, our results may be helpful in addressing this problem.

CONCLUSIONS

In some patient groups, thoughts of suing a physician are not uncommon, and this has been determined to be associated with some patient attributes. While events that occur in a medical setting, such as poor outcomes or medical errors, probably contribute substantially to thoughts of suing, research shows that only a small percentage of patients involved in alleged medical errors resort to litigation. The presence of psychologic risk factors may influence who sues and who does not. Future research in this area may now begin to test the validity of these patient attribute variables in those patients who actually proceed with lawsuits.

Acknowledgments: We thank the staff of Pearson Assessments for its invaluable help in collecting the data. Pearson Assessments was involved in data collection in the original study, in which data were gathered for the purpose of developing the psychologic scale. Pearson Assessments, however, had no role in the design or statistical analysis of this study, nor did Pearson Assessments provide any funding, support, or input. This study’s findings were the result of a reanalysis of the data after the original study was completed.

References

27. Lindberg MA. The role of suggestions and personality characteristics in producing illness reports and desires for suing the responsible party. J Psychol 2002;136:125-40.

Supplier
a. SPSS Inc, 233 S Wacker Dr, 11th Fl, Chicago, IL 60606.
Cancer Screening Behaviors Among Canadian Women Living With Physical Disabilities

Nicole S. Cooper, MSc, Karen K. Yoshida, PhD


Objective: To report the prevalence and factors associated with ever having had a Papanicolaou (Pap) test or pelvic examination among Canadian women with physical disabilities and the barriers to having the tests.

Design: Cross-sectional survey.

Setting: General community.

Participants: Convenience sample of 1095 women between the ages of 18 to 93 completed the survey. The most frequently reported health conditions were musculoskeletal (44%), neurologic (17%), and sensory (13%).

Interventions: Not applicable.

Main Outcome Measures: Outcomes included prevalence of ever having a Pap test or pelvic examination and odds ratios of having the tests.

Results: Prevalence of ever having a Pap test was 90% and 91% for a pelvic examination. The most common barriers to the screening tests were “not being sexually active,” “my doctor told me I do not need one,” and “the exam table is too high/narrow.”

Conclusions: Although the prevalence of ever having a Pap test or pelvic examination was at or above 90%, women with physical disabilities need further education on the necessity and benefits of having regular cancer screening behaviors, especially among those who may not be sexually active. Further research is also required into why these women are informed they do not require cancer screening tests.

Key Words: Disabled persons; Preventive health services; Rehabilitation; Vaginal smears; Women.

© 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

Can cancer screening behaviors among women with physical disabilities be studied extensively in the Canadian population? However, no studies have specifically examined women living with physical disabilities and their utilization of cancer screening tests within Canada. Cervical cancer is largely preventable through the Papanicolaou test (Pap test) to detect pre-cancerous lesions. Although the test has existed for over 50 years, research has shown that this health procedure remains underused by some portions of the Canadian population, mainly those women who are economically and/or socially disadvantaged, and possibly including women living with physical disabilities.

Cancer screening guidelines differ by province in Canada; however, review of literature from Canada and the United States indicates that cervical cancer screening should begin once a woman is sexually active or at the age of 18, and continue yearly until there are 2 annual negative tests, after which the test can continue every 3 years. The necessity and importance of regular pelvic examinations for women living with disabilities is identical to that for women without disabilities. Evidence from the United States indicates those with functional limitations are less likely than those without functional limitations to have had a Pap test within the past 3 years, and the more functional limitations a woman reports, the less likely she is to have had a Pap test within the past year. In addition, those women who were older than 65 with functional limitation were less likely to have had a Pap test than those between 18 and 44 with functional limitations.

A study by Nosek and Howland investigated cancer screening behaviors among women with physical disabilities in the United States and found that women living with disabilities are less likely than women without disabilities to receive pelvic examinations on a regular basis, and women with more severe functional limitations are significantly less able to comply with recommended guidelines. They examined the barriers to having a cancer screening test and found that the most frequently reported barriers included: “difficulty getting onto an examination table,” “lack of time,” and “the inability to find a doctor who suits them.” Disability specific barriers that were reported included: “not needing the test due to their disability” or “not being able to find doctor who is knowledgeable about their disability.”

When specifically examining those women with multiple sclerosis (MS), Shabas and Weinreb found that 25% were not having regular Pap tests or pelvic examinations, with 11% not having had a Pap test within the last 3 to 5 years. Another study of women with MS found that those who were nonambulatory had lower use of preventive health services in comparison with women who were fully or partially ambulatory.

A small survey of 45 women of the deaf community whose primary language was American Sign Language found that only 63% had had a Pap test or pelvic examination within the past year, despite all of them having health insurance and 31% having completed college. This indicates that not all barriers to the Pap test and pelvic examination are due to physical limitations but may also be related to the environment with respect to language.

The literature from the United States shows the need for research into the prevalence of cancer screening behaviors within Canada as well as the specific barriers that women with physical disabilities face with respect to having these tests. This study examines the prevalence and timing of cancer screening behaviors for the Pap test and pelvic examination for a sample of Canadian women living with physical disabilities.
In addition, we attempted to determine some of the barriers that these women face in having cancer screening tests performed.

**METHODS**

**Questionnaire Development**

The questionnaire consisted of 12 sections addressing the following areas: level of disability or long-term health condition(s), sociodemographic information, use of personal assistants, use of disability-related aids, cancer screening behaviors, and general health.

The questionnaire incorporated health-related questions from existing population health surveys including the National Population Health Survey (NPHS) \(^1\) from which we took the questions regarding ever having a Pap test or pelvic examination (table 1). Questions related to limitation of activity were also adapted from the NPHS. Questions related to the barriers that women face in having these tests were developed for this survey.

Focus groups occurred within the community to ensure that questionnaire items were relevant and to elicit additional questions. Forty-five women with a wide range of physical disabilities between the ages of 18 and 80 participated in the 6 focus groups in Ontario, Canada (2 in rural centers, 4 in urban centers).

Finally, the questionnaire was pilot tested in Ontario, Canada, to 2 convenience samples of socially and ethnically diverse women with physical disabilities (n=11, n=14) to determine clarity of the questions. Results of the pilot studies were used to revise the content, length, format, and language of the questionnaire.

**Eligibility**

Participants were eligible for this study if they were women over the age of 18, possessed a self-determined physical or sensory disability, and resided in the community.

**Participant Recruitment**

We developed a convenience sample through the mailing list of subscribers to the *Abilities* magazine (published by Disability Women’s Network of Canada) and was the initial recruitment of participants (N=2352). This included the members from all provinces of Canada with the exception of Quebec (due to a lack of funds for proper translation of the survey). Additional recruitment occurred through advertisements placed in the following: *Horizon's* magazine (a women’s publication), *VIBES* magazine (Canadian Hearing Society), flyers distributed to the mailing list of the Canadian Hard of Hearing Association, and through a notice placed on the website of the Canadian National Institute for the Blind. Alternative methods to the self-completed mailed survey were made available by audocassette and computer disk to allow those with a variety of physical disabilities to complete the questionnaire. A modified Dillman methodology \(^2\) was used to facilitate response and returns, allowing for extra time between mailings for those who may have required additional time to complete their survey because of their health condition.

Ethics approval for the survey was granted by the Ethics Review Board of the University of Toronto.

**Statistical Analysis**

We used descriptive statistics to describe the sample. Bivariate analyses were run between the dependent variables of ever having a Pap test or a pelvic examination and sociodemographic variables, disability measure variables, and barrier variables. All applicable variables were first examined on bivariate analysis with the dependent variable. If during bivariate analysis it was determined that there were no subjects within any cells of the 2×2 framework, the variable was then not included in the multivariate analysis. All variables were assessed for colinearity. A multivariate logistic regression was used to determine odds ratios (ORs) for the dependent variables of having a Pap test or pelvic examination.

We developed a model for the logistic regressions for ever having a Pap test and ever having a pelvic examination (fig 1). Variables selected for the model were based on the literature and on previous works in the area of disability and preventive health, as well as the bivariate analysis. No variables displayed colinearity and so they were all retained for multivariate analyses. Blocks of variables were added to the logistic regression

<table>
<thead>
<tr>
<th>Question</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you ever had a Pap smear test? (this involves the physician inserting an instrument into your vagina to collect cells from your cervix) The Pap smear test checks for cancer of the cervix.</td>
<td>Yes</td>
</tr>
<tr>
<td>If yes, when was the last time you had a Pap smear test?</td>
<td>Less than 6 months ago</td>
</tr>
<tr>
<td></td>
<td>6 months to less than 1 year ago</td>
</tr>
<tr>
<td></td>
<td>1 year to less than 3 years ago</td>
</tr>
<tr>
<td></td>
<td>3 years to less than 5 years ago</td>
</tr>
<tr>
<td></td>
<td>5 years or more ago</td>
</tr>
<tr>
<td>Have you ever had a pelvic examination? (the physician places two gloved fingers into your vagina and presses on your lower abdomen with the other hand)</td>
<td>Yes</td>
</tr>
<tr>
<td>If yes, when was the last time you had a pelvic exam?</td>
<td>Less than 6 months ago</td>
</tr>
<tr>
<td></td>
<td>6 months to less than 1 year ago</td>
</tr>
<tr>
<td></td>
<td>1 year to less than 3 years ago</td>
</tr>
<tr>
<td></td>
<td>3 years to less than 5 years ago</td>
</tr>
<tr>
<td></td>
<td>5 years or more ago</td>
</tr>
</tbody>
</table>

---

**Table 1: Variables From the NPHS for Pap Test and Pelvic Examination**

| Question Responses |
|-------------------|---|
| Yes               | No | Don't know |
| Less than 6 months ago          | 6 months to less than 1 year ago |
| 1 year to less than 3 years ago | 3 years to less than 5 years ago |
| 5 years or more ago                         | | |
sequentially. The first regression included sociodemographic variables. The second regression included the sociodemographic variables and the disability-related variables, and the third regression added the barrier-related variables. For each of the logistic regressions, the same sample of subjects was used for each of the 3 blocks of regressions.

RESULTS

Using the Abilities mailing list, we mailed out 2352 surveys, of which 926 eligible surveys were returned, 427 ineligible surveys were returned, and 999 surveys were not returned. The conservative response rate using the mailing list of Abilities is 48.1%. An additional 250 surveys were mailed out to women recruited from other sources and 225 eligible surveys were returned with a 90% response rate. The overall response rate was 53% from both sources. From the 1151 respondents, additional surveys were excluded if the participants stated no limitation of activities and/or reported no medical condition or physical disability. The resultant sample was 1095 surveys eligible for analysis.

The majority of women in this sample were between the ages of 30 to 59, with a mean age of 49 ± 14.3 years. Forty-eight percent of the women were married or partnered. Overall the sample was well educated, with 31% having completed university education which may have accounted for the high levels of income with 20% having a household income of $60,000 or more per year. Table 2 provides the breakdown of sociodemographic information on the total sample.

Frequencies and percentages of the disability variables are provided in table 3. The most frequently reported health conditions were 44% musculoskeletal, 17% neurologic, 13% sensory, and 10% cardiopulmonary. Sixty-nine percent of the women surveyed acquired their disability versus the 31% who were born with their disability. The majority of women (48%) did not report having pain that limited their activity.

Barriers to having a Pap test or pelvic examination were identical within the questionnaire and were collapsed into 3 categories of barriers based on type of barrier prior to analysis. Three categories of barriers included individual barriers, disability-related barriers, and structural barriers. A barrier was considered an individual barrier if the characteristics of the person prevented the screening behavior (eg, “Lack of time/too busy”). Disability-related barriers were those where the disability itself prevented the screening tests (eg, “It is not easy to get into the position because of pain, spasms, etc”). Structural barriers were those where the environment or the interface between the disability and the environment was the reason for

Table 2: Sociodemographic-Related Variables (N=1095)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18–29</td>
<td>87</td>
<td>8.1</td>
</tr>
<tr>
<td>30–39</td>
<td>211</td>
<td>19.6</td>
</tr>
<tr>
<td>40–49</td>
<td>293</td>
<td>27.1</td>
</tr>
<tr>
<td>50–59</td>
<td>248</td>
<td>23.0</td>
</tr>
<tr>
<td>60–69</td>
<td>132</td>
<td>12.2</td>
</tr>
<tr>
<td>70 and over</td>
<td>108</td>
<td>10.0</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never married</td>
<td>308</td>
<td>28.6</td>
</tr>
<tr>
<td>Partner/married</td>
<td>519</td>
<td>48.1</td>
</tr>
<tr>
<td>Separated/divorced/widowed</td>
<td>251</td>
<td>23.3</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some secondary or less</td>
<td>142</td>
<td>13.2</td>
</tr>
<tr>
<td>Secondary</td>
<td>146</td>
<td>13.5</td>
</tr>
<tr>
<td>Some college</td>
<td>126</td>
<td>11.7</td>
</tr>
<tr>
<td>College</td>
<td>182</td>
<td>16.9</td>
</tr>
<tr>
<td>Some university</td>
<td>151</td>
<td>14.0</td>
</tr>
<tr>
<td>University</td>
<td>331</td>
<td>30.7</td>
</tr>
<tr>
<td>Residence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>929</td>
<td>88.0</td>
</tr>
<tr>
<td>Rural</td>
<td>129</td>
<td>12.0</td>
</tr>
<tr>
<td>Total household income ($)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;7,000</td>
<td>47</td>
<td>4.8</td>
</tr>
<tr>
<td>7,000–11,999</td>
<td>154</td>
<td>15.7</td>
</tr>
<tr>
<td>12,000–19,999</td>
<td>126</td>
<td>12.9</td>
</tr>
<tr>
<td>20,000–29,999</td>
<td>135</td>
<td>13.8</td>
</tr>
<tr>
<td>30,000–39,999</td>
<td>123</td>
<td>12.6</td>
</tr>
<tr>
<td>40,000–49,999</td>
<td>114</td>
<td>11.6</td>
</tr>
<tr>
<td>50,000–59,999</td>
<td>83</td>
<td>8.5</td>
</tr>
<tr>
<td>≥60,000</td>
<td>198</td>
<td>20.2</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canadian only</td>
<td>475</td>
<td>43.4</td>
</tr>
<tr>
<td>Canadian and other</td>
<td>432</td>
<td>39.4</td>
</tr>
<tr>
<td>Non-Canadian only</td>
<td>188</td>
<td>17.2</td>
</tr>
</tbody>
</table>
The results of the logistic regressions for the Pap test were provided in table 5 with a sample of 564 used within the models. Only the participants who responded to the combination of all the sociodemographic, disability-related, and barrier-related variables in addition to not having a Pap test within the last year were included in the logistic regressions. The same participants were then included in each of the 3 modeled logistic regressions. Table 6 provides the results of the logistic regressions with ORs for the pelvic examination, using a sample of 544 participants. The same participants were included in the 3 modeled logistics and all participants were required to have responded to the sociodemographic, disability-related, and barrier-related variables in addition to not having a pelvic exam within the last year.

Regarding the Pap test and pelvic examination and the sociodemographic variables, those who were married or partnered, separated, divorced, or widowed had much higher odds of ever having had a Pap test or pelvic exam compared with those who were single and had never been partnered or married (for ORs, see tables 4, 5). Age was not related to ever having a Pap test in the final model, but age was associated with higher odds of having a pelvic exam (OR = 1.32; 95% CI, 1.02–1.87). Turning to the disability-related variables, for the Pap test, those with the health condition of cerebral palsy (CP) had decreased odds (OR = 0.39; 95% CI, 0.17–0.85). For pelvic exams, the only disability-related variable included was “I am not sexually active” (Pap, 19%; pelvic, 21%). Age was associated with higher odds of having a pelvic examination (OR = 1.32; 95% CI, 1.02–1.87). Turning to the barrier-related variables, for the Pap test, those with “I have problems getting onto the table, the table is too high/narrow” (Pap, 12%; pelvic, 12%) were more likely to have had the Pap test (OR = 3.93; 95% CI, 3.32–11.67). For pelvic exams, the only disability-related barrier included was “I am not sexually active” (Pap, 19%; pelvic, 21%).

### Table 3: Disability-Related Variables (N=1095)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Born vs acquired</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Born with disability</td>
<td>335</td>
<td>30.7</td>
</tr>
<tr>
<td>Acquired disability</td>
<td>757</td>
<td>69.3</td>
</tr>
<tr>
<td>No. of reported health conditions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>286</td>
<td>26.2</td>
</tr>
<tr>
<td>2</td>
<td>239</td>
<td>21.9</td>
</tr>
<tr>
<td>3</td>
<td>214</td>
<td>19.6</td>
</tr>
<tr>
<td>4</td>
<td>135</td>
<td>12.4</td>
</tr>
<tr>
<td>5</td>
<td>91</td>
<td>8.3</td>
</tr>
<tr>
<td>6+</td>
<td>128</td>
<td>11.7</td>
</tr>
<tr>
<td>Use of personal assistance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>682</td>
<td>62.9</td>
</tr>
<tr>
<td>No</td>
<td>402</td>
<td>37.1</td>
</tr>
<tr>
<td>Use of assistive device</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>878</td>
<td>80.6</td>
</tr>
<tr>
<td>No</td>
<td>212</td>
<td>19.4</td>
</tr>
<tr>
<td>Activities prevented by pain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No pain</td>
<td>467</td>
<td>43.7</td>
</tr>
<tr>
<td>None</td>
<td>36</td>
<td>3.4</td>
</tr>
<tr>
<td>A few</td>
<td>162</td>
<td>15.2</td>
</tr>
<tr>
<td>Some</td>
<td>270</td>
<td>25.3</td>
</tr>
<tr>
<td>Most</td>
<td>134</td>
<td>12.5</td>
</tr>
</tbody>
</table>

### Table 4: Reported Barriers to Having a Regular (once a year) Pap Test or Pelvic Examination

<table>
<thead>
<tr>
<th>Barriers</th>
<th>Pap Frequency (%) (n=680)</th>
<th>Pelvic Frequency (%) (n=651)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual barriers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am not sexually active</td>
<td>211 (19.5)</td>
<td>203 (20.7)</td>
</tr>
<tr>
<td>I no longer have a cervix or ovaries</td>
<td>116 (10.7)</td>
<td>90 (9.2)</td>
</tr>
<tr>
<td>Exams are painful</td>
<td>70 (6.5)</td>
<td>53 (5.4)</td>
</tr>
<tr>
<td>Lack of time/too busy</td>
<td>69 (6.4)</td>
<td>66 (6.7)</td>
</tr>
<tr>
<td>I cannot find a doctor I am comfortable with</td>
<td>55 (5.1)</td>
<td>49 (5.0)</td>
</tr>
<tr>
<td>I wish to avoid hearing bad news</td>
<td>34 (3.1)</td>
<td>24 (2.5)</td>
</tr>
<tr>
<td>I have a history of sexual abuse and I’m not comfortable with the procedure</td>
<td>23 (2.1)</td>
<td>23 (2.3)</td>
</tr>
<tr>
<td>I use alternatives to traditional health care</td>
<td>6 (0.6)</td>
<td>9 (0.9)</td>
</tr>
<tr>
<td>Disability-related barriers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>My doctor told me I do not need one</td>
<td>108 (10.0)</td>
<td>103 (10.5)</td>
</tr>
<tr>
<td>It is not easy to get into the position because of pain, spasms, etc.</td>
<td>78 (7.2)</td>
<td>65 (6.6)</td>
</tr>
<tr>
<td>Because of bladder/bowel incontinence I’m too embarrassed to go</td>
<td>31 (2.9)</td>
<td>33 (3.4)</td>
</tr>
<tr>
<td>I do not need them because of my disability</td>
<td>18 (1.7)</td>
<td>22 (2.2)</td>
</tr>
<tr>
<td>I cannot find a doctor who knows about my disability</td>
<td>12 (1.1)</td>
<td>11 (1.1)</td>
</tr>
<tr>
<td>Structural barriers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I have problems getting onto the table, the table is too high/narrow</td>
<td>128 (11.8)</td>
<td>118 (12.0)</td>
</tr>
<tr>
<td>Inconvenient to schedule visits</td>
<td>42 (3.9)</td>
<td>32 (3.3)</td>
</tr>
<tr>
<td>I had problems finding an accessible doctor’s office or clinic</td>
<td>37 (3.4)</td>
<td>34 (3.5)</td>
</tr>
<tr>
<td>I had problems finding transportation to get me there</td>
<td>25 (2.3)</td>
<td>28 (2.9)</td>
</tr>
<tr>
<td>The exam room is too cold</td>
<td>22 (2.0)</td>
<td>18 (1.8)</td>
</tr>
</tbody>
</table>
related variable was that participants who used personal assistance were more likely to have had a pelvic exam (OR = 2.59; 95% CI, 1.25–5.37).

When we examined the barriers to having had a Pap test, 3 barriers had increased or decreased odds of ever having had a Pap test: not sexually active (OR = 0.31; 95% CI, 0.13–0.71), problems getting onto the examination table (OR = 3.05; 95% CI, 1.05–8.86), and the examinations are too painful (OR = 6.37; 95% CI, 1.59–25.54). The barriers to having a pelvic exam included; not sexually active (OR = 0.34; 95% CI, 0.16–0.73) and a trend for I do not need them due to my disability (OR = 0.14; 95% CI, 0.02–1.03).

**DISCUSSION**

This study found that the prevalence of women ever having a Pap test was 90%, which is very similar to data from the United States where, in 1994, 91% of the women living with functional limitations had ever had a Pap test. The prevalence of ever having had a pelvic examination was 91%. Information is also provided on the barriers that these women face in having cancer screening tests performed.

Within this model the overall logistic regressions for the Pap test and pelvic examination indicate that for women with physical disabilities those who are married, with a partner, or have previously been married compared with single or never married have increased odds of having had a Pap test or pelvic examination. These findings are consistent with the results of the 1996–1997 NPHS in Canada, but are inconsistent with reports of those with mobility impairments in a national study in the United States. Age, however, was not uniformly related to increased odds of having the Pap test or pelvic exam. Interestingly, increases in education or income did not have statistically significant change in ORs for the Pap test or pelvic examination, as had been previously reported in the literature for Canadian women.

With respect to the disability-related variables, the Pap test and pelvic examination logistic regressions are very different. The ORs for barrier-related questions also have distinct features for the Pap test and pelvic examination. The following discussion will first examine the findings related to having ever had a Pap test followed by the pelvic examination. The following discussion will first examine the findings related to having ever had a Pap test followed by the pelvic examination.

The only 2 specific disabilities that displayed a relationship to ever having had the Pap test were those who reported having CP and those reporting depression. As previously discussed, other studies had examined the prevalence of screening among the deaf community and those with MS; however, within

**Table 5: ORs for Pap Test With 95% CI***

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 1: Sociodemographic</th>
<th>Model 2: Sociodemographic and Disability</th>
<th>Model 3: Sociodemographic, Disability, and Barriers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sociodemographic variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (in decades)</td>
<td>NS</td>
<td>NS</td>
<td>1.38 (1.02–1.87)</td>
</tr>
<tr>
<td>Marital status</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Single/never married</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married/partner</td>
<td>7.43 (3.53–15.65)</td>
<td>7.74 (3.31–18.09)</td>
<td>6.11 (2.28–16.38)</td>
</tr>
<tr>
<td>Separated/divorce/widowed</td>
<td>6.55 (2.43–17.64)</td>
<td>6.51 (2.19–19.36)</td>
<td>6.48 (1.93–21.70)</td>
</tr>
<tr>
<td>Disability variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>†</td>
<td>0.26 (0.08–0.80)</td>
<td>0.17 (0.05–0.64)</td>
</tr>
<tr>
<td>Depression</td>
<td>†</td>
<td>2.97 (1.10–8.04)</td>
<td>3.93 (1.32–11.67)</td>
</tr>
<tr>
<td>Barrier variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not sexually active</td>
<td>†</td>
<td></td>
<td>0.31 (0.14–0.71)</td>
</tr>
<tr>
<td>Problems getting onto the table</td>
<td>†</td>
<td></td>
<td>3.05 (1.05–8.86)</td>
</tr>
<tr>
<td>Exam too painful</td>
<td>†</td>
<td></td>
<td>6.37 (1.59–25.54)</td>
</tr>
</tbody>
</table>

**Table 6: ORs for Pelvic Exam (95% CI)**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 1: Sociodemographic</th>
<th>Model 2: Sociodemographic and Disability</th>
<th>Model 3: Sociodemographic, Disability, and Barriers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sociodemographic variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (in decades)</td>
<td>NS</td>
<td>NS</td>
<td>1.38 (1.02–1.87)</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single/never married</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Married/partner</td>
<td>6.36 (3.15–12.84)</td>
<td>6.50 (2.94–14.37)</td>
<td>4.12 (1.68–10.10)</td>
</tr>
<tr>
<td>Separated/divorce/widowed</td>
<td>10.20 (3.58–29.10)</td>
<td>8.56 (2.76–26.59)</td>
<td>7.57 (2.29–25.01)</td>
</tr>
<tr>
<td>Disability variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assistance with daily activities</td>
<td>†</td>
<td>2.18 (1.09–4.36)</td>
<td>2.59 (1.25–5.37)</td>
</tr>
<tr>
<td>Barrier variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not sexually active</td>
<td>†</td>
<td></td>
<td>0.34 (0.16–0.73)</td>
</tr>
<tr>
<td>I do not need them due to my disability</td>
<td>†</td>
<td></td>
<td>0.14 (0.02–1.03)</td>
</tr>
</tbody>
</table>

Abbreviation: NS, not significant.

*P < .05 (*n* = 564).
†Did not enter the model.
this study those specific disabilities did not display a relationship with having ever had a Pap test.

Women who reported having CP had decreased odds of ever having a Pap test within the model used. This may indicate that their health condition takes priority when they see their physicians and that they never address the issue of cancer screening tests during their physician visits. This has been reported within subjective narratives by Thomas and by Canadian women with physical disabilities in a report from British Columbia.

Conversely, having reported depression as a disability condition greatly increases the odds of ever having a Pap test. Possible explanations for the increased odds may be that women with depression undergo more clinical screening tests prior to their diagnoses. Alternatively, these women may be more aware of their health and are more likely to have the cancer screening tests. Given that we do not know the history of depression with these women it is impossible to determine the direct relationship between having had a Pap test and self-reporting the health condition of depression.

Women who reported not being sexually active as a barrier or reason for not having a Pap test were very unlikely to have ever had the test. This is consistent with findings from the British Columbia report. Cancer screening guidelines from various institutes, however, all report that screening should begin once a woman is sexually active or at the age of 18. Thus, these women who feel that not having sex obviates a Pap test need further education on the necessity of being screened.

The barriers of “the exam is too painful” and “I have problems getting onto the exam table or table is too narrow” have increased odds of having a Pap test. This indicates that these women have had the test and may even realize the importance of the test but have individual barriers or disability-related barriers to having the test again. This also confirms the inaccessibility of examination tables quantitatively that has been reported qualitatively in studies examining the barriers to preventive services among those living with physical disabilities.

Regarding pelvic examinations, no health conditions exhibited increased or decreased odds of ever having a pelvic examination within this model. The use of personal assistance was the only disability-related variable that had increased odds of having ever had a pelvic examination. Women with personal assistance may use their assistance when going to the physician for appointments, and thus may have fewer physical barriers to receiving a pelvic examination (eg, help with getting onto the table). Personal assistance in the context of this study could be from 3 major sources. First, a woman may have a family member or friend as an informal or formal personal assistant. Second, personal assistance may be government supported. Some government supported programs have personal assistance that is portable, such that the person living with a disability can get their needs met outside of the home (eg, school or health care). Third, a woman may pay for attendant services strictly out of pocket. In Canada, people with more activities of daily living needs will be more likely to have some form of government supported personal assistance. In this sample of women there was a mix of personal assistance and women may use a combination of informal and formal sources of assistance. Similarly to the Pap test, the barrier of “not being sexually active” resulted in decreased odds of ever having had a pelvic examination. Again these women may need further education on the need and benefits of regular screening tests.

The barrier “I do not need the test due to my disability,” although not statistically significant, did display a trend for decreased odds in ever having a pelvic exam. This explanation for not having a pelvic examination needs further investigation. Is it that these women have been told that they do not need the test due to their disability? And if so who told them they did not require the test? Or do these women believe that they do not require the test, in which case they need further education on the benefits of pelvic examinations.

**Study Limitations**

Although the survey attempted to recruit people from across Canada, the questionnaire was not translated into French due to budgetary constraints and thus the province of Quebec was not sampled. Therefore, the results of the survey cannot be generalized to all women with physical disabilities in Canada, given the lack of representation of women from Quebec and those who communicate only in French within other parts of Canada.

Because this survey used a convenience sample, the subjects of this survey are also well educated, representing few minorities, and mainly from the community in urban areas which may not accurately represent all Canadian women with disabilities. This has been a reported limitation of other work with convenience samples of women with physical disabilities.

As a result the findings are not applicable to all Canadian women with disabilities and different prevalence and barriers may be present among those with different sociodemographic factors including education and income.

The questionnaire was also quite long and as a result women with fewer physical disabilities may have been better able to complete the survey, thus biasing the results toward those with few limitations, even though alternative methods to completing the survey were offered to participants to help include those with more functional limitations.

**CONCLUSIONS**

Further studies are required to examine if women with physical disabilities are having cancer screening tests on a regular basis. This study only investigated if these women had ever had a Pap test or pelvic examination and because it was a cross-sectional design, it did not assess if the women were meeting suggested guidelines for frequencies of the tests. Different barrier patterns may be evident among women with physical disabilities who have only ever had 1 Pap test or pelvic examination compared with those who continue with the screening tests on a regular basis. In addition, there is no information available to report on the mortality or morbidity of cervical cancer for women living with physical disabilities.

Qualitative work may also help to determine the specific reasons that these women feel they do not require a Pap test or pelvic examination. As discussed earlier, is the lack of screening due to the fact that someone told them they do not require the test, and if so who provided them with that information? Or do these women believe that their disability precludes them from having cancer and thus they do not require any cancer screening tests?

Additional work is also required to examine women with physical disabilities living in institutions, because they may have very different responses to having ever had the cancer screening tests and more importantly may have many different barriers to receiving preventive health screening tests. Also French-speaking Canada needs to be surveyed to determine if there are cultural differences among women with physical disabilities in Canada.

This study, though, does provide us with information regarding the prevalence of Pap tests and pelvic examination among this sample of women with physical disabilities within Canada. Additionally, information related to barriers is important, be-
cause the specific barrier of “not needing the tests due to sexual inactivity” or believing that cancer screening tests are “not required due to their disabilities” indicates the need for further education of Canadian women with physical disabilities and their health care providers on the benefits of cancer screening tests.

**Acknowledgment:** We thank Fran Odette, project coordinator for the survey.

**References**


Objective: To test the clinical relevance of the stair climb power test (SCPT) as a measure of leg power impairments in mobility-limited older adults.

Design: Cross-sectional analysis of baseline data from participants within a randomized controlled trial.

Setting: Rehabilitation research gym.

Participants: Community-dwelling older adults (N=138; mean age, 75.4 y) with mobility limitations as defined by the Short Physical Performance Battery (SPPB).

Interventions: Not applicable.

Main Outcome Measures: Leg power measures included the SCPT and double leg press power measured at 40% (DLP40) and 70% (DLP70) of the 1 repetition maximum. Mobility performance tests included the SPPB and its 3 components: gait speed, chair stand time, and standing balance.

Results: Stair climb power per kilogram (SCP/kg) had correlations of moderate strength ($r=.47$, $r=.52$) with DLP40/kg and DLP70/kg, respectively. All 3 leg power measures correlated with each of the mobility performance measures with the exception of DLP40/kg ($r=.11$, $P=.27$) and DLP70/kg ($r=.11$, $P=.18$) with standing balance. Magnitudes of association, as described by the Pearson correlation coefficient, did not differ substantively among the separate power measures as they related to SPPB performance overall. Separate adjusted multivariate models evaluating the relationship between leg power and SPPB performance were all statistically significant and described equivalent amounts of the total variance ($R^2$) in SPPB performance (SCP/kg, $R^2=.30$; DLP40, $R^2=.32$; DLP70, $R^2=.31$). Analyses of the components of the SPPB show that the SCPT had stronger associations than the other leg power impairment measures with models predicting chair stand (SCP/kg, $R^2=.25$; DLP40, $R^2=.12$; DLP70, $R^2=.13$), whereas both types of leg press power testing had stronger associations with models predicting gait speed (SCP/kg, $R^2=.16$; DLP40, $R^2=.34$; DLP70, $R^2=.34$). Stair climb power was the only power measure that was a significant component of models predicting standing balance (SCP/kg $R^2=.20$).

Conclusions: The SCPT is a clinically relevant measure of leg power impairments. It is associated with more complex modes of testing leg power impairments and is meaningfully associated with mobility performance, making it suitable for clinical settings in which impairment-mobility relationships are of interest.

Key Words: Aged; Exercise test; Rehabilitation; Task performance and analysis.

© 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation
tance machines and (2) the separate respective associations of each leg power measure with the SPPB and its component measures of mobility performance.

METHODS

This study was a cross-sectional analysis of baseline data from a randomized controlled trial of exercise among mobility-limited older adults.

Conceptual Paradigm

For the purpose of this investigation, we evaluated disability outcomes using the paradigm originally established by Nagi and currently accepted by the Institute of Medicine. In this paradigm (1) impairments are defined as deficits at the level of the organ system (eg, loss of muscle power); (2) functional limitations are defined as alterations in the performance of a functional task (eg, lower mobility performance); and (3) disability is defined as the inability for a person to perform his/her social and environmental roles because of impairments and functional limitations.

Recruitment of Participants

Initially, 590 inquiries were solicited via advertising in newspapers, direct mailings, referrals from primary care providers, and telephone screenings. Of these, 260 people were identified as potentially eligible and attended an initial screening assessment at our facility. Outcomes testing for the intervention study was completed over 1 to 2 subsequent visits depending on participant availability. Measures used for this analysis were all completed within the first 2 visits, which were scheduled within 1 week of each other. On completion of the initial screening, 92 people could not participate in the study because of exclusion criteria, and 30 chose not to commit to the study, leaving 138 participants.

Screening Process

Subjects included in the study were community-dwelling older adults (age ≥ 65y) with SPPB scores between 4 and 10 who were able to climb a flight of stairs independently with a device such as a cane. Exclusion criteria were unstable acute or chronic disease, a score of less than 23 on the Folstein Mini-Mental State Examination, a neuromusculoskeletal impairment limiting participation in further outcomes testing, or an exercise tolerance test with positive findings for cardiovascular disease.

After providing informed consent, participants underwent a comprehensive history and physical examination that was conducted by the principal investigator. At the completion of the physical examination, the total number of active medical conditions was recorded for each participant. Active medical conditions were defined as either (1) any condition for which a patient was actively receiving treatment or (2) a condition requiring medical treatment within the past 1 year. Medical records were requested from participants’ primary care physicians to corroborate these findings. The total number of prescription and over-the-counter medications was recorded.

Impairment Measures

Lower-limb strength and power was measured with a pneumatic double leg press resistance machine as previously described. Briefly, the 1 repetition maximum (1-RM) was determined by progressively increasing the resistance for successive repetitions until each participant could no longer move the lever arm 1 time through the full range of motion. Peak power was measured as the best of 5 repetitions performed at 40% and 70% of the 1-RM, in which participants performed the concentric action of 1 repetition as quickly as possible. These 2 intensities were chosen to represent double leg press muscle power production at relatively high force and low velocity—70% 1-RM (DLP70)—and low force and high velocity—40% 1-RM (DLP40).

Stair climb power was calculated with the following formula: power equals force times velocity. Stair climb time (see below) and vertical height of the stairs were used to calculate velocity (distance/time), and body mass and acceleration due to gravity were used to calculate force. Stair climb time was measured by the following procedures. The examiner stood with each subject at the base of a well-lighted, 10-stair flight of stairs and (1) instructed the subject to safely ascend the stairs as fast as they could. They were further instructed that they could use the handrail if they thought it was necessary for safety purposes and to begin when the examiner said “Ready, set, go.” Timing began after the examiner said “go” and once each subject began moving. When both feet of a subject reached the top step, the timing stopped. Time was recorded to the nearest .01 second, and the average of 2 trials was taken. Test-retest reliability for this measure within our clinical lab is excellent (R=.99). Stair climb power testing was initiated after the inception of the study, and therefore only 124 of the 138 subjects underwent testing.

Mobility Performance Testing

Short Physical Performance Battery. The SPPB is a well-established, reliable, and valid measure of lower-extremity performance. Testing involves an assessment of standing balance, the timed 4.0-m walk, and a timed test of 5 repetitions of rising from a chair and sitting down. All times are measured to the nearest .01 second with a stopwatch. Each of the aforementioned tests is scored between 0 and 4 and summed to arrive at a maximum score, 12. SPPB scores have been found to predict disability over 1 to 6 years in several older populations. Testing was completed over a subsequent visit at which point a second SPPB was performed. Although the initial SPPB was the test that determined eligibility, the average of the 2 SPPB scores was used in this analysis. In addition, to better understand the respective associations between each leg power measure and the SPPB, each of its subcomponents (gait speed, chair stand time, standing balance score) were evaluated as unique secondary outcomes.

Adjustment Variables

Age, mass, height, sex, number of chronic conditions, and number of medications were assessed as part of the screening examination and served as potential adjustment variables.

Statistical Analysis

Descriptive statistics were calculated, including frequencies and proportions for categorical variables and mean and standard deviations for continuous variables. To account for body size and sex differences among participants, we normalized all power measures by mass. On inspection of the primary and secondary outcomes, to ensure that chair stand time was normally distributed, 1 value that was an outlier (69.9s) was recoded to the whole number value (37s) immediately above the next longest value. In addition, 4 subjects who attempted the chair stand test but were unable to complete it were converted to values of 37 seconds as well. Inclusion of these observations through this transformation did not materially alter the analyses. To compare the bivariate associations between each leg power measure (aim 1) and their respective
association with the SPPB and its components (aim 2), simple correlations were calculated between each of the impairment measures and each of the physical performance measures. Next, we evaluated the bivariate relationships between each of the potential adjustment variables and the mobility measures, ensuring a statistically significant association (P < .05). In addition, we inspected adjustment variables for colinearity, eliminating those colinear covariates that had the weaker association with the outcome. Prior reports show that impairment-function relationships may best be characterized as curvilinear.3,26 We therefore inspected plots of the bivariate relationships to determine whether linear or curvilinear models would best characterize our data. Next, we constructed 12 separate multivariate regression models, controlling for those covariates that were included in the final model. For all models, we inspected the distribution of the predictors and outcomes to ensure normality and inspected residuals to evaluate for influential data points. To ensure normality and ease of comparison between the secondary outcomes, chair stand time values were allowed a maximum value of 37 seconds. Last, the results were reanalyzed without those subjects who used the handrail during stair climb testing to ascertain whether this in any way altered the findings.

RESULTS

Participants had a mean age of 75.4 years and were predominately women (69%), predominantly white (85%) (15% black), and overweight, with a mean body mass index of 27.5 kg/m² (table 1). On average, participants reported 5.6 chronic medical conditions and were prescribed 4.3 medications. Participants had a mean SPPB of 8.7, which has been characterized as mobility limitations of moderate severity.25 Consistent with this, performance measures included the following mean values: gait speed of .93 m/s, chair stand time of 15.94 seconds, and average standing balance score of 3.15 out of 4. Twenty-one of the 124 participants completing the SCPT (17%) used the handrail during stair climb testing because of self-perceived safety concerns. Normalized values of stair climb power per kilogram (SCP/kg) were significantly associated (P < .001) with both normalized values of leg press power (DLP40/kg, R² = .47; DLP70/kg, R² = .52). Both leg press power values were highly associated (r = .93). With respect to the SCPT, through the course of the study there were no falls or adverse events, and no subjects refused to attempt the test.

The bivariate associations between each normalized value of leg power and the SPPB and each of its components are presented in table 2. Pearson correlation coefficients between mean SPPB score and SCP/kg, DLP40/kg, and DLP70/kg were all statistically significant (P < .001) with magnitudes of .51, .42, and .44, respectively. In evaluating gait speed, both DLP40 (r = .54) and DLP70 (r = .56) had greater associations than those seen with the SCPT (r = .29). The pattern of association was opposite with measures of chair stand, with r values being −.43, −.32, and −.31 for SCP/kg, DLP40/kg, and DLP70/kg, respectively (P < .001). Last, bivariate associations with standing balance were greater for SCP/kg (r = .31, P = .004) and were of lesser magnitude with both leg press measures, never reaching statistical significance (DLP40/kg, r = .11, P = .20; DLP70/kg r = .11, P = .18).

Inspection of the bivariate plots of the respective leg power–mobility measure associations did not suggest that curvilinear models were superior in characterizing these relationships. Among the potential adjustment variables, age and height were included within the multivariate models. The number of chronic medical conditions and number of regular medications are both established measures characterizing health status. They were sufficiently colinear (r = .67, P < .001) to justify using only the number of regular medications as a covariate, given that it had a stronger association with the outcomes than did chronic medical conditions. The resulting 12 separate multivariate linear regression models are presented in table 3. Within the 3 models predicting SPPB, the magnitude of the estimate measured in watts per kilogram was greater for SCP/kg (.59 vs .22–.23) compared with the other leg power measures; however, the respective models of leg power impairments described equivalent levels of the variance in SPPB performance, ranging between 30% and 32%.

To better understand these relationships, separate models were constructed predicting performance on each component of the SPPB. With respect to gait speed, although the magnitude of effect (parameter estimates) in watts per kilogram was similar (between .03 and .04) for all 3 power measures, models of DLP40/kg (R² = .34) and DLP70/kg (R² = .34) described a larger proportion of the variance than did SCP/kg (R² = .16). Compared with the bivariate associations, the magnitude of the effect of SCP/kg on gait speed was not statistically significant (P = .13). In the evaluation of chair stand performance, SCP/kg had a much greater effect (estimate, −2.62 W/kg), compared with an estimate of −.71 W/kg for both DLP40/kg and DLP70/kg. Although all 3 models were statistically significant, SCP/kg described a larger proportion of the variance in chair stand performance, having an R² value of .25 compared with an R² value of .12 for DLP40/kg and an R² value of .13 for DLP70/kg.
Using a chair stand ceiling value of 37 seconds, as mentioned in the Methods section, did not materially influence the results. For standing balance, the magnitude of the estimate was greatest for SCP/kg (SCP/kg estimate, .14; DLP40/kg estimate, .03; DLP70/kg estimate, −.03). Although the total variance in standing balance performance was similar for all 3 separate models, with $R^2$ ranging between .19 and .20, SCP/kg was the only power impairment measure that had statistical significance (SCP/kg, $P = .04$; DLP40/kg, $P = .33$; DLP70/kg, $P = .67$). When we re-evaluated the 12 statistical models excluding those subjects who used the handrail during stair climb (data not shown), we found that handrail use did not weaken the association between SCP/kg and each of the performance measures. Similarly, these multivariate relationships were not materially different when evaluated using curvilinear models (data not shown).

**Discussion**

Our investigation is the first to attempt to formally evaluate the SCPT as a potential measure of lower-extremity power and a predictor of important outcomes of mobility performance. The major finding of our study is that models using the SCPT predict approximately one third of SPPB performance, which is comparable with that predicted by leg power measured using pneumatic isotonic resistance machines. In addressing the 3 components of the SPPB, compared with both DLP40 and DLP70, the SCPT had a weaker association with gait speed performance and a stronger association with chair stand performance, and it was the only impairment measure to achieve statistical significance with standing balance performance.

The importance of these findings should be interpreted within the larger understanding of the clinical relevance of mobility performance testing. In a recent editorial, it was the only performance measure to achieve statistical significance with standing balance performance. Although not measured in our investigation, it may be that the power output and velocities of movement to maintain standing balance and to presents plots of mean recorded maximal power and subsequently derived mean maximal velocities (see fig 1 legend for formulae). Although the derived percentage of the 1-RM at which SCP/kg production occurred was 52%, clearly in the middle between 40% and 70%, both leg press impairment measures generated higher velocities of movement compared with the stair climb power. The mean gait speed velocity of participants (95m/s) is indicated on figure 1. The stronger association between leg press measures of power and gait speed seen in table 3 may be mechanistically linked to the similarity between velocity and power generated during these activities relative to stair climb power testing. Also, prior reports evaluating gait among functionally limited older adults have reported combined power generation from the hip and leg extensors to be in the range of 6 to 8W/kg, more closely approximating the power output seen with leg press testing.

This line of thinking may also clarify the other findings in table 3, in which models with the SCPT explained greater variances in chair stand and standing balance performance. Although not measured in our investigation, it may be that the power output and velocities of movement to maintain standing balance and to avoid falling are more closely aligned with leg power compared with gait speed. The mechanistic differences among SCP/kg, DLP40/kg, and DLP70/kg with respect to gait speed and the other components of SPPB performance may best be seen in figure 1, which

![Fig 1. Values for maximal mean power with the corresponding percentage of the 1-RM at which the maximal mean velocity of movement was derived. Double leg press (DLP) velocity was derived from the formula velocity = power/(%1-RM max velocity). Stair climb power (SCP) velocity was derived from velocity = SCP/(mean mass × 9.8m/s²). The derived percentage of 1-RM for SCP was calculated by %1-RM = mean DLP 1-RM/mean mass × 9.8m/s²). The mean gait speed for all subjects is noted.](Image 317x159 to 554x303)
perform the chair stand are more consistent with those generated during stair climb power testing than with leg press power measurements.

Cuoco et al\textsuperscript{29} have suggested that walking tasks may be better characterized by power output at lower resistance and higher velocities. Although our findings with regard to double leg press power are not fully consistent with this supposition, this may reflect different methodologies used to measure gait speed. The overall message of figure 1 supports a general conclusion that can be derived from the report by Cuoco.\textsuperscript{29} When considering rehabilitation methods, impairment and functional measures in general should be regarded in terms of their mechanistic similarities.

The weaker associations observed between the stair climb power and gait speed should be considered if gait speed is a sole outcome of interest. Gait speed has been shown to be almost as strong a predictor of mortality and disability as the SPPB.\textsuperscript{6} However, leg power generation at more distal segments of the leg, such as the ankle, play a larger role in gait.\textsuperscript{30} In addition, stair climb activity emphasizes more proximal lower-extremity muscle power generation.\textsuperscript{51,52} If gait speed is the only outcome of interest, the SCPT may be a less optimal power measurement. One potential solution to this concern would be to consider future measurement of leg power while walking up a ramp. Although not a focus of the current study, this other means of leg power measurement may more favorably evaluate leg power generation through the more distal lower-extremity muscle groups. This option may also limit the confounding effect of height observed between stair climb power and gait speed, which was likely related to the relative disadvantage with stair climbing for people of short stature.

**Study Limitations**

Our study has limitations. It might be suggested that use of the handrail could allow upper-extremity power output to be included within this lower-extremity power measure. However, the association between the SCPT and these lower-extremity mobility performance measures was strongest when people who used the handrail for safety purposes were included in the analyses. Our own observations suggest that use of the handrail primarily served the purpose of mitigating balance concerns and ensuring confidence to perform optimally. This allows people with balance impairments to more closely approximate the testing experience of leg press testing, in which they are seated. Furthermore, although not measured, upper-extremity weight bearing and push off from the handrail appeared to be minimal. In addition, our study was not able to compare stair climb power with other methods of leg power generation (eg, isokinetic testing). Although pneumatic resistance equipment is in our opinion superior, our investigation would have been strengthened with a direct comparison with these other leg power measures. Also, this was a cross-sectional analysis among a relatively homogenous cohort. A longitudinal investigation among a more racially diverse population would better characterize the temporal relationships between our leg power and performance measures, allowing clinicians to determine the sensitivity of the SCPT to detect training-induced changes and, in turn, to identify critical and clinically relevant values of the SCPT.

Beyond the mechanistic considerations and potential limitations, it is important to reconsider the primary aim of the study: to determine the clinical usefulness of the SCPT as a measure of leg power impairments among older adults with mobility problems. It is not uncommon for sound clinical measures to lack the precision and accuracy of their sophisticated and expensive lab-based counterparts.

**CONCLUSIONS**

Our study has shown that the associations between the SCPT and the SPPB are sufficiently strong to consider the SCPT a relevant clinical measure of leg muscle power impairments. The true test of the feasibility of the SCPT as an impairment measure would be the evaluation of its use within clinical settings that use mobility performance testing as a means of screening for those at risk for disability. Most clinicians caring for older adults do not have access to lab-based power measures. We recognize that the SPPB and similar mobility tests will serve as clinical screening tools in the future; therefore, for older adults with mobility limitations, the SCPT may be a useful impairment measure guiding care.

**References**


Supplier
Exercise Testing and Training in a Cancer Rehabilitation Program: The Advantage of the Steep Ramp Test

Ingrid C. De Backer, MSc, Goof Schep, MD, PhD, Adwin Hoogeveen, MD, PhD, Gerard Vreugdenhil, MD, PhD, Arnold D. Kester, PhD, Eric van Breda, PhD


Objective: To compare the short maximal exercise capacity test (steep ramp test) with the submaximal test to determine the most appropriate exercise test in cancer rehabilitation.

Design: A prospective study in which a submaximal test, a maximal short exercise capacity test (steep ramp test), and a maximal oxygen consumption test (\(\dot{V}O_2\max\) test) were performed before and after an 18-week training program. \(\dot{V}O_2\max\) testing, the criterion standard for the measurement of physical capacity, was compared with the submaximal test and the steep ramp test.

Setting: Community hospital and physiotherapy.

Participants: Thirty-seven cancer survivors (10 men, 27 women) treated with chemotherapy. The subjects’ mean age ± standard deviation (SD) was 48±11 years.

Intervention: An 18-week training program including strength training, interval aerobic training, and home-based activities (endurance).

Main Outcome Measures: Estimated \(\dot{V}O_2\max\) (submaximal test) and maximal workload (steep ramp test) were assessed during the exercise tests and compared with the results of the \(\dot{V}O_2\max\) test.

Results: A paired \(t\) test showed a significant improvement in \(\dot{V}O_2\max\) (+13%, \(P<.001\)) and maximal workload (+19%, \(P<.001\)) after the training program. This improvement was confirmed in the steep ramp test (maximal workload, +13%, \(P<.001\)) but not in the submaximal test (estimated \(\dot{V}O_2\max\), +4%, \(P=.192\)). Pearson correlation quantified only a moderate correlation between the \(\dot{V}O_2\max\) test and the submaximal test and a high correlation between the \(\dot{V}O_2\max\) test and the steep ramp test. Intraclass correlation determined the test-retest reliability of the submaximal test (.873) and the steep ramp test (.996). A linear regression model (\(\dot{V}O_2\max\), 6.7; steep ramp Wmax, +356.7) was estimated to predict \(\dot{V}O_2\max\) from the steep ramp test outcome, implying a prediction margin of ±2 SDs (616mL/min).

Conclusions: The submaximal test proved to be invalid, whereas the steep ramp test seems to be a practicable, reliable, and valid test for the assessment of the training dose. The steep ramp test can be regularly repeated during the training program, providing information needed to readjust the training dose according to the progress made.

Key Words: Cancer; Ergometry; Exercise test; Oxygen consumption; Rehabilitation.

© 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation
limitations during exercise as well. In daily clinical practice, however, direct measurements of VO\(_2\)max during a graded exercise test are expensive in terms of equipment and staff and put a serious burden on the patient as far as exercise time is concerned.

For this reason, submaximal exercise tests are often used. Maximum exercise capacity is then deduced from the heart rate at a submaximal workload. Although submaximal testing is widely used, the outcome in terms of determining the patient’s loading capacity is inaccurate. The major physiologic reason for this inaccuracy is that the heart rate varies substantially at any level of submaximal work, independently of the oxygen uptake. This variation in heart rate can be caused, for example, by emotional state, degree of physical conditioning, elapsed time after the previous meal, total circulating hemoglobin, the degree of hydration, alterations in ambient temperature, and hydrostatically induced changes resulting from prolonged erect posture.

Recently, a short maximal exercise capacity test (steep ramp test) has been described in the exercise rehabilitation of patients with chronic heart failure. This test proved to be safe, practical in its use, and reproducible.

Therefore, the aim of the present study is to compare the short maximal exercise capacity test (steep ramp test) with the submaximal test to discover the most valid, reliable, and feasible exercise test in cancer rehabilitation. To make a good comparison, we analyzed whether the tests had enough discriminative power to detect the progress that was made in a training program, assessed the test-retest reliability in a representative group of patients, and compared the test results of the steep ramp test and the submaximal test with VO\(_2\)max testing as a criterion standard.

METHODS

Participants

Thirty-seven patients treated curatively with chemotherapy were included in the study. Patient characteristics are depicted in Table 1. Training started a minimum of 6 weeks after completing chemotherapy to counteract bias resulting from spontaneous recovery after chemotherapy.

The following patients were excluded: (1) patients who were not capable of performing basic skills like sitting or lying down, (2) patients who had cognitive disorders or severe emotional instability, and (3) patients who were known to have other serious diseases that might hamper physical performance capacity (eg, heart failure, chronic obstructive pulmonary disease).

The project was approved by the medical ethics committee of the Máxima Medical Centre Hospital, and informed consent was obtained from all subjects.

Exercise Testing

All exercise testing was performed on a cycle ergometer. Before and after the training program, a maximal exercise test (VO\(_2\)max test) was performed. The submaximal and steep ramp tests were performed before, during, and after the training program.

VO\(_2\)max Test

A VO\(_2\)max test was performed on a cycle ergometer, and an oximeter was used to collect expired gas that was sampled and analyzed breath by breath for oxygen (VO\(_2\)), carbon dioxide (VO\(_2\)), and volume. The oximeter was coupled to a computer, which plotted workload against VO\(_2\), VO\(_2\), and heart rate. The heart rate and VO\(_2\) used for sampling and statistical calculations were taken from the means of every 30 seconds. Electrocardiographic activity was continuously monitored by using a 12-lead Jaeger electrocardiograph. Ventilatory threshold was determined by using the oxygen equivalent method.

Before the test, an estimation of the maximal workload (Wmax, in watts) was made by an experienced sports physician based on the patient’s sex, height, age, and history of physical exercise capacity. Based on this estimation, a ramp protocol was applied in which the subject was expected to reach the maximum short exercise capacity, the time cycled at that load, and the heart rate.

### Table 1: Subject Characteristics (N=37)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)</td>
<td>Mean ± SD 48±11 Range 24–71</td>
</tr>
<tr>
<td>Sex (n)</td>
<td>Male 10 27.0 Female 27 73.0</td>
</tr>
<tr>
<td>Cancer type (n)</td>
<td>Breast cancer 20 54.1 Ovarian cancer 5 13.5 Non-Hodgkin’s lymphoma 5 13.5 Colorectal 4 10.8 Testis 2 5.4 Hodgkin’s lymphoma 1 2.7</td>
</tr>
<tr>
<td>Treatment (n)</td>
<td>Chemotherapy 37 100.0 Plus radiotherapy 1 2.7 Plus surgery 11 29.7 Plus radiotherapy and surgery 20 54.1</td>
</tr>
<tr>
<td>Chemotherapy (n)</td>
<td>AC, breast cancer 6 16.2 CMF, breast cancer 5 13.5 FEC, breast cancer 9 24.3 Carboplatin-paclitaxel, ovarian cancer 5 13.5 CHOP/CVP, non-Hodgkin’s lymphoma 5 13.5 5-FU leucovorin, colorectal cancer 4 10.8 BEP/EP, testis 2 5.4 ABVD, Hodgkin’s lymphoma 1 2.7</td>
</tr>
</tbody>
</table>

Abbreviations: ABVD, doxorubicin, bleomycin, vinblasticine, dacarbazine; AC, Adriamycin, cyclophosphamide; BEP, bleomycin, etoposide, cisplatin; CHOP, cyclophosphamide, doxorubicin, vincristine, prednisone; CMF, cyclophosphamide, methotrexate, fluorouracil; CVP, cyclophosphamide, vincristine, prednisone; EP, etoposide, cisplatin; FEC, fluorouracil, epirubicin, cyclophosphamide; SD, standard deviation.
rate at the end of the test were reported. This test was performed at the start of the program, week 5, week 9, week 13, and week 18.

**Submaximal Test**

The submaximal test was based on the maximal workload and the ventilatory threshold obtained with the $\dot{V}O_2$max test. The test started at 50% of maximal workload and was increased by 10% every 3 minutes. Heart rate was reported in the last 15 seconds of each section. At a heart rate corresponding to the ventilatory threshold (as assessed by the previous $\dot{V}O_2$max test), time and workload were recorded. The submaximal test ended when the patient reached a heart rate that was 10 bpm higher than the heart rate at the ventilatory threshold. This test was performed at the start of the program, week 13, and week 18.

**Training Program**

The training program used in this study consisted of a supervised strength and interval aerobic training program lasting 18 weeks. For the first 12 weeks, patients underwent an intensive strength training program and interval aerobic training on a bicycle under supervision of a physiotherapist twice a week. In addition to the supervised training, patients were asked to perform homework activities (endurance training). During the last 6 weeks, subjects trained once a week under supervision, and the instructions for homework activities were intensified.

**Strength Training**

The strength program consisted of 6 exercises targeting the large muscle groups as follows: (1) vertical row (focusing on the longissimus, biceps brachii, rhomboideus), (2) leg press (quadriceps, glutei, gastrocnemius), (3) bench press (pectoralis major, triceps), (4) pull over (pectoralis, triceps brachii, deltoideus, trapezius), (5) abdominal crunch (rectus abdominis), and (6) lunge (quadriceps, glutei, hamstrings). At first, strength exercises were performed at 65% to 80% of the 1-repetition maximum (1-RM), consisting of 2 sets of 10 repetitions. After week 12, the emphasis shifted from muscle strength to muscle endurance involving training with less resistance (35%–40% of 1-RM) but more (20) repetitions. Every 4 weeks, the training progress was evaluated, and the training result was adjusted by means of a 1-RM test.

**Interval Aerobic Training**

The interval aerobic training consisted of cycling 2 times 8 minutes, before and after the strength program. The first 8 weeks, the 8-minute periods consisted of 30 seconds at 65% of the maximal workload of the steep ramp test and 60 seconds at 30%; from week 9, the 8-minute periods consisted of 30 seconds at 65% and 30 seconds at 30% of the maximal workload of the steep ramp test.

**Endurance Training: Home-Based Activities**

In addition to the supervised training, the participants had to do endurance training at home (eg, walking, cycling, or swimming). The first 4 weeks, patients were advised to train at least 3 times a week, for between 30 and 60 minutes. After week 12, they were recommended to walk, cycle, or swim for 60 minutes, at least 4 times a week. These home activities were registered in an exercise log in which the patients subjectively determined the training intensity by Borg rating of perceived exertion and the duration of the exercise. The Borg rating was used for the home program for practical reasons; it was easy for the patients to register, and there was no need for equipment.

**Data Processing and Statistical Analysis**

Means and standard deviations (SDs) for physical characteristics and test results were calculated. Student t tests for paired samples were applied to compare the outcomes of the $\dot{V}O_2$max test, submaximal test, and steep ramp test before and after training. Significance was set at $P < 0.05$. In the submaximal test, heart rates at different submaximal workloads (50%, 60%, 70% of Wmax) were used to calculate predicted $\dot{V}O_2$max values with a computer solution for the Astrand nomogram of Shephard.

The mean of these values was taken as the predicted $\dot{V}O_2$max.

Intraclass correlation coefficients (ICCs) were calculated to examine the test-retest reliability of the steep ramp test and the submaximal test in a separate population of patients ($n = 23$).

---

**Table 2: V\text{\small{O}}\textsubscript{max} Test Before and After Training (N=37)**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Before Training (week 0)</th>
<th>After Training (week 18)</th>
<th>Difference</th>
<th>%</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass (kg)</td>
<td>77.9±14.6</td>
<td>78.9±15.1</td>
<td>+1.0</td>
<td>1.3</td>
<td>.053</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>26.2±4.4</td>
<td>26.5±4.5</td>
<td>+0.3</td>
<td>1.1</td>
<td>.067</td>
</tr>
<tr>
<td>$\dot{V}O_2$\textsubscript{max} (mL·kg⁻¹·min⁻¹)</td>
<td>27.6±6.4</td>
<td>31.2±6.7</td>
<td>+3.6</td>
<td>13.0</td>
<td>.000*</td>
</tr>
<tr>
<td>Wmax (W)</td>
<td>167.4±50.2</td>
<td>198.6±58.8</td>
<td>+31.2</td>
<td>18.6</td>
<td>.000*</td>
</tr>
<tr>
<td>$\dot{V}O_2$ VT (mL·kg⁻¹·min⁻¹)</td>
<td>19.9±5.5</td>
<td>24.6±5.7</td>
<td>+4.7</td>
<td>23.6</td>
<td>.000*</td>
</tr>
<tr>
<td>Heart rate VT (bpm)</td>
<td>141±15</td>
<td>153±14</td>
<td>+12</td>
<td>8.3</td>
<td>.000*</td>
</tr>
<tr>
<td>Maximal heart rate (bpm)</td>
<td>170±14</td>
<td>174±13</td>
<td>+4</td>
<td>2.5</td>
<td>.005*</td>
</tr>
<tr>
<td>Respiratory quotient ($\dot{V}CO_2$/\dot{V}O₂)</td>
<td>1.17±0.14</td>
<td>1.20±0.13</td>
<td>+0.03</td>
<td>2.6</td>
<td>.335</td>
</tr>
</tbody>
</table>

**Abbreviations:** BMI, body mass index; VT, ventilatory threshold.

*P<.01.

---

**Table 3: Submaximal Test Before, During, and After Training (N=37)**

<table>
<thead>
<tr>
<th>Percentage of Exertion</th>
<th>Variables</th>
<th>Before Training (week 1)</th>
<th>Week 13</th>
<th>After Training (week 18)</th>
<th>Difference (18–1)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>50% Wmax</td>
<td>Heart rate (bpm)</td>
<td>132±12</td>
<td>125±12</td>
<td>128±11</td>
<td>−4</td>
<td>.182</td>
</tr>
<tr>
<td>60% Wmax</td>
<td>Heart rate (bpm)</td>
<td>139±12</td>
<td>136±16</td>
<td>138±13</td>
<td>−1</td>
<td>.694</td>
</tr>
<tr>
<td>70% Wmax</td>
<td>Heart rate (bpm)</td>
<td>149±14</td>
<td>143±17</td>
<td>148±12</td>
<td>−1</td>
<td>.739</td>
</tr>
<tr>
<td></td>
<td>Estimated $\dot{V}O_2$\textsubscript{max} (mL/min)</td>
<td>2130±586</td>
<td>2284±550</td>
<td>2215±597</td>
<td>+85 (4%)</td>
<td>.192</td>
</tr>
</tbody>
</table>
The 95% confidence intervals (CIs) were determined for each ICC.

The Pearson correlation coefficient quantified the relation between measured \( V_{O2} \max \) and (1) measured Wmax of the steep ramp test and (2) estimated \( V_{O2} \max \) of the submaximal test, at different test moments (before and after the rehabilitation program). Correlations were categorized by Munro,32 in which .26 to .49 is a low correlation, .50 to .69 is moderate, .70 to .89 is high, and .90 to 1.00 is very high. CIs were calculated by using a Fisher z transformation.33

A linear regression model was used to predict the maximal power and \( V_{O2} \max \) in the \( V_{O2} \max \) test from the steep ramp test outcome in a mixed linear analysis of data of 2 test moments (before and after training), including a random intercept term for each patient.

Detailed statistical analyses showed a similar pattern between male and female patients. Because all patients were provided with a training load independently of sex, analyses and data were performed without distinction between men and women. All statistical analyses were performed with SPSS.5

RESULTS

All patients selected for the study were able to complete the training program. All exercise tests were tolerated well without complication.

\( V_{O2} \max \) Test

Table 2 shows the means and SDs of different variables measured during \( V_{O2} \max \) testing before and after training. A total of 37 patients with an average body weight of 78kg and body mass index of 26kg/m² completed a \( V_{O2} \max \) test. There was a 13% increase \((P<.001)\) in \( V_{O2} \max \) after training. A 19% increase \((P<.001)\) in maximal workload was noted after the training program. Oxygen consumption at the ventilatory threshold also increased significantly by 24% \((P<.001)\). Heart rate at the ventilatory threshold and maximal heart rate increased significantly by 12 \((P<.001)\) and 4bpm \((P<.01)\), respectively. There was no significant difference in respiratory quotient \(\left(\frac{V_{CO2}}{V_{O2}}\right)\) at the end of the test whether conducted before or after the program.

Submaximal Test

Table 3 shows the estimated \( V_{O2} \max \) based on heart rate at different submaximal workloads (Astrand nomogram). The heart rate at different submaximal workloads did not decrease after training, and, consequently, there was no significant change \((+4\%,\ P=.192)\) in estimated \( V_{O2} \max \). The ICC to measure the test-retest reliability of the submaximal test was moderate (.873), with a 95% CI of .713 to .948.

Table 4 displays the Pearson correlation coefficients between estimated \( V_{O2} \max \) (submaximal test) and measured \( V_{O2} \max \). The Pearson correlation coefficient is .79 before and .71 after the training program and was categorized as high according to Munro.32

Steep Ramp Test

Data of the steep ramp test conducted at different test moments are presented in table 5. Maximal workload, maximal workload per kilogram of body weight, and maximal heart rate increased by 13%, 15%, and 5% \((P<.01)\), respectively, after the training program. In accordance with the results of the \( V_{O2} \max \) test, there is a significant increase after training.

The ICC (to measure the test-retest reliability) of the steep ramp test is .996 (95% CI,.989-.998), which is categorized as very high.

Pearson correlation coefficients between the \( V_{O2} \max \) test and the steep ramp test at different test moments (before and after training) are presented in table 4. The correlations between maximal workload of the steep ramp and \( V_{O2} \max \) are categorized as very high according to Munro.32

Prediction of \( V_{O2} \max \) From the Steep Ramp Test Outcome

A mixed linear regression model was fitted to allow a prediction of maximal workload and \( V_{O2} \max \) with the \( V_{O2} \max \) test from the measurements of the steep ramp test. Results are presented in figures 1 and 2. The resulting regression equation to calculate maximal workload was as follows:

---

**Table 4: Pearson Correlation Coefficient Between the \( V_{O2} \max \) Test and Steep Ramp Test and Submaximal Test (N=37)**

<table>
<thead>
<tr>
<th>Time Moment</th>
<th>Variables</th>
<th>Pearson r</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before (week 0)</td>
<td>( V_{O2} \max ) ( (\text{Wmax}) ) – steep ramp ( (\text{Wmax}) )</td>
<td>.85*</td>
<td>.72–.92</td>
</tr>
<tr>
<td></td>
<td>( V_{O2} \max ) ( (\text{V} ) ) – steep ramp ( (\text{Wmax}) )</td>
<td>.82*</td>
<td>.67–.90</td>
</tr>
<tr>
<td></td>
<td>( V_{O2} \max ) ( (\text{Wmax}) ) – submaximal ( (\text{estimated ( V_{O2} )}) )</td>
<td>.79*</td>
<td>.63–.89</td>
</tr>
<tr>
<td></td>
<td>( V_{O2} \max ) ( (\text{V} ) ) – submaximal ( (\text{estimated ( V_{O2} )}) )</td>
<td>.79*</td>
<td>.62–.89</td>
</tr>
<tr>
<td>After (week 18)</td>
<td>( V_{O2} \max ) ( (\text{Wmax}) ) – steep ramp ( (\text{Wmax}) )</td>
<td>.86*</td>
<td>.75–.93</td>
</tr>
<tr>
<td></td>
<td>( V_{O2} \max ) ( (\text{V} ) ) – steep ramp ( (\text{Wmax}) )</td>
<td>.85*</td>
<td>.72–.92</td>
</tr>
<tr>
<td></td>
<td>( V_{O2} \max ) ( (\text{Wmax}) ) – submaximal ( (\text{estimated ( V_{O2} )}) )</td>
<td>.77*</td>
<td>.60–.88</td>
</tr>
<tr>
<td></td>
<td>( V_{O2} \max ) ( (\text{V} ) ) – submaximal ( (\text{estimated ( V_{O2} )}) )</td>
<td>.71*</td>
<td>.50–.84</td>
</tr>
</tbody>
</table>

\* \( P<.01 \).

**Table 5: Steep Ramp Test Before, During, and After Training (N=37)**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Before Training (week 1)</th>
<th>Week 5</th>
<th>Week 9</th>
<th>Week 13</th>
<th>After Training (week 18)</th>
<th>Difference (18–1)</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wmax</td>
<td>270±75</td>
<td>295±78</td>
<td>303±76</td>
<td>309±79</td>
<td>307±73</td>
<td>+37 (13%)</td>
<td>.000*</td>
</tr>
<tr>
<td>Wmax/kg</td>
<td>3.50±0.90</td>
<td>3.80±0.94</td>
<td>3.90±0.90</td>
<td>4.01±0.95</td>
<td>4.01±0.88</td>
<td>0.51 (15%)</td>
<td>.000*</td>
</tr>
<tr>
<td>Maximal heart rate</td>
<td>154±18</td>
<td>154±18</td>
<td>160±25</td>
<td>155±31</td>
<td>161±17</td>
<td>+7 (5%)</td>
<td>.003*</td>
</tr>
</tbody>
</table>

\* \( P<.01 \).
The residual SD (including the random intercept variance) was 26.7, implying a prediction margin of ±2 SDs (53.4 W). The resulting regression equation to calculate V\textsubscript{O2max} was as follows:

\[
V_{\text{O2max}} = 6.7 \text{ steep ramp Wmax} + 356.71
\]

The residual SD (including random intercept variance) was 308, implying a prediction margin of ±2 SDs (616 mL/min).

**DISCUSSION**

The physical limitations after treatment for cancer vary widely among patients.\textsuperscript{34} Physical capacity may be affected by a variety of factors such as stage and type of cancer, previous history of physical activity, psychologic variables, type of treatment, and obesity.\textsuperscript{12} Consequently, there is a need to evaluate exercise tests used for assessing the initial physical fitness state, readjustment of the training load, and monitoring of the effects of a training intervention.

\(V_{\text{O2max}}\) testing is widely accepted as the criterion standard for the evaluation of physical capacity.\textsuperscript{35} In this category of patients, \(V_{\text{O2max}}\) testing using gas exchange measurements and electrocardiographic registration, under supervision of a physician, is especially indicated because cancer survivors may have additional pulmonary or cardiac limitations caused by cardiotoxic (eg, anthracyclins) or pulmotoxic (eg, bleomycin) medications or radiation therapy to the breast.\textsuperscript{8,27,38-36} \(V_{\text{O2max}}\) testing serves, in this case, as a diagnostic tool. A shortcoming of both the steep ramp test and the submaximal test is that they do not have this diagnostic value. However, \(V_{\text{O2max}}\) testing cannot be used frequently during rehabilitation because of its cost (personnel and equipment) and impact on the patient. There is a need, therefore, for exercise tests that can be easily applied during training. These tests have to be valid and reliable and provide sufficient information about the cancer patient’s physical capacity and training progress.

In the present study, \(V_{\text{O2max}}\) testing was performed before and after the training program to provide a criterion standard of the physical exercise capacity. As we expected, the group was very heterogeneous in terms of physical exercise capacity (resulting in large SDs and range), which emphasizes the importance of individual training assessment by means of appropriate tests. After the training program, there was a significant increase in \(V_{\text{O2max}}\) (+13%), maximal workload (+19%), and \(V_{\text{O2}}\) at the ventilatory threshold (+24%). The respiratory quotient at the end of the test did not change after training and was above 1.15, indicating that maximal exhaustion was reached in both tests.\textsuperscript{29} These results all indicate that there is a substantial training progress in the cancer patients included in the training program. These results are in accordance with other training intervention studies in cancer patients that use \(V_{\text{O2max}}\) as an outcome measure for physical capacity. Significant changes in \(V_{\text{O2max}}\) fall within a range of 6% to 19%.\textsuperscript{3,11,14,38} Remarkably, the training results appear to level off starting from week 13 in both the submaximal test and the steep ramp test. The lesser degree of supervision starting from week 13 can be considered as possible explanation; however, normal leveling off of the training effect is possible.

The results of the submaximal tests were analyzed with the Astrand nomogram. Based on the assumptions of Astrand, there is a linear relationship between heart rate, exercise intensity, and oxygen consumption. Given this relation, the measurement of heart rate combined with workload and the age-related maximal heart rate provides a reasonable estimation of exercise capacity.\textsuperscript{23} The training progress has occurred when the heart rate decreases at a fixed submaximal workload, and, as a result, estimated \(V_{\text{O2max}}\) will increase. Interestingly, the results of the present study show no significant decrease in heart rate at different submaximal workloads (50%, 60%, 70%). In contrast to the \(V_{\text{O2max}}\) tests, based on the results of the submaximal test, we were not able to monitor training progress. Our results are in line with other investigations\textsuperscript{37} that found that submaximal tests designed to predict \(V_{\text{O2max}}\) for the majority of disabled populations are not accurate because of the altered disability-specific physiologic responses. Furthermore, Viniegra et al\textsuperscript{41} found in patients with chronic obstructive pulmonary disease that estimation of \(V_{\text{O2max}}\) by submaximal testing might be appropriate at the group level but is inaccurate in predicting the outcomes of each patient. Especially in cancer patients, the relationship between predicted \(V_{\text{O2max}}\) and real \(V_{\text{O2max}}\) may be distorted by several causes. For instance, Viniegra et al\textsuperscript{41} investigated cardiovascular autonomic function in 55 breast cancer patients. Thirty-one patients had been treated with cardiotoxic anthracycin—containing chemotherapy regimens, and abnormal variations in heart rate were found in 81%. Finally, Greiwe et al\textsuperscript{25} reported that a submaximal test overestimates \(V_{\text{O2max}}\) in healthy subjects and should not be used when an accurate assessment of \(V_{\text{O2max}}\) is required.

Given these results from the literature, one may wonder why in our research Pearson correlations of the submaximal test with the \(V_{\text{O2max}}\) test are relatively high (0.79, 0.79, 0.77, 0.71). A plausible explanation for this obvious contradiction is the fact that a large range in \(V_{\text{O2max}}\) (17.4—51.6 mL kg\(^{-1}\) min\(^{-1}\)) is observed in our research population, which would enhance the
The steep ramp test should be recommended in cancer rehabilitation programs as it is more reliable than the steep ramp test. Consequently, a steep ramp test is recommended for assessing exercise capacity and is less feasible and less practical means for prescribing the training load and for monitoring training progress in the rehabilitation of cancer patients. Submaximal testing proves to have only limited value in the assessment of exercise capacity and for monitoring the effects of a training intervention in cancer patients.

By using regression analysis, we can deduce the maximal wattage obtained by VO2max from the steep ramp test results with a margin of ±2 SDs (53W), as presented in figure 1. VO2max can also be deduced from the steep ramp test results within a margin of ±2 SDs (614mL/min) (see fig 2).

Because of its short duration (on average, 2min excluding recovery), the steep ramp test can be repeated regularly, and, consequently, it is easy to incorporate this test in a training program. The submaximal test lasts much longer (10min, excluding recovery) and thus is more difficult to integrate regularly into the program. In the present study, the steep ramp test was performed once every 4 weeks. Because there is a gradual improvement in the maximal workload (see table 5), the training can be readjusted according to the individual progress achieved. Maximal workload achieved in the steep ramp test can be easily translated into training intensity by calculating a percentage of the maximal workload. Because the steep ramp test has been used in patients with severe chronic heart failure and was tolerated by all of the patients in this study, it seems to be safe when it is correctly applied. 27,28

CONCLUSIONS

The steep ramp test has proved to be a valid, safe, and practical means for prescribing the training load and for monitoring training progress in the rehabilitation of cancer patients. Submaximal testing proves to have only limited value in the assessment of exercise capacity and is less feasible and less reliable than the steep ramp test. Consequently, a steep ramp test should be recommended in cancer rehabilitation programs for the individual assessment of training.

References

Suppliers

c. Version 11.5; SPSS Inc, 233 S Wacker Dr, 11th Fl, Chicago, IL 60606.
Recovery of Function in Skeletal Muscle Following 2 Different Contraction-Induced Injuries

Richard M. Lovering, PT, PhD, Joseph A. Roche, PT, Robert J. Bloch, PhD, Patrick G. De Deyne, MPT, PhD


Objective: To determine if the proliferation of myogenic cells is equally important to recovery of contractile function after 2 different types of contraction-induced muscle injuries.

Design: Randomized trial.

Setting: Muscle biology laboratory.

Animals: Adult male Sprague-Dawley rats.

Interventions: Tibialis anterior muscles were injured by a single lengthening contraction with large strain (1R) or multiple lengthening contractions with small strain (MR). The hindlimbs of some animals in each group were irradiated before injury to prevent proliferation of myogenic cells during recovery.

Main Outcome Measures: Contractile tension was measured immediately after injury and 3, 7, 14, and 21 days after injury. Permeation to Evans blue dye was used to assay membrane damage. Centrally nucleated fibers and reverse transcriptase-polymerase chain reaction were used as measures of myogenesis.

Results: Inhibiting myogenesis prevented the recovery of contractile function after MR, but not after 1R. Both protocols caused Evans blue dye uptake immediately after injury, but Evans blue dye was only retained in fibers for several days after 1R. This suggests that membranes reseal after 1R, but not after MR.

Conclusions: The mechanisms that underlie recovery after injuries caused by repeated lengthening contractions and injuries caused by a single lengthening contraction are different. The differences may be important when planning targeted rehabilitation strategies for each type of injury.

Key Words: Muscles; Regeneration; Rehabilitation; Satellite cells, skeletal muscle.

© 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation
METHODS

Injury Induced by a Single, Maximal Lengthening Contraction or Multiple Lengthening Contractions

Injury induced by a single lengthening contraction was performed as previously described. Briefly, male age-matched Sprague-Dawley rats, weighing 383 ± 12g, were anesthetized with intraperitoneal ketamine and xylazine (at 40 and 10mg/kg of body mass, respectively). With the animal supine, the hindlimb was stabilized and the foot was secured onto a plate, the axis of which was attached to a stepper motor, a potentiometer to measure range of motion of the ankle, and a torque sensor to ensure that equivalent dorsiflexion forces were exerted. A custom program was used to synchronize contractile activation and the onset of ankle rotation. For convenience, the left hindlimb was designated as the experimental side and the right hindlimb was designated as the control side.

For the single repetition (1R), the foot was placed orthogonal to the tibia and moved into plantarflexion through a 90° arc of motion at an angular velocity of 900°, beginning 200ms after tetanic stimulation of the tibialis anterior, in a procedure that produced a large and reproducible injury. For the multiple repetitions (MR), we reproduced an established injury protocol that uses 150 lengthening contractions, equally spaced apart over 30 minutes, through a smaller arc of motion, and with the muscle in a shorter start position. The ankle was dorsiflexed 20° from the perpendicular position and moved into plantarflexion through a 40° arc at 1700°/s. Ankle movement was superimposed onto a tetanic contraction of the dorsiflexors, which was initiated 100ms before movement.

For both injury protocols, the peroneal nerve was dissected free through a small incision at the lateral aspect of the knee and clamped with a subminiature electrode. Monophasic square pulses, 1ms in duration, were delivered with an S48 Stimulator. Pulse amplitude was adjusted to optimize twitch tension, and the frequency of pulses was increased until a maximal fused tetany was obtained (usually 75Hz). A stimulation isolation unit (model PSIU6) was used between the stimulator and electrode to minimize artifact and to ensure that the peak current delivered was no greater than 15mA.

Rats were used according to the guidelines set by the National Institutes of Health Guide for the Care and Use of Laboratory Animals. The University of Maryland Institutional Animal Care and Use Committee approved our procedures.

Gamma Irradiation

Hindlimbs were subjected to a single, localized dose of ionizing radiation (25Gy at 2.5Gy/min) 30 minutes before the injury. We used an irradiation dose and dose rate identical to that used in previous studies to eliminate proliferation of satellite cells. The irradiation was delivered with a Pantak-Seifert 250KpV X-Irradiator. The radiation beam was focused onto the lower hindlimb, while shielding protected the rest of the body composed of thick lead (fig 1). Ion chamber dosimetry was performed outside the collimator to ensure that the exact dose was delivered to the hindlimb, as well as inside the collimator (lead shielding), to monitor backscatter of radiation. In addition to animals subjected to injury after irradiation, we irradiated another group of animals (n=15) as controls to determine the effects of gamma irradiation alone.

Cardiotoxin and Bromodeoxyuridine Injections

We used 6 rats to test the effectiveness of our irradiation protocol in inhibiting proliferation of myogenic cells after injection of cardiotoxin (0.1mL 10µmol in phosphate-buffered saline [PBS]) into the tibialis anterior muscle, which induces a massive myogenic response. Three animals received an intraperitoneal injection of bromodeoxyuridine (BrDU) (500mg/kg) 6 hours before muscles were harvested, and the number of nuclei with BrDU uptake and centrally nucleated fibers were counted in injured muscles and compared to muscles that were irradiated before injury.

Functional Data

Contractile force was measured from tibialis anterior muscles of each group on the day of injury (day 0) and at 3, 7, 14, and 21 days after the injury (n=5 animals tested at each time point), as described. Briefly, the distal tendon of the tibialis anterior was cut and the muscle was dissected entirely free except for its origin on the tibia. The proximal portion of the distal tendon was secured in a custom-made metal clamp and attached to a load cell (FT03) with a suture tie (4.0-coated Vicryl). The load cell was mounted onto a micromanipulator so that the tibialis anterior could be aligned and adjusted to resting length (L0). The parameters for electric stimulation were the same as those used in the injury protocol. The hardware was calibrated after a 30-minute stabilization period to minimize thermal drift. To assess functional recovery, the injured tibialis anterior was compared with the noninjured tibialis anterior of the same animal.

Hematoxylin and Eosin Staining

After functional data were collected at the selected time points, tibialis anterior muscles were harvested from the anesthetized rat, snap frozen in liquid nitrogen, and stored at −80°C. The animal was then euthanized with pentobarbital sodium (200mg/kg) administered intraperitoneally. For standard histopathologic evaluation and for counting centrally nucleated fibers, 7µm-thick frozen sections of snap frozen muscle were stained with hematoxylin and eosin. Sections were ran-
homogenized in Trizol. Ribonucleic acid (RNA) was extracted and isolated according to the manufacturer’s protocol, based on published methods.

Aliquots containing 2 μg of total RNA from each sample were reverse transcribed using Thermoscript RT mixture with oligo-dT in a total volume of 20 μL, according to the manufacturer’s protocol. The complementary deoxyribonucleic acid (cDNA) was then mixed with Platinum PCR Supermix plus 20 pmol of the specific primers and amplified for 30 cycles in a GeneAmp PCR System 9700. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as an internal control. PCR products were separated on a 1.5% agarose gel by electrophoresis, stained with ethidium bromide, and photographed. Forward and reverse primers were designed using the National Center for Biotechnology database with Omiga 2.0 software. The primers used to amplify the reverse-transcribed cDNAs are outlined in table 1.

### Immunofluorescence Labeling

Injured and noninjured tibialis anterior muscles were harvested and snap frozen, as described above. Cryosections (cross-section, 10 μm) were prepared, collected onto chromium-coated slides, incubated with 3% bovine serum albumin in PBS for 1 hour, and then labeled with antidystrophin antibodies, followed by fluoresceinated donkey anti-mouse immunoglobulin G, as described. Tissue sections were washed and mounted in Vectashield. Digital images were obtained with a Zeiss 410 confocal laser-scanning microscope and assembled with CorelDraw.

### Injection of Evans Blue Dye

One day before injury or control treatments, the rats received an intraperitoneal injection of 1% (wt/vol) Evans blue dye in PBS at a volume of 1% body mass (1 mg of Evans blue dye/0.1 mL of PBS/10 g of body mass). This solution was sterilized by passage through a Millex-GP 0.22 μm filter. The presence of Evans blue dye does not interfere with normal muscle function or recovery after injury (data not shown).

Muscules of animals injected with Evans blue dye were collected as described above and the presence of Evans blue dye was assessed under confocal optics at 568 nm. To determine the number of Evans blue dye–positive fibers, at least 4 sections from every control and injured muscle were examined. We administered equal amounts of Evans blue dye per body weight and maintained identical optical settings during all confocal microscopy. Cells were counted as Evans blue dye positive if the fluorescent intensity was at least 50% of the signal intensity of the brightest region of the labeling in the extracellular space. With an average of 47 ± 12 fibers/field, almost 600 fibers from 3 different animals were counted at each time point to determine the presence of Evans blue dye. In most experiments, samples labeled with Evans blue dye were also labeled for dystrophin by immunofluorescence (see above).

### Statistical Analysis

Contractile data were analyzed using a 2-factor analysis of variance (ANOVA), with irradiation and time as the independent variables, and percentage of control force as the dependent variable. Cell counts from light and fluorescent micrographs were analyzed with a 1-way ANOVA. We performed a Tukey post hoc analysis to determine significant differences. Significance was set at P less than 0.05 and all results are reported as mean ± standard deviation (SD).

### Results

Despite very different muscle strain protocols, 1R and MR both resulted in similar force deficits (~40%), thus permitting the study of differences in the process of recovery from each without concern for differences in the extent of injury. To determine the effects of irradiation on recovery, we assessed the loss and recovery of tibialis anterior muscle function over a 3-week period after the injury by measuring maximal tetanic tension (\(P_0\)). \(P_0\) is a strong indicator of the overall status of a muscle and a reliable indicator of injury. Control data were obtained from the uninjured tibialis anterior of the opposite hindlimb.

### Effects of Gamma Irradiation on Functional Recovery

Compared with controls (\(P_0 = 8.68 ± 1.04\)N), \(P_0\) in tibialis anterior muscles from animals that were injured using the MR protocol decreased significantly on the day of injury (41% force deficit at day 0) and continued to decline at day 3 and day 7, at which times it remained significantly different from \(P_0\) in noninjured controls (fig 2). Force recovered thereafter and returned to normative values over the next 1 to 2 weeks, such that by day 14 there was no significant difference from the controls. Tibialis anterior muscles irradiated before injury had a similar, statistically significant loss of force on day 0 (46% force deficit) and had further losses on days 3 and 7 (see fig 2). In the 3-week period that we studied them, irradiated muscles failed to recover normally: 3 weeks after injury they still showed a significant force deficit of 35%. Additional controls showed that irradiation alone had no effect on the contractile function of uninjured muscle over the 3-week period we studied (see fig 1), nor did it alter overall growth of the animals, as indicated by body weight (table 2).

Tibialis anterior muscles in hindlimbs subjected to the 1R protocol experienced a comparable amount of immediate injury (43% decrease in \(P_0\) at day 0), but they began to recover within 7 days (see fig 2). Muscles that had been irradiated before the

### Table 1: Primers Used for RT-PCR

<table>
<thead>
<tr>
<th>Primer</th>
<th>NCBI Reference Number</th>
<th>Molecular Weight</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
<th>(T_a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myogenin</td>
<td>M017008</td>
<td>666 BP</td>
<td>CACCTTCCAGATGAAACC</td>
<td>AAGAAGTCACCCCAAGACG</td>
<td>59.4</td>
</tr>
<tr>
<td>MyoD</td>
<td>M24293</td>
<td>854 BP</td>
<td>CACTCTCCAAATGTC</td>
<td>CTTATTCCAACACCTGAGC</td>
<td>58.8</td>
</tr>
<tr>
<td>GAPDH</td>
<td>M64176</td>
<td>496 BP</td>
<td>ACGACCCCCCTTCAATTGACC</td>
<td>ATCAGCCCAACAGCTTCCC</td>
<td>57.3</td>
</tr>
</tbody>
</table>

Abbreviations: BP, base pairs; \(T_a\), annealing temperature.
1R injury experienced a similar loss of contractile force initially (see fig 2) and then recovered complete contractile function with a time course indistinguishable from injured muscles that had not been irradiated.

Effectiveness of Gamma Irradiation

We confirmed the effectiveness of the gamma-irradiation protocol on the myogenic response by studying tibialis anterior muscles injected with cardiotoxin. Cardiotoxin induces extensive muscle degeneration, which is followed by regeneration through proliferation of myogenic cells.21,24,25 BrdU is a thymidine analogue that is incorporated into the DNA of replicating cells. After cardiotoxin injection (see Methods), the number of nuclei with BrdU uptake increased significantly 3 days after cardiotoxin injection (29% of nuclei were labeled), but was dramatically reduced in muscles that had been irradiated (0%). The number of centrally nucleated fibers peaked 14 days after cardiotoxin injection (40%), which was significantly different from controls (P < .05), but there were no centrally nucleated fibers detectable after irradiation (0%). Thus, our irradiation protocol disrupts mitotic activity (BrdU uptake) and the formation of newly formed muscle fibers, indicated by the presence of central nuclei. This inhibition of myogenic cell proliferation by irradiation is in agreement with previous reports.19,21

Effects of Irradiation on Centrally Nucleated Fibers and Messenger RNA After Injury

We counted centrally nucleated fibers to assess the contribution of the myogenic response to irradiated and uninjured tibialis anterior muscles recovering from 1R and MR protocols. Centrally nucleated fibers did not increase significantly at any time after the 1R injury, compared with controls, whether or not they were irradiated before injury (fig 3A). The MR injury protocol resulted in the appearance of centrally nucleated fibers on day 7 and a marked increase on day 14, most of which were lost over the following week (fig 3B). Irradiation of muscles before the MR injury completely eliminated the appearance of centrally nucleated fibers over the following 3 weeks (see fig 3B).

The messenger RNA (mRNA) for myoD, a muscle-specific transcription factor, was present in controls and increased on days 3 and 7 after the MR protocol (fig 4). A much smaller increase in the myoD PCR product was detected after the 1R injury. The mRNA encoding myogenin, another muscle-specific transcription factor, gave similar results. Increases in both mRNAs were nearly completely inhibited by irradiation before injury (see fig 4), although a very faint band for myogenin persisted in irradiated muscles at day 21 after the MR injury.

Labeling With Evans Blue Dye and Antibodies

Evans blue dye only enters fibers that sustain sarcolemmal damage as a result of disease or injury.26 We injected Evans blue dye into rats 24 hours before inducing injury to test whether injured muscle fibers that took up the dye disappeared as a function of time after injury, or if the dye persisted over the period in which recovery occurred (fig 5). After the 1R proto-

Table 2: Body Weights and Muscle Weights

<table>
<thead>
<tr>
<th>Weight</th>
<th>Injury Group (n=50)</th>
<th>Injury + Irradiation Group (n=50)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day 0</td>
<td>Day 3</td>
</tr>
<tr>
<td>Initial BW (g)</td>
<td>385 ± 7</td>
<td>397 ± 11</td>
</tr>
<tr>
<td>Final BW (g)</td>
<td>NA</td>
<td>409 ± 14</td>
</tr>
<tr>
<td>Injured TA weight (mg)</td>
<td>769 ± 69</td>
<td>817 ± 41</td>
</tr>
<tr>
<td>Opposite TA (control)</td>
<td>weight (mg)</td>
<td>708 ± 30</td>
</tr>
</tbody>
</table>

NOTE: Values are mean ± SD. There were 5 animals at each of the 5 time points, and 2 of each group (MR and 1R Injury, MR and 1R Inj + Irradiation). The irradiation-only group is not included in this table. No significant differences were found in body weight between groups or in rate of growth between groups.

Abbreviations: BW, body weight; NA, not available; TA, tibialis anterior muscle.
col, the number of Evans blue dye–positive fibers (30%) was not significantly different 1 week after injury (28%) (fig 6).

These findings support the idea that membrane resealing occurs after a contraction-induced injury from a single acute lengthening contraction. Although the number of Evans blue dye–positive fibers was even higher in the MR protocol on day 0 (46%) (see figs 5, 6), the number of Evans blue dye–positive fibers gradually diminished over time (36% on day 3, 10% on day 7).

We also monitored the presence of dystrophin, which is lost from fibers injured by the 1R protocol. Immunolabeling of injured muscles with antidystrophin antibodies confirmed the loss of dystrophin from the sarcolemma immediately after injury (26%) (see figs 5, 6). Three days after 1R injury, Evans blue dye–labeled fibers showed partial restoration of dystrophin (16%), and within a week (5%), dystrophin was nearly completely restored to fibers that retained Evans blue dye. Immediately after the MR protocol, only about half as many fiber had a loss of labeling for dystrophin at the sarcolemma (13%) (see figs 5, 6), which suggests that the nature of the damage to the sarcolemma caused by MR and 1R protocols differed substantially.

DISCUSSION

Proliferation of myogenic cells in adult skeletal muscle is presumed to be a prerequisite for the recovery of function after injury, but given the differences among injury protocols, it seems likely that the myogenic response may not be essential for recovery from all types of injury. We addressed this question by irradiating muscles before injuring them with 2 different protocols for contraction-induced injury that had different biomechanic features, but that led to a similar loss of contractile function. Recovery of muscle after injury by the MR protocol has already been shown to require proliferation of myogenic cells, which can be effectively blunted by gamma irradiation. We confirm that full recovery after multiple lengthening contractions requires a myogenic response and we report for the first time in the literature that recovery from a single contraction-induced injury does not.

The similar loss of contractile function after each of the protocols we used was the result of different biomechanic perturbations—one that causes injury by repeated lengthening contractions using a small strain, and the other by a single lengthening contraction using a large strain. We therefore expected to induce different mechanisms of recovery. The efficacy of our irradiation method in preventing proliferation of myogenic cells was key to the interpretation of our results. We tested this rigorously in muscle treated with cardiotoxin, which causes massive degeneration and regeneration of skeletal muscle and found that irradiation completely inhibited the formation of new skeletal myofibers by 2 measures: incorporation of BrdU into myonuclei and the appearance of centrally nucleated fibers.

By every measure, the MR injury protocol induced proliferation of myogenic cells. Myogenic markers were sharply elevated after MR injury. Likewise, the number of centrally nu...
nucleated fibers increased greatly after MR injury, consistent with the generation of new muscle fibers. The appearance of centrally nucleated fibers peaked at approximately 2 weeks after the MR protocol and dropped thereafter, suggesting that most of the newly formed, centrally nucleated muscle fibers were transformed into more mature fibers (with myonuclei located at the cell periphery, not centrally) over the following week. Such a transient peak in centrally nucleated fibers in the days after muscle injury has also been observed by others. Applying our irradiation protocol to muscle immediately before MR blocked the appearance of centrally nucleated fibers after the MR injury (see fig 3), suppressed the expression of myogenic markers—myoD and myogenin (see fig 4)—and attenuated functional recovery. This is consistent with previous reports that show local gamma irradiation of muscle before repeated lengthening contractions prevents proliferation of myogenic cells and regeneration of skeletal muscle.

By contrast, recovery of contractile force after injury by a single, maximal lengthening contraction (1R) is independent of the myogenic response, by the same criteria. The mRNAs encoding myoD and myogenin are only slightly elevated after the 1R protocol, but even this meager increase was suppressed by irradiation. Centrally nucleated fibers did not increase after 1R, with or without irradiation before injury. Nonetheless, recovery of contractile force in irradiated, 1R-injured muscle was not impeded and was, in fact, identical to that in unirradiated, injured tibialis anterior muscles.

Although myoD and myogenin are expressed specifically in myogenic cells and are down-regulated after muscle fibers form, they are expressed at low levels in mature myofibers. Their slight elevation after the 1R protocol may be part of the normal response of mature muscle to damage associated with injury. Alternatively, they may increase slightly because of the persistence of radiation-resistant stem cells within the muscle. If these confounding factors contributed to our results, their effects were clearly minimal.

Our 1R and MR protocols for injuring tibialis anterior muscles were based on established methods and were designed to induce a similar amount of force loss and comparable period for full recovery. Thus, the differences in the role of myogenic cell proliferation in the recovery after these 2 protocols could not simply be ascribed to quantitative differences in the extent or timing of the injury or the recovery period.

The fact that recovery from 1R injury occurs without a detectable contribution of myogenic cell proliferation indicates that other mechanisms must predominate. Our data suggest that 1 contributing mechanism is the ability of the sarcolemma to reseal shortly after 1R injury. Disruption and repair of the plasma membrane is a normal physiologic process that occurs in many cell types, including skeletal muscle. The sarcolemma is clearly breached by the 1R injury protocol, as indicated by Evans blue dye labeling of the myoplasm and a loss of dystrophin (see also fig 5). Remarkably, however, damaged myofibers recovering from the 1R protocol retain...
Evans blue dye. This suggests that the sarcolemma becomes impermeable to the dye shortly after the injury and that the damaged areas of the membrane are therefore "resealed." Consistent with this assumption, dystrophin is gradually restored in fibers that were damaged during the 1R protocol and that remain labeled with Evans blue dye. Resealing is likely to require proteins such as dysferlin and myoferlin, which have been shown to promote membrane repair after other forms of injury. 6,39

By contrast, most of the intracellular Evans blue dye label was lost over the recovery period after the MR protocol, consistent with the loss of the fibers that sustained damage to their sarcolemma during the repetitive injury. We propose that the majority of these injured fibers were replaced by myogenic cells. Mild eccentric stretch injury in skeletal muscle causes transient effects on tensile load and cell proliferation. Scan J Med Sci Sports 2004;14:367-72. Hameed M, Orrell RW, Cobbold M, Goldspink G, Harridge SD. Expression of IGF-I splice variants in young and old human skeletal muscle after high resistance exercise. J Physiol 2003;547: 247-54.


Suppliers

a. Charles River Laboratories, 251 Ballardvale St, Wilmington, MA 01887-1000.
b. Model T8904; NMB Technologies, 9730 Independence Ave, Chatsworth, CA 91311.
c. LabView version 4.1; National Instruments, 11500 N Mopac Expy, Austin, TX 78759-3504.
d. Harvard Apparatus, 84 October Hill Rd, Holliston, MA 01746.
e. Grass Instruments, Astro-Med Industrial Park, 600 E Greenwich Ave, West Warwick, RI 02893.
f. Bipolar series model HF 320; Pantak-Seifert GmbH & Co, Bogenstrasse 41, 22926 Ahrensburg, Germany.
g. Model 31006; PTW-Freiburg, Löracher Strasse 7, 79115 Freiburg, Germany.
h. Kite Manipulator; World Precision Instruments Inc, 175 Sarasota Center Blvd, Sarasota, FL 34240.
i. Carl Zeiss AG, 07740 Jena, Germany.
j. Invitrogen Corp, 1600 Faraday Ave, PO Box 6482, Carlsbad, CA 92008.
k. Applied Biosystems, 850 Lincoln Centre Dr, Foster City, CA 94404.
m. Oxford Molecular Ltd, Accelrys Ltd, 334 Cambridge Science Pk, Cambridge, CB4 0WN, UK.
n. Jackson ImmunoResearch Laboratories, PO Box 9, 872 West Baltimore Pike, West Grove, PA 19390.
o. Vector Laboratories, 30 Ingold Rd, Burlingame, CA 94010.
p. CorelDraw; Corel Corp, 1600 Carling Ave, Ottawa, ON, K1Z 8R7, Canada.
q. Sigma-Aldrich, 3050 Spruce St, St. Louis, MO 63103.
r. Millipore, 290 Concord Rd, Billerica, MA 01821.
s. SigmaStat; Systat Software Inc, 1735 Technology Dr, Ste 430, San Jose, CA 95110.

Objective: To assess reliability of isokinetic peak torque and work for knee flexion and extension.

Setting: University laboratory.

Participants: Eleven men and 7 women (mean age, 21y).

Interventions: Not applicable.

Main Outcome Measure: Peak torque and work for concentric and eccentric knee extension and flexion were recorded at 60°/s for 3 trials on 2 occasions. Intraclass correlation coefficient (ICC) values were calculated for the maximum and for the mean peak torque and work of the 3 repetitions.

Results: Reliability was “very high” for peak torque and work (ICC range, >.90). The SE measurements ranged between 5% and 10% of the initial values for both peak torque and work. The smallest change that indicates a real improvement for a single subject (smallest real differences) ranged from 12% to 25% for peak torque and work variables and from 25% to 30% for the peak torque ratios.

Conclusions: Isokinetic concentric and eccentric knee extensor and flexor strength variables are reliable when measured by the same examiner in asymptomatic subjects.

Key Words: Test-retest reliability; Reproducibility of results; Thigh; Treatment outcome.

© 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

ISOKINETIC DYNAMOMETRY IS A method commonly used in the assessment of muscle performance and pathology, both in research and in clinical practice.1 To be clinically meaningful, the assessment procedure must be reliable and sensitive enough to assess whether a finding indicates impairment, and to evaluate outcomes of therapeutic intervention. Values must be defined so as to provide guidance in deciding whether an observed change on reassessment is within the boundaries of assessment error or whether there has been a true change.

Most studies of the test-retest reliability of isokinetic knee strength2-7 have reported only relative reliability, such as the intraclass correlation coefficient (ICC). These statistics indicate the degree of the relationship between 2 or more measures,8 but they do not provide clinical guidance for assessing real changes.9,10 Absolute reliability reflects the magnitude of the differences between 2 measures.11 Examples of these statistics are the standard error (SE) of measurement and the smallest real difference.9,12,13 To be clinically useful, an assessment procedure must have a small measurement error to detect a real change. A retest difference in a subject with a value smaller than the SE of measurement is likely to be the result of “measurement noise” and is unlikely to be detected reliably in practice; a difference greater than the smallest real difference is 95% likely to be a real difference. A retest difference that lies between the SE of measurement and smallest real difference is less certain (between 68% and 95%), whether or not there is a real difference.11 It has been suggested that the SE of measurement can be used to indicate the limit for the smallest change that indicates a real improvement for groups of subjects, whereas for a single person, any retest measurement should exceed the smallest real difference to indicate a real change.9

Previous reports13-15 have suggested “high” to “very high” relative reliability for peak torque and work for the knee extensors and flexors. Absolute reliability has been documented in only a few studies related to isokinetic parameters of knee muscles.16,17 These studies determined the SE of measurement for peak torque of knee flexors and extensor during concentric reciprocal movements, ranging between 2.4 and 18.0Nm or 4.8% and 12.4% of the means.15,17 The absolute reliability of eccentric knee extensor and flexor contractions has not been determined, thus the smallest change necessary to indicate a real change is unknown.

Clinicians often use ratios of quadriceps and hamstring muscle peak torques to determine risk of injury or whether an athlete can safely return to a sport. Two ratios have been described, namely, the conventional ratio (Hc:Qc), which calculates the ratio between concentric hamstring peak torque to concentric quadriceps peak torque, and the dynamic control ratio (He:Qc), which calculates the ratio between eccentric hamstring peak torque and concentric quadriceps peak torque.1,18 To our knowledge, however, both the relative and the absolute reliability of these ratios have not been determined.

Our objectives in this study were to determine the relative and absolute test-retest reliability of: (1) peak torque and the work for isokinetic concentric and eccentric knee extension and flexion in uninjured subjects, and (2) of the Hc:Qc and He:Qc ratios.

METHODS

Subjects were included in the study if they participated in elite, subelite, or recreational running-related sports at least twice weekly and, at the time of testing, were participating fully in their planned sports training and/or competition. Ex-
clusion criteria included having sustained a lower limb, pelvic, or back injury in the past 6 months that prevented them from participating in their training for more than 1 week, or if they had any neurologic or systemic disease affecting a lower limb. They were asked not to do any strenuous exercise in the 48 hours preceding each testing day. Subjects underwent a musculoskeletal screening examination before the study to confirm their eligibility. All subjects read and signed an informed consent document approved by the University of Otago Human Ethics Committee.

Instrumentation

We used the Kinetic Communicator (KinCom) 500H isokinetic dynamometer and the KinCom system software to determine the peak torque and the work performed during the knee extension and flexion movements. The dynamometer was calibrated according to the manufacturer’s instructions before the testing. The system reliability of the KinCom dynamometer has been shown to be high, with ICCs of .99, .99, and .95 for the functions of lever arm position, lever arm velocity, and force measurement, respectively.19

Procedures

The same test procedure was performed on 2 separate occasions (day 1, day 2), 7 days apart. This time period between tests was used in previous studies of the reliability of isokinetic parameters.20 Furthermore, subject-linked variability is more likely to be controlled if testing is conducted on the same weekday relative to their respective sports programs. The tests were conducted during the same time of the day in order to reduce the effect of diurnal variation influences.

We tested only the dominant leg, defined as the preferred kicking leg. Subjects were seated in a comfortable position with the backrest angled at 100° to the seat. Self-adhesive (Velcro) straps were placed across the thigh, the pelvis, and chest to minimize body movements and to optimally isolate the movement to the knee joint.1,21 Subjects folded their arms across their chest and were not permitted to hold on to the equipment during the test. The mechanical axis of the dynamometer was aligned with the knee’s axis of rotation, with the lateral femoral epicondyle used as the bony landmark. The shin pad was placed 2 cm above the medial malleoli and the length of the lever arm was recorded. The weight of the leg was recorded and gravity adjustment was made using the computer software. The range of movement was from 0° (anatomic 0) to 85° of knee flexion.

Subjects performed 10 consecutive submaximal and 2 maximal concentric and eccentric contractions as a specific warm-up and also to become familiar with the movement. There was a 1-minute break between the warm-up and the testing.

Three concentric and eccentric maximum knee extensions were performed at 60°/s.22 Each concentric contraction was followed by an eccentric contraction,23,24 with a 15-second rest between the contractions.23,26 Subjects were instructed to extend the knee against the shin pad during concentric extensions and to resist the lever during eccentric extension. The dynamometer was then set for knee flexion and the same procedures were followed during the specific warm-up and data collection. Subjects were instructed to flex the knee during concentric flexion and to resist the dynamometer during eccentric flexion.

To reduce examiner variability, the same investigator (GS) conducted the tests for all subjects on both occasions. Subjects were told to abort the test if they felt any discomfort or pain. During the test, all subjects were given visual feedback from the system’s monitor. They were also verbally encouraged by the investigator to give their maximal effort.

Measures

Two variables were extracted for each of the direction (flexion, extension) and contraction types (concentric, eccentric): peak torque measured in newton meters, and work measured in joules. The peak torque is the single highest torque output recorded throughout the range of motion of each repetition.27 Work is defined as the output of mechanical energy and is represented by the area under the torque versus angular displacement curve.28 It is thus a “whole curve” parameter, rather than a “peak” or “single-point” parameter.29 In addition to presenting absolute measures of peak torque and work, these variables were normalized to body weight (in kilograms); this method has been used when comparing different groups of subjects30,31 and in determining outcomes of rehabilitation regimens.32

Statistical Analysis

From each set of 3 repetitions, we determined the means and maxima of the variables for each participant. Mean peak torque data were used to calculate the He:Qc and He-Qc. Group data are presented as mean ± standard deviation (SD). Data were analyzed using the ICC,31 with 95% confidence intervals (CIs) to determine relative reliability across the 2 test sessions for the respective movement directions and for the conventional and functional ratios.

Absolute reliability was determined with the SE of measurement and smallest real difference. These were calculated with the following formulas:8,12,30 SE of measurement = SD × (1−ICC), where SD is the mean SD of day 1 and day 2 to represent total measurement variability8,11, and smallest real difference = 1.96 × √2 × (SE of measurement). The SE of measurement and smallest real difference were also expressed as a percentage of the group mean for both test sessions for each of the variables.9,11

We calculated differences between day 1 and day 2 for all variables for each participant. The agreements between measurements of day 1 and day 2 were verified qualitatively using Bland and Altman plots.31 All statistics were performed using SPSS.8

RESULTS

Twenty healthy subjects (13 men, 7 women) volunteered for this study. Two men (aged 22y and 35y) did not attend the second session because of a soccer-related injury and a work commitment, respectively, and were excluded from the study. The mean age ± SD for the 11 men was 20±1 years and for the women it was 22±3 years. The mean body mass index ± SD for the men was 23.0±2.5 kg/m² and for the women was 22.3±2.6 kg/m². Two of the men were experienced in resistance training but none of the subjects was familiar with isokinetic dynamometry.

Table 1 shows the means and SDs of the peak torque and work measures. There were statistically significant differences (P<.05) between the values of day 1 and day 2 for maximal and mean peak torque and work of concentric extensor contractions. Table 2 shows the ICC values and their 95% CIs for all measurements. All peak torque and work measures had an ICC greater than .90 and thus were classified as “very high,” with the exception of work for the maximal concentric flexion (ICC = .88).32

The ICC for Hc:Qc was “low,” but was “high” for the He-Qc. The means ± SDs of the Hc:Qc on day 1 and day 2 were 62.84±9.15% and 62.52±0.41%, respectively. For the
Reliability of Isokinetic Knee Measures, Sole

He:Qc, means were 76.27 ± 13.66 on day 1 and 72.37 ± 13.51 on day 2. The reliability statistics of these parameters are shown in Table 2.

Bland-Altman plots for mean peak torque of concentric and eccentric extension and flexion contractions (Fig 1) illustrated a random relationship between the individual differences and the averages of the 2 testing sessions. The bias represents the average difference between day 1 and day 2 for the subjects, with a negative figure indicating that day 2 had a higher average than day 1. There were similar findings for plots of work (not illustrated here).

### DISCUSSION

Our subjects formed a heterogeneous group with regard to sex, experience with strength testing and training, and sports backgrounds and were a reflection of patients most commonly seen by clinicians at the community level. The men were generally stronger than the women, as is apparent from the Bland-Altman plots (see Fig 1). No other systematic differences existed between the sexes, however, thus we pooled the reliability statistics for the men and the women.

Correlations from .50 to .69 have been described as being “moderate,” those from .70 to .89 as being “high,” and from 0.9 and above as being “very high.” The ICC values for peak torque that we found can thus be described as being “very high” and in agreement with results of earlier studies with healthy, uninjured subjects. Phillips et al found slightly higher ICC values for the extensor and flexor maximal peak torque of 3 contractions (.98 for both directions). Harding et al also found high ICC values (.95) and lower SE of measurement (2.40–5.46 Nm) scores for peak torque when testing concentric knee flexion and extension. These 2 studies used reciprocal movements, that is, concentric extension testing concentric knee flexion and extension. These 2 studies used reciprocal movements, that is, concentric extension testing concentric knee flexion and extension.

### Table 1: Summary of Isokinetic Peak Torque and Work at 60°/s on 2 Occasions (N=18)

<table>
<thead>
<tr>
<th>Test Measurement</th>
<th>Day 1 Mean ± SD</th>
<th>Day 2 Mean ± SD</th>
<th>P</th>
<th>Day 1 Mean ± SD</th>
<th>Day 2 Mean ± SD</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Peak Torque (Nm)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximal*</td>
<td>121.39 ± 30.53</td>
<td>132.11 ± 31.45</td>
<td>.009</td>
<td>115.56 ± 26.37</td>
<td>125.50 ± 26.53</td>
<td>.005</td>
</tr>
<tr>
<td>Mean†</td>
<td>114.08 ± 29.08</td>
<td>123.31 ± 28.69</td>
<td>.004</td>
<td>110.53 ± 26.50</td>
<td>119.08 ± 27.46</td>
<td>.002</td>
</tr>
<tr>
<td><strong>Concentric flexor contractions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximal*</td>
<td>77.39 ± 20.62</td>
<td>81.28 ± 21.91</td>
<td>.158</td>
<td>86.5 ± 21.33</td>
<td>92.6 ± 26.06</td>
<td>.112</td>
</tr>
<tr>
<td>Mean†</td>
<td>71.18 ± 18.82</td>
<td>76.71 ± 20.03</td>
<td>.189</td>
<td>82.20 ± 19.85</td>
<td>87.73 ± 23.70</td>
<td>.056</td>
</tr>
<tr>
<td><strong>Eccentric extensor contractions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximal*</td>
<td>182.28 ± 50.25</td>
<td>181.06 ± 43.35</td>
<td>.989</td>
<td>162.28 ± 37.31</td>
<td>162.44 ± 33.88</td>
<td>.988</td>
</tr>
<tr>
<td>Mean†</td>
<td>172.49 ± 48.44</td>
<td>169.70 ± 43.05</td>
<td>.726</td>
<td>155.67 ± 38.24</td>
<td>154.44 ± 36.12</td>
<td>.944</td>
</tr>
<tr>
<td><strong>Eccentric flexor contractions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximal*</td>
<td>94.44 ± 26.25</td>
<td>94.00 ± 27.43</td>
<td>.616</td>
<td>105.00 ± 29.90</td>
<td>107.83 ± 31.98</td>
<td>.472</td>
</tr>
<tr>
<td>Mean†</td>
<td>86.77 ± 23.56</td>
<td>88.98 ± 27.87</td>
<td>.785</td>
<td>97.82 ± 28.81</td>
<td>102.06 ± 31.55</td>
<td>.250</td>
</tr>
</tbody>
</table>

*Maximum of 3 repetitions; †mean of 3 repetitions.

### Table 2: Reliability of Isokinetic Concentric and Eccentric Knee Measurements (N=18)

<table>
<thead>
<tr>
<th>Test Measurement</th>
<th>ICC (95% CI)</th>
<th>SEM (Nm)</th>
<th>SEM%</th>
<th>SRD (Nm)</th>
<th>SRD%</th>
<th>ICC (95% CI)</th>
<th>SEM (J)</th>
<th>SEM</th>
<th>SRD (J)</th>
<th>SRD%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Concentric extension contractions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximal*</td>
<td>.93 (.81 – .97)</td>
<td>8.21</td>
<td>6.48</td>
<td>22.75</td>
<td>17.95</td>
<td>.94 (.83 – .98)</td>
<td>6.50</td>
<td>5.39</td>
<td>18.01</td>
<td>14.94</td>
</tr>
<tr>
<td>Mean†</td>
<td>.95 (.85 – .98)</td>
<td>6.45</td>
<td>5.44</td>
<td>17.88</td>
<td>15.07</td>
<td>.96 (.88 – .98)</td>
<td>5.12</td>
<td>4.45</td>
<td>14.19</td>
<td>12.34</td>
</tr>
<tr>
<td><strong>Concentric flexion contractions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximal*</td>
<td>.93 (.80 – .97)</td>
<td>5.57</td>
<td>7.02</td>
<td>15.45</td>
<td>19.47</td>
<td>.88 (.69 – .96)</td>
<td>8.20</td>
<td>9.16</td>
<td>22.73</td>
<td>25.38</td>
</tr>
<tr>
<td>Mean†</td>
<td>.94 (.83 – .98)</td>
<td>4.74</td>
<td>6.41</td>
<td>13.14</td>
<td>17.78</td>
<td>.91 (.76 – .97)</td>
<td>6.55</td>
<td>7.73</td>
<td>18.15</td>
<td>21.42</td>
</tr>
<tr>
<td><strong>Eccentric extension contractions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximal*</td>
<td>.93 (.81 – .97)</td>
<td>12.24</td>
<td>6.74</td>
<td>33.92</td>
<td>18.67</td>
<td>.95 (.87 – .98)</td>
<td>7.87</td>
<td>4.84</td>
<td>21.81</td>
<td>13.43</td>
</tr>
<tr>
<td>Mean†</td>
<td>.94 (.85 – .98)</td>
<td>11.20</td>
<td>6.54</td>
<td>31.04</td>
<td>18.14</td>
<td>.96 (.90 – .98)</td>
<td>6.86</td>
<td>4.37</td>
<td>19.01</td>
<td>12.12</td>
</tr>
<tr>
<td><strong>Eccentric flexion contractions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximal*</td>
<td>.94 (.85 – .98)</td>
<td>6.48</td>
<td>6.88</td>
<td>17.97</td>
<td>19.07</td>
<td>.94 (.84 – .98)</td>
<td>7.46</td>
<td>7.00</td>
<td>2.69</td>
<td>19.41</td>
</tr>
<tr>
<td>Mean†</td>
<td>.92 (.79 – .97)</td>
<td>7.20</td>
<td>8.20</td>
<td>19.19</td>
<td>22.72</td>
<td>.93 (.80 – .97)</td>
<td>7.82</td>
<td>7.83</td>
<td>21.66</td>
<td>21.70</td>
</tr>
</tbody>
</table>

Abbreviations: SEM, standard error of measurement; SEM%, SEM as percentage of group average; SRD, smallest real difference; SRD%, SRD as percentage of group average.

*Maximum of 3 repetitions; †mean of 3 repetitions.
When interpreting changes in variables on reassessment after an intervention, clinicians must decide whether a true change has occurred or whether the changes reflect measurement noise or assessment error. If the change is less than the SE of measurement (and percentage of SE of measurement), it is most likely that the change reflects measurement noise and is unlikely to be of clinical significance. Based on the percentage of SE of measurement we found in this study, differences less than 8% and 10% for peak torque and work, respectively, should be considered to be measurement noise and most likely to be meaningless. Alternatively, the magnitude of a retest difference that is less than the SE of measurement and percentage of SE of measurement cannot be reliably determined. A retest difference that lies between the SE of measurement and smallest real difference is less certain (between 68% and 95%), whether or not there is a real difference. For concentric extension, concentric flexion, and eccentric extension peak torque differences between 8% and 20%, a clinical decision about whether a real change has occurred would be needed, taking into account all aspects of patient assessment (eg, prior familiarity with testing procedures). For eccentric flexion peak torque and all work measures, this would apply for differences up to 25%. Based on the percentage of smallest real difference in this study, a general guideline would be that a change of 15% to 20% is necessary for peak torque of concentric and eccentric extension and concentric flexion, and up to 23% for eccentric flexion peak torque to be 95% confident that there has been real change. For work measures, this difference is 12% to 15% for concentric and eccentric extension, but is 19% to 25% for concentric and eccentric flexion.

Reliability of eccentric flexion contractions has not been previously reported and our findings indicate that these also have a “very high” reliability for peak torque and work measures. There were no statistically significant differences for the group means of eccentric extensor and flexor peak torque and work between the 2 occasions. This differed from the differences found for peak torque and work for concentric extensor contractions. Subjects often reported that they found the eccentric contractions more difficult to perform than the concentric contractions in both movement directions. Eccentric contractions thus may require more skill and motor control. The increased group means of the variables for the concentric extensor contractions may indicate a learning effect for these among some of the subjects.

For both peak torque and work measures, the ICCs of mean variables (ICC range, .91 – .96) were slightly higher than for the maximal variables (ICC range, .88 – .95), with the exception of eccentric flexion contractions. There was a similar pattern for SE of measurement and percentage of SE of measurement, but these differences were very small. Because of the small differences in reliability, clinicians and researchers can

---

**Fig 1.** The differences between day 1 and day 2 sessions plotted against the means of the 2 test sessions for the peak torque of the men (○) and the women (●): (A) concentric extensor contractions; (B) eccentric extensor contractions; (C) concentric flexor contractions; and (D) eccentric flexor contractions.
choose to use either the maximum score of a set of repetitions or the mean of the set.

The reliability of the hamstring to quadriceps ratios has not, to our knowledge, been reported elsewhere. In this group of subjects, the He:Qc had “low” relative test-retest reliability with an ICC of .43. For the He:Qc, the ICC of .73 can be classified as “high” relative reliability. The 95% CIs for both ratios, however, are wide and the smallest real differences indicate that a difference of 28% and 30% for the He:Qc and Hc:Qc, respectively, is needed to be 95% confident that there has been a real change. Individual changes in extension and flexion of individual subjects from day 1 to day 2 were not equal, which would explain the low reliability of the ratios. These ratios thus cannot be considered in isolation when the outcomes of isokinetic tests of knee flexors and extensors in people who are active in sports are assessed.

Reliability of measurement can be affected by instrument, data processing, examiner and subject-linked variability, test procedure, and protocol errors. The reliability of the KinCom operating system has been shown to be excellent, with ICCs of .99 for force recorded at the strain gauge. When calibrated, the accuracy of the force measuring system was within 3% of an applied actual load. The KinCom 500H alignment of the mechanical axis of the dynamometer to the knee is performed manually and depends on visual placement of the seat relative to the dynamometer. The precise placement is not documented for subjects, thus it can affect intersession reliability. Variability between sessions may be smaller than our findings for dynamometers where placement is adjusted mechanically and where this can be individually recorded.

To reduce examiner variability in this study, 1 experienced examiner conducted all procedures and gave standardized instructions and verbal encouragement. Protocol-linked errors were kept to a minimum by following standardized procedures in regard to warm-up, stabilization, seating, and alignment of the dynamometer and lever. Various factors can affect subject-linked variability. It has been suggested that a prior familiarization session may decrease learning effects. In an earlier study, however, lower ICCs than what we found were found for eccentric and concentric knee contractions at 60°/s despite a pretrial familiarization session. Further, an investigation on the reliability of isokinetic isometric elbow flexion did not show a decrease in variability on 5 consecutive days. In clinical practice and in the screening, it is unlikely that a familiarization session would be held because of time commitments and financial costs. We thus made the decision not to include an additional familiarization session before day 1.

Study Limitations
This study included only healthy, young adults who participate in running-related sports at various levels. In a group of men and women with hemiparesis after stroke, the percentage of smallest real difference ranged from 26% to 33% for knee extension and from 39% to 55% for concentric flexion. Similarly, variability is likely to be greater in subjects with musculoskeletal injuries than in the subjects in this study. In the absence of reports of absolute and relative reliability in subjects with specific impairments, a decision about a minimal acceptable level of change still must be based mainly on clinical reasoning.

CONCLUSIONS
With the Kin-Com 500H isokinetic dynamometer, the relative reliability for all knee flexion and extension variables at a velocity of 60°/s was very high and was high for the Hc:Qc, but low for the He:Qc. In uninjured subjects, the smallest change that can indicate a real improvement (smallest real difference) ranges from 13 to 34Nm (15%–23%) for peak torque and from 14 to 23J (12%–25%) for work variables. The peak torque ratios are less sensitive in detecting a real change, with the smallest necessary change ranging from 25% to 30% of the initial value.

References

Suppliers
a. Chattecx Corp, 4717 Adams Rd, PO Box 489, Hixson, TN 37343.
b. Version 10.1; SPSS Inc, 233 S Wacker Dr, 11th Fl, Chicago, IL 60606.
Reliability of the Clinical Outcome Variables Scale When Administered Via Telephone to Assess Mobility in People With Spinal Cord Injury

Ruth N. Barker, PhD, Delena I. Amsters, MPhyt, Melissa D. Kendall, MHumSrv, Kiley J. Pershouse, BSocWk, Terry P. Haines, PhD

ABSTRACT. Barker RN, Amsters DJ, Kendall MD, Pershouse KJ, Haines TP. Reliability of the Clinical Outcome Variables Scale (COVS) when administered via telephone (TCOVS) to people with spinal cord injury (SCI).

Objective: To examine the equivalence reliability and test-retest reliability of the Clinical Outcome Variables Scale (COVS) when administered via telephone (TCOVS) to people with spinal cord injury (SCI).

Design: Equivalence (telephone administration vs in-person) and test-retest reliability study.

Setting: Assessments conducted in participants’ home environment.

Participants: Equivalence reliability was examined in a convenience sample of 37 people with a diagnosis of traumatic SCI who had been discharged from the Queensland Spinal Injuries Unit to the community. In a separate group of participants, test-retest reliability of COVS when administered via telephone was examined in 43 people with SCI who were randomly selected from the Queensland Spinal Cord Injuries Service records.

Interventions: Not applicable.

Main Outcome Measures: Reliability was assessed at the subscale and composite score level using intraclass correlation coefficients (ICC2,1) and Bland-Altman limits of agreement.

Results: Reliability was good for TCOVS and COVS for the composite score (ICC = .98), mobility subscale (ICC = .97), and ambulation subscale (ICC = .99). Reliability was also good for TCOVS test and retest assessments for the composite score (ICC = 1), mobility subscale (ICC = 1), and ambulation subscale (ICC = 1). For all comparisons, most data points were within the 95% limits of agreement and the width of limits of agreement were considered to be clinically acceptable.

Conclusions: The study findings confirm the equivalence and test-retest reliability of the TCOVS in an SCI population when administered by trained raters.

Key Words: Outcome assessment (health care); Rehabilitation; Reproducibility of results; Spinal cord injuries; Telephone.

© 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

PEOPLE WITH TRAUMATIC spinal cord injury (SCI) can now expect to live close to a normal life span; however, many may experience age-related health issues at an earlier age than the general population. To investigate these health issues, the Queensland Spinal Cord Injuries Service (QSCIS) is undertaking a longitudinal study to track changes in people with SCI across the lifespan. Assessment occurs via telephone interview, a method of data collection that is easily administered, cost effective, and inclusive of a geographically dispersed group of people. Many outcome measures that are included in the assessment are traditionally administered in-person, particularly functional measures. It is necessary, therefore, to confirm that information obtained by telephone interview is similar to information obtained by in-person observation (equivalence reliability) and that information obtained by telephone interview is consistent over time (test-retest reliability). Only then can any variability in assessment findings be attributed to real change, and not simply to variability in reporting.

One of the key functional outcome measures to be used in the longitudinal study is the Clinical Outcome Variables Scale (COVS), a clinician rated, composite measure of mobility that is used routinely across the continuum of care provided by QSCIS. It has also been used for general rehabilitation populations and specific diagnostic groups such as stroke, traumatic brain injury, amputations, and musculoskeletal injuries, and has been applied in acute, inpatient, and outpatient rehabilitation settings, and community settings.

In its current form, it consists of 13 items, which comprise rolling (2 items), lying to sitting (1 item), sitting balance (1 item), transfers (2 items), ambulation (4 items), wheelchair mobility (1 item), and arm function (2 items). All 13 items are rated by a clinician through observation and assessment of task performance according to detailed guidelines. Each COVS item is scored on a 7-point scale ranging from 1 (fully dependent mobility) to 7 (normal independent mobility). COVS scores are generally reported as a single composite score ranging from 13 to 91. In an SCI population, 2 subscales have also been reported, of which the first is a general mobility score ranging from 7 to 49 and the second is an ambulation score ranging from 5 to 35. As such, COVS offers a more relevant and more complete profile of mobility after SCI when compared with other mobility tools. A list of individual items and those that are included in the composite score and subscale scores is outlined in Table 1. The findings of previous studies have shown that COVS has acceptable construct validity when used in an SCI population, particularly when the 2 subscales are used. It has been found to discriminate between respondents across lesion levels, completeness of injury level, AmericanSpinal Injury Association impairment grade and walking status at the individual item, composite score, and subscale score levels. In addition, COVS has been found to have acceptable internal consistency and inter-rater reliability between hospital and community settings and greater sensitivity than the FIM instrument in measuring changes over time for assessment of mobility in people with SCI.
The purpose of the present study was (1) to examine the equivalence reliability of COVS scores obtained through in-person functional assessment and through telephone interview and (2) to examine test-retest reliability of the COVS when administered via telephone (TCOVS).

METHODS

We examined equivalence reliability and test-retest reliability in 2 study phases, using 2 different study samples and study procedures. Informed consent was obtained from participants prior to data collection. Ethics approval for this project was provided by the Princess Alexandra Hospital Human Research Ethics Committee.

Phase 1: Equivalence Reliability of COVS and TCOVS

Participants. We examined equivalence reliability of COVS in a convenience sample of 41 people recently discharged from the Queensland Spinal Injuries Unit after either their primary rehabilitation or their readmission for management of secondary conditions. All people who attended the Transitional Rehabilitation Program between September 2004 and November 2005 were invited to participate. This program assists people with SCI in the transition from hospital rehabilitation to community living. Because it is part of the QSCIS continuum of care, all people undergoing rehabilitation with QSCIS were eligible. The criteria for inclusion consisted of (1) a diagnosis of traumatic SCI, (2) aged over 15 years at time of injury, (3) discharged from hospital to a community setting more than 4 weeks previously, (4) living within the same home environment on each testing occasion, and (5) adequate cognitive and verbal skills to complete a telephone interview. The criteria for exclusion were (1) ventilator dependent, (2) difficult psychosocial situation (eg, substance abuse), (3) poor knowledge of English, or (4) limited access to a telephone. A total of 53 people were eligible to participate, of whom 41 agreed to participate and 37 completed both assessments within the 1-week period. The final study sample was considered representative of people with SCI on the QSCIS records in terms of age, sex, and level and completeness of injury. The average age ± standard deviation (SD) of participants was 41±18 years, of which 31 (84%) were men and 6 (14%) were women, 18 (49%) had tetraplegia and 19 (52%) paraplegia, and 19 (51%) had a complete injury and 18 (49%) an incomplete injury. The average time since injury was 8±3.7 months with the exception of 1 participant whose injury had occurred 41 years previously.

Procedure. Three to 4 weeks after discharge from hospital, we approached potential participants in person, provided with a verbal and written explanation of the study purpose, procedure, and consent process, and asked if they were willing to participate. To prevent temporal bias and order effects, each person who agreed to participate was randomly allocated (with the throw of a die) to 1 of 2 groups: group A (TCOVS assessment first, COVS assessment second) or group B (COVS assessment first, TCOVS assessment second). Each participant was then assessed within the home environment on 2 occasions by trained raters according to group allocation. Both assessments were completed within a 1-week period to ensure that the functional ability of participants would not change substantially. To prevent recall on the part of the trained raters, 1 rater conducted the COVS in-person assessments and the second rater conducted the telephone assessment. For the in-person assessment, participants were asked to perform each task and were then rated on their performance. The TCOVS assessments were performed via a guided telephone interview format in which written questions were provided to participants in advance. The rater encouraged participants to describe their current abilities and prompted them to do so with the statement “If I was to visit you today and ask you to perform these tasks, describe how you would do them.” Both raters were blind to previous assessments. All data were collected according to standard procedures by the 2 raters, both of whom were physiotherapists trained in the administration of both the COVS and the TCOVS and experienced in administration of the COVS within an SCI population. Time taken to conduct both in-person and telephone assessments was recorded.

Phase 2: Test-Retest Reliability of the TCOVS

Participants. We examined test-retest reliability of the TCOVS in a random sample of 43 people with SCI from the QSCIS records who met the same inclusion criteria as for phase 1. Prior to randomization, stratification into 6 groups according to year of injury occurred (<5y, 5–10y, 10–15y, 15–20y, 20–25y, >25y), with 7 participants then selected from each group. This selection process was used to ensure equivalence across strata and representation of the SCI population over the last 50 years. Contact details were obtained by searching QSCIS records and telephone directories. Potential partic-
patients were contacted by telephone, provided with a verbal explanation of the study purpose, procedure, and consent process, and invited to participate. The selection process was repeated until 7 participants were recruited to each group. Contact was made with 56 people, of whom 45 agreed to participate and 43 completed both assessments within an average time period of 15±2 days. Two people failed to complete the second assessment due to a major change in their personal situation (n=1) and because they were unavailable (n=1). The final study sample was considered representative of people with SCI on the QSCIS records in terms of age, sex, and level and completeness of injury. The average age of participants was 46±10 of which 35 (81%) were men and 8 (19%) women, 18 (42%) had tetraplegia and 25 (58%) paraplegia, and 21 (49%) had a complete injury and 22 (51%) an incomplete injury.

Procedure. Test-retest reliability of TCOVS was examined together with the core of established measures included in the QSCIS longitudinal study. Each participant completed assessments twice. To prevent recall, each occasion was separated by a 2-week period, as recommended by psychometric texts.10 It was not anticipated that the functional ability of participants would change substantially over this period, given the rate of progression in the SCI population observed clinically. Once verbal consent had been provided, an appointment for the first interview was made and information sheets, consent forms, and the interview questionnaire were sent to the participant. Immediately after the first interview, the participant was asked to destroy or dispose of the interview questions. A week later, the participant was contacted a second time to make an appointment for the second telephone interview, scheduled to occur 2 weeks after the initial interview. A copy of the interview questionnaire was then mailed to the participant a second time. At the beginning of the second interview, the participant was asked if they had experienced any major change in their health or life, physically or emotionally, to rule out the possibility that a functional change had occurred between the 2 testing sessions.

Analysis

We analyzed the data using SPSS software.12 Analysis was performed for the total composite score and 2 subscales consistent with the use and interpretation of COVS in clinical practice and research. Analysis of individual items was also performed in order to identify sources of error.

Equivalence reliability and test-retest reliability were assessed using 2 methods: intraclass correlation coefficients (ICCs)11 and the Bland-Altman limits of agreement.12 An ICC2,1 was selected on the basis that each participant was assessed by the same raters and that the raters were expected to represent the population of raters who may be using this procedure in practice (as we assume people to use this in practice will also be trained appropriately). Although the ICC is designed primarily for use with interval data, the ICC can be applied without distortion to data on the ordinal scale when intervals between such measurements are assumed to be equivalent.13 Therefore, the ICC was used on the assumption that the intervals between categories were equivalent, given that ordinal subscales were added together to get the overall score. The ICC values higher than .75 were considered to represent good reliability and values below .75 to represent poor to moderate reliability.14 Limits of agreement were examined by plotting the mean score and the difference between the 2 scores. The 95% limits of agreement were computed as the mean difference ±1.96 SD of the difference scores. Plots were generated using Microsoft Excel.8

RESULTS

Equivalence Reliability of TCOVS and COVS

The mean time period between assessments ± SD was 4±4 days. The average time taken to administer the TCOVS assessment (5min) was substantially less than the average time taken to administer the COVS in-person assessments (35min).

Correlation between the TCOVS and COVS was good for the composite score (ICC=.98), the general mobility subscale (ICC=.97), and for the ambulation subscale (ICC=.99). ICCs, mean scores, mean difference between paired scores, and limits of agreement for the composite scores and individual items are displayed in table 2. Correlation between TCOVS and COVS individual items was also good (ICC range .89-.99) with the

Table 2: Equivalence Reliability of TCOVS and COVS for the Composite Score, Mobility Subscale, Ambulation Subscale, and Individual Item Scores

<table>
<thead>
<tr>
<th>Item Aggregate</th>
<th>ICC2,1</th>
<th>TCOVS</th>
<th>COVS</th>
<th>TCOVS &amp; COVS</th>
<th>TCOVS – COVS</th>
<th>95% Limits of Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composite score</td>
<td>.98 (.96–.99)</td>
<td>.54±18</td>
<td>.55±20</td>
<td>.55±19</td>
<td>-.2±4</td>
<td>-9.8 to 7.5 to -12.1</td>
</tr>
<tr>
<td>Mobility subscale</td>
<td>.97 (.94–.98)</td>
<td>.37±12</td>
<td>.39±13</td>
<td>.38±13</td>
<td>-.2±3</td>
<td>-8.3 to 6.5 to -10.2</td>
</tr>
<tr>
<td>Ambulation subscale</td>
<td>.99 (.97–.99)</td>
<td>.11±9</td>
<td>.11±9</td>
<td>.11±9</td>
<td>0±2</td>
<td>-3.1 to 2.3 to -3.9</td>
</tr>
<tr>
<td>Individual items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roll right</td>
<td>.95 (.91–.97)</td>
<td>.53±23</td>
<td>.53±22</td>
<td>.54±22</td>
<td>-.0±2</td>
<td>-1.6 to 1.2 to -2.0</td>
</tr>
<tr>
<td>Roll left</td>
<td>.95 (.91–.97)</td>
<td>.54±23</td>
<td>.53±22</td>
<td>.53±22</td>
<td>-.0±2</td>
<td>-1.6 to 1.2 to -2.0</td>
</tr>
<tr>
<td>Supine to sit</td>
<td>.91 (.83–.95)</td>
<td>.51±22</td>
<td>.54±22</td>
<td>.53±22</td>
<td>.0±3</td>
<td>-2.2 to 1.6 to -2.7</td>
</tr>
<tr>
<td>Sitting balance</td>
<td>.52 (.24–.72)</td>
<td>.37±21</td>
<td>.49±23</td>
<td>.43±19</td>
<td>-.1±1</td>
<td>-5.3 to 4.1 to -6.5</td>
</tr>
<tr>
<td>Horizontal transfer</td>
<td>.91 (.84–.95)</td>
<td>.49±24</td>
<td>.51±23</td>
<td>.50±23</td>
<td>-.0±2</td>
<td>-2.1 to 1.6 to -2.7</td>
</tr>
<tr>
<td>Vertical transfer</td>
<td>.94 (.88–.97)</td>
<td>.32±23</td>
<td>.33±24</td>
<td>.33±23</td>
<td>-.0±1</td>
<td>-1.7 to 1.2 to -2.2</td>
</tr>
<tr>
<td>Walking</td>
<td>.97 (.94–.98)</td>
<td>.20±18</td>
<td>.19±19</td>
<td>.20±18</td>
<td>.0±1</td>
<td>-0.9 to .0±6 to -1.1</td>
</tr>
<tr>
<td>Walking aids</td>
<td>.99 (.98–.99)</td>
<td>.18±16</td>
<td>.18±16</td>
<td>.18±16</td>
<td>.0±0</td>
<td>-0.5 to 0±3 to -0.6</td>
</tr>
<tr>
<td>Walking distance</td>
<td>.96 (.93–.98)</td>
<td>.19±18</td>
<td>.20±20</td>
<td>.20±19</td>
<td>-.0±1</td>
<td>-1.1 to 0±8 to -1.4</td>
</tr>
<tr>
<td>Walking speed</td>
<td>.89 (.80–.94)</td>
<td>.20±17</td>
<td>.21±20</td>
<td>.20±18</td>
<td>.0±0</td>
<td>-1.8 to 1.3 to -2.3</td>
</tr>
<tr>
<td>Wheelchair mobility</td>
<td>.42 (.11–.65)</td>
<td>.61±68</td>
<td>.57±10</td>
<td>.59±7</td>
<td>0±4</td>
<td>-1.3 to 0±8 to -1.8</td>
</tr>
<tr>
<td>Right arm function</td>
<td>.89 (.79–.94)</td>
<td>.62±12</td>
<td>.60±14</td>
<td>.61±13</td>
<td>0±1</td>
<td>-1.1 to 0±8 to -1.5</td>
</tr>
<tr>
<td>Left arm function</td>
<td>.90 (.81–.95)</td>
<td>.60±16</td>
<td>.61±14</td>
<td>.60±15</td>
<td>-0±1</td>
<td>-1.4 to 1±0 to -1.8</td>
</tr>
</tbody>
</table>

NOTE. For each of the paired scores, the columns represent the ICC [95% confidence interval (CI)], the mean ± SD of the TCOVS, COVS, and TCOVS and COVS, the difference between TCOVS and COVS, and the 95% limits of agreement (CI).
exception of sitting balance (ICC = .52) and wheelchair mobility (ICC = .42), for which there was only moderate correlation.

Bland-Altman plots (fig 1) showed good agreement between TCOVS and COVS. The difference and 95% limits of agreement between the paired means were $-2 (-9.8$ to $6.5)$ for sitting balance item, $-2 (-8.3$ to $4.5)$ and for ambulation scale, which represents the mean of TCOVS and COVS paired scores. Dashed lines represent 95% limits of agreement.

TCOVS Test-Retest Reliability

Correlation between TCOVS test and TCOVS retest was good for the composite score (ICC = 1), the general mobility subscale (ICC = 1), the ambulation subscale (ICC = 1), and for all individual items (ICC = .95 for all items) (table 3). Bland-Altman plots (fig 2) showed good agreement between TCOVS test and TCOVS retest. The difference and 95% limits of agreement between the paired means were $0 (-4.1$ to $3.7)$ for the composite score, $0 (-2.5$ to $2.8)$ for the mobility subscale, and $0 (-3.1$ to $2.6)$ for the ambulation subscale.

DISCUSSION

The results of this study indicate that COVS assessment of people with SCI by trained raters is similar when COVS is administered via telephone interview and by in-person observation. Assessment via telephone interview was also found to be consistent over time. These findings support the use of TCOVS as a reliable measure of mobility when administered by trained raters. Coupled with the fact that it can be administered quickly and to a geographically dispersed group of people, TCOVS provides a convenient, time efficient, and geographically inclusive method for the collection of data across the lifespan for people with SCI. The reliability of the TCOVS also compares more than favorably with the reliability of the telephone version of the FIM instrument motor subscale, which appears to be the only other functional measure with demonstrated reliability when administered by telephone to people with SCI.

The equivalence reliability of TCOVS composite score and 2 subscales was good, illustrated by both high correlation coefficients and more than 90% of data points within the 95% limits of agreement. Although we would argue that the width of the limits of agreement is clinically acceptable, there may be some scope for reducing these limits, however. Inspection of the individual items revealed that 1 item, the sitting balance item, had only moderate correlation and wide limits of agreement. Although this may reflect real differences in the SCI person’s experience, recall, and perception of sitting balance compared with the physiotherapist’s direct observation, it may also reflect the shortcomings of the operational definitions for the sitting balance item. According to the standard COVS guidelines, sitting balance is assessed with legs over the side of the bed with no hands for support, a position that would not be used functionally by the majority of people with SCI. Surprisingly, no assessment is made of balance in long sitting or of sitting with 1 hand for support, both of which represent functional sitting balance strategies used by people with SCI and incorporated into training during rehabilitation. It is probable, therefore, that the reliability of this item could be improved by modifying the operational definitions to include balance in long sitting and with support of 1 hand, thereby making the item more relevant to everyday life for people with SCI. Given that the COVS is used widely in other diagnostic groups, however, further investigation of this item in diagnostic groups other than SCI would be necessary prior to item modification.

A second individual item that also appeared to show a sizeable amount of error between in-person and telephone assessment was the wheelchair mobility item. Interestingly, there was only moderate correlation between TCOVS and COVS scores, but acceptable limits of agreement. A small mean difference between TCOVS and COVS scores ± SD (0.4±0.9) for the wheelchair mobility item would suggest that the level of error was low. High mean scores for TCOVS ± SD (6.1±0.68) and COVS (5.7±1.0) on a scale of 7 suggests that scores were aggregated in the top part of the scale, however,
which led to a lower correlation coefficient. Therefore, rather than a high level of error, the low correlation coefficient reflects a skewed distribution likely to have poor sensitivity to further improvements, as previously reported.\(^8\)

### Study Limitations

It is important to highlight a number of methodologic limitations that may have influenced the results of this study. First, the findings may represent an underestimate of equivalence reliability because 2 different raters were used to conduct the assessments. On that basis, differences between paired scores may reflect not only a different method of administration, but also different raters or a combination of both. Even so, any difference is likely to be small, because interviewers were highly trained and experienced in the use of the TCOVS, familiar with functional abilities after SCI, and comfortable with administering this tool with people with SCI. Second, participants in this study had all been discharged from hospital in the previous 2 months and would therefore have undergone in-person COVS assessments immediately prior to discharge from hospital. As a result, it is possible that reliability estimates could have been biased. On the one hand, participants would have had a recent opportunity to attempt each task in a manner consistent with the operational definitions for COVS assessment. If, instead, the sample had consisted of people with SCI of longer duration, the reliability might have been lower for items that measured tasks not performed on a regular basis, such as vertical transfers. On the other hand, people with longer duration SCI may have been more aware of their exact functional limitations. In contrast, participants who had been recently discharged from hospital may still have been exploring their functional limits while taking on new challenges and opportunities within their home and community environment. Additionally, if this was the case, the period between assessments may not have been sufficiently brief to ensure that functional status had not changed. Ideally, future research should be conducted to confirm the equivalence reliability of TCOVS in a population of people with SCI of longer duration.

Test-retest reliability was good to almost perfect for the composite score, the subscales and individual item scores. This suggests that people with SCI are consistent in the way in which they score their mobility when both tests are performed by the same rater. Because the findings arose from people who have sustained SCI over the past 50 years, it is reasonable to suggest that they may be generalized to the population of people with SCI of traumatic origin.

The strength of the findings may be due to attributes of the COVS that are known to enhance the reliability of self-report measures.\(^17\) With the exception of the sitting balance item, the COVS includes functional items that are relevant to day-to-day life of people with SCI, with each item defined using nontechnical jargon. The strength of the findings may also be attributed to the standardized method of administration by a single trained rater who was a physiotherapist with experience in the use of the TCOVS and familiar with function after SCI. Further research would be required to determine the reliability of TCOVS when performed by physiotherapists unfamiliar with function after SCI or physiotherapists inexperienced in the use of COVS.

Alternatively, it is possible that the test-retest reliability is inflated due to a number of methodologic limitations. Participants may not have destroyed the original survey as requested, allowing them to repeat their responses from the first interview. In addition, the 2-week period between each test may have been brief enough to help ensure no change in mobility status, but may not have been sufficiently long to prevent the participant from recalling their responses on the first test. This is unlikely, however, given that TCOVS was administered as part of a large survey questionnaire that included many functional measures.\(^18\)

For both components of this study, like other similar studies, analysis was based on persons who agreed to participate, who had access to the telephone, and whose status could be considered uncomplicated. Although this is likely to have led to nonresponse bias, it is not clear how this bias would have influenced reliability. This highlights the need to ensure the reliability of TCOVS when used to assess the functional status of a more representative sample of people with SCI, including people who experience difficulties with the telephone interview.

---

### Table 3: Test-Retest Reliability of TCOVS 1 and TCOVS 2 for the Composite Score, Mobility Subscale, Ambulation Subscale, and the Individual Item Scores

<table>
<thead>
<tr>
<th>Item Aggregate</th>
<th>ICC(_{2,1})</th>
<th>TCOVS 1</th>
<th>TCOVS 2</th>
<th>TCOVS 1 &amp; 2</th>
<th>TCOVS 1 (–) 2</th>
<th>95% Limits of Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composite score</td>
<td>1.00</td>
<td>58±21</td>
<td>58±21</td>
<td>58±21</td>
<td>0.0±1.5</td>
<td>-4.1 (-3.1 to -5.2)</td>
</tr>
<tr>
<td>Mobility subscale</td>
<td>1 (0.99–1.00)</td>
<td>37±12</td>
<td>37±12</td>
<td>37±12</td>
<td>0.0±1.1</td>
<td>-2.5 (-1.8 to -3.2)</td>
</tr>
<tr>
<td>Ambulation subscale</td>
<td>1 (0.99–1.00)</td>
<td>15±11</td>
<td>15±11</td>
<td>15±11</td>
<td>0.0±1.5</td>
<td>-3.1 (-2.4 to -3.9)</td>
</tr>
<tr>
<td>Individual items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roll right</td>
<td>0.99 (0.98–0.99)</td>
<td>5.5±2.0</td>
<td>5.4±2.1</td>
<td>5.5±2.0</td>
<td>0.1±0.3</td>
<td>-0.6 (-0.4 to -0.8)</td>
</tr>
<tr>
<td>Roll left</td>
<td>0.99 (0.98–0.99)</td>
<td>5.3±2.2</td>
<td>5.3±2.2</td>
<td>5.3±2.2</td>
<td>0.0±0.3</td>
<td>-0.7 (-0.5 to -0.8)</td>
</tr>
<tr>
<td>Supine to sit</td>
<td>0.99 (0.98–1.00)</td>
<td>5.2±2.2</td>
<td>5.2±2.2</td>
<td>5.2±2.2</td>
<td>0.1±0.3</td>
<td>-0.6 (-0.5 to -0.8)</td>
</tr>
<tr>
<td>Sitting balance</td>
<td>0.99 (0.98–0.99)</td>
<td>3.6±2.4</td>
<td>3.7±2.4</td>
<td>3.6±2.4</td>
<td>0.1±0.4</td>
<td>-0.8 (-0.6 to -1.0)</td>
</tr>
<tr>
<td>Horizontal transfer</td>
<td>1 (0.99–1.00)</td>
<td>5.2±2.6</td>
<td>5.2±2.6</td>
<td>5.2±2.6</td>
<td>0.0±0.2</td>
<td>-0.4 (-0.3 to -0.5)</td>
</tr>
<tr>
<td>Vertical transfer</td>
<td>1.00</td>
<td>4.0±2.4</td>
<td>4.0±2.4</td>
<td>4.0±2.4</td>
<td>0.0±0.2</td>
<td>-0.5 (-0.4 to -0.6)</td>
</tr>
<tr>
<td>Walking</td>
<td>0.95 (0.91–0.97)</td>
<td>2.8±2.4</td>
<td>2.9±2.4</td>
<td>2.9±2.4</td>
<td>-0.1±0.8</td>
<td>-1.6 (-1.2 to -2.0)</td>
</tr>
<tr>
<td>Walking aids</td>
<td>1.00</td>
<td>2.7±2.4</td>
<td>2.7±2.4</td>
<td>2.7±2.4</td>
<td>0.0±0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Walking distance</td>
<td>1.00</td>
<td>2.5±2.2</td>
<td>2.5±2.2</td>
<td>2.5±2.2</td>
<td>0.0±0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Walking speed</td>
<td>0.96 (0.93–0.98)</td>
<td>2.6±2.3</td>
<td>2.6±2.3</td>
<td>2.6±2.3</td>
<td>0.0±0.9</td>
<td>-1.7 (-1.2 to -2.1)</td>
</tr>
<tr>
<td>Wheelchair mobility</td>
<td>0.98 (0.96–0.99)</td>
<td>6.4±0.7</td>
<td>6.3±0.7</td>
<td>6.3±0.7</td>
<td>0.0±0.2</td>
<td>-1.2 (-0.9 to -1.6)</td>
</tr>
<tr>
<td>Right arm function</td>
<td>0.99 (0.98–0.99)</td>
<td>6.2±1.3</td>
<td>6.3±1.3</td>
<td>6.2±1.3</td>
<td>0.0±0.2</td>
<td>-0.5 (-0.4 to -0.6)</td>
</tr>
<tr>
<td>Left arm function</td>
<td>1.00</td>
<td>6.2±1.1</td>
<td>6.2±1.1</td>
<td>6.2±1.1</td>
<td>0.0±0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

**NOTE.** For each of the paired scores, the columns represent the ICC (95% CI), the mean ± SD of the TCOVS 1, TCOVS 2, and TCOVS 1 and TCOVS 2, the difference between TCOVS and COVS and the 95% limits of agreement (CI).
format, such as people with hearing, visual, literacy, or communication problems or psychologic barriers. In addition, further research for this purpose could include administration of the TCOVS with or without a proxy.

CONCLUSIONS
The findings of this study show that TCOVS, when used by trained raters, offers a feasible and highly reliable method for assessing mobility in the home environment. The TCOVS can be considered, therefore, as a suitable tool to track outcomes for people with SCI across the lifespan.

Acknowledgments: We thank Jennifer Campbell, BPhTy, for expert advice regarding the COVS and to Sarita Schuurs, BPhTy, and Brooke Wadsworth, BPhTy, for assistance with data collection.

References
Symptom Burden in Persons With Spinal Cord Injury

Mark P. Jensen, PhD, Carrie M. Kuehn, MA, MPH, Dagmar Amtmann, Diane D. Cardenas, MD, MHA


Objective: To determine (1) the frequency, severity, and reported course of 7 symptoms in persons with spinal cord injury (SCI) and (2) the association between these symptoms and patient functioning.

Design: Postal survey.

Setting: Community.

Intervention: A survey that included measures of the frequency, severity, and recalled course of pain, fatigue, numbness, weakness, shortness of breath, vision loss, and memory loss, as well as a measure of community integration and psychological functioning was mailed to a sample of persons with SCI. One hundred forty-seven usable surveys were returned (response rate, 43% of surveys mailed).

Main Outcome Measures: The frequency and average severity of each symptom was computed, and the frequencies of each type of reported course were noted. Analyses estimated the associations among the symptoms, and between symptom severity and measures of patient functioning.

Results: The most common symptoms were pain, weakness, fatigue, and numbness. All symptoms were reported to remain the same or to get worse more often than they were reported to improve once they began. Pain, weakness, fatigue, and memory loss were the symptoms most closely associated with patient functioning.

Conclusions: Patients with SCI must deal with a number of secondary complications in addition to any disability caused by the injury itself. Of 7 symptoms studied, pain, weakness, and fatigue appeared to be most common and most closely linked to patient social and mental health functioning. Research is needed to identify the causal relationships between perceived symptoms and quality of life in patients with SCI and to identify effective treatments for those symptoms shown to impact patient functioning.

Key Words: Fatigue; Pain; Rehabilitation; Signs and symptoms; Spinal cord injuries.

© 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

ALTHOUGH SECONDARY complications in persons with spinal cord injury (SCI), such as pressure ulcers,1-2 bowel dysfunction,2-3 bladder dysfunction,4 urinary tract infections,5 and obesity,6 as well as a limited number of complications that are related to subjective experience, such as pain and fatigue, have received a great deal of attention in the research literature, the scope and impact of a number of other secondary complications associated with SCI have not yet been the focus of much empirical study.

As indicated above, of the symptoms reported by persons with SCI, pain and fatigue have received the greatest empirical attention. Research on pain, for example, indicates that about one third of patients with SCI report severe pain.7 Research also indicates that pain can have a significant negative impact on the daily functioning of persons with SCI, including psychological functioning8-12 and sleep.9,13 Moreover, research indicates that once pain problems develop, they rarely if ever resolve over time and in fact, for shoulder pain, show a tendency to worsen over time.14

Fatigue also is common in patients with SCI.15 Fatigue in persons with SCI has been typically discussed in terms of muscle fatigue and lack of endurance; that is, fatigue among persons with SCI has mostly been viewed as a physical construct.16-18 However, to our knowledge, the frequency of vision problems also indicates that pain can have a significant negative impact on the daily lives of persons with SCI.20 However, to our knowledge, no research has examined the extent to which fatigue improves or worsens over time in persons with SCI.

Prior research and our own clinical experience indicate that other symptoms, such as dyspnea (breathlessness), memory problems, vision problems, numbness, and perceived weakness, are reported by persons with SCI. Dyspnea has been reported in 24% and 6% of users and nonusers of motorized wheelchairs, respectively, in 1 sample of persons with SCI.21,22 Dyspnea also is common in patients with SCI.15 Like pain, research indicates that fatigue may have serious adverse impacts in the daily lives of persons with SCI.15,20 However, to our knowledge, no research has examined the extent to which fatigue improves or worsens over time in persons with SCI.

Concerning potential memory problems, the Model Spinal Cord Injury Systems (MSCIS) reported that 28% of patients with acute SCI had at least a minor brain injury, with 12% reported to have cognitive or behavioral changes associated with these injuries.23 Kreutzer et al24 performed neuropsychologic testing in 30 consecutive SCI patients with no obvious traumatic brain injury and found problems in visual learning, verbal learning, visual organization, and attention.24,28

Sherman et al25 reported a case of traumatic optic neuropathy documented by magnetic resonance imaging in a man with T4 paraplegia after a motorcycle crash who denied any loss of consciousness. Vaccaro et al26 similarly reported a case of delayed cortical blindness in a patient with a C1 burst fracture (that resolved somewhat after cervical fusion). These case reports are consistent with our clinical experience that some patients with SCI report significant problems with their vision. However, to our knowledge, the frequency of vision problems and the potential impact of these problems on functioning have not yet been systematically examined in patients with SCI.

The subjective sensation of numbness can also be reported by persons with SCI, particularly at or below the level of
injury, because of nerve damage or dysfunction. In a previous survey study of 84 community residents with SCI and chronic pain, 44 (52%) reported numb sensations, and 15 of these (18% of the sample) reported that the numb sensations were severe. As with most other symptoms that could be experienced by patients with SCI, the association between numb sensations and patient functioning has not been adequately studied.

Finally, perceived weakness has also been reported to be a common problem in patients with SCI. However, to our knowledge, its frequency and association with patient functioning relative to other symptoms, such as pain and fatigue, have not yet been examined in samples of patients with SCI.

Although a great deal of research has studied a number of secondary medical conditions in persons with SCI, such as pressure ulcers and bladder and bowel dysfunction, less research has studied the many other symptoms often reported by patients with SCI. There is especially a lack of empirical information about the frequency and severity of symptoms such as dyspnea, memory problems, visual problems, numbness, and perceived weakness in persons with SCI. Moreover, the associations between these symptoms and measures of patient functioning have not yet been systematically examined. In addition, although some information about the course of pain in persons with SCI is available, very little, if any, information is available concerning the course of other symptoms when they occur, either in the long (ie, since onset) or short (ie, over the past 6 mo) term. Therefore, we do not know if fatigue, memory loss, vision loss, numbness, and shortness of breath tend to improve, stay the same, or get worse when and if they develop in persons with SCI. If research shows that these symptoms improve after they develop, then clinicians can use this knowledge to inform patients that any problems with symptoms such as weakness or fatigue may be expected to resolve on their own. On the other hand, if research shows that these symptoms do not resolve on their own or even get worse over time, then this knowledge can be used by clinicians as an indication for effectively treating such symptoms as soon as possible when they are reported. Knowledge about the natural course of symptoms is also helpful for clinical researchers; any symptoms that are shown to be consistently associated with poor quality of life (QOL) and that either do not resolve or get worse over time should be targeted for treatment development, especially when and if there is a lack of effective treatments for these symptoms.

The primary purpose of this study was to increase our knowledge about the nature and impact of a number of symptoms commonly reported by persons with SCI. Specifically, we sought (1) to determine the relative frequency and severity of pain, fatigue, numbness, weakness, shortness of breath, vision loss, and memory loss, in a sample of persons with SCI; (2) to examine the extent to which these symptoms are reported to have a tendency to improve or resolve, get worse, or stay the same over time; and (3) to estimate the associations among the symptoms and the associations between the severity of these symptoms and both community integration and psychologic functioning in a sample of persons with SCI.

Based on previous research, we developed a number of hypotheses regarding the analyses about which we are reporting. First, we anticipated that the frequency of pain would be around 80% in this sample, based on the findings from a previous study that used a similar sample. We were also interested in exploring the frequency of other symptoms reported by the sample. We did not make any hypotheses concerning their specific frequency, because the frequency of occurrence of these symptoms had not been reported by previous researchers. In addition, we predicted that most respondents would report that their pain either remained stable or got worse over time, given that pain in persons with SCI is relatively refractory. Again, however, because of the lack of previous research on the natural course of other symptoms, we did not make any hypotheses concerning the reported course of weakness, fatigue, numbness, memory loss, vision loss, or dyspnea. Finally, because each symptom plays an important role in patient functioning, we predicted that the severity of each symptom would show independent (ie, when controlling for the other symptoms) associations with patient functioning.

METHODS

Subject Characteristics

Participants for the current study completed a survey study that focused on pain in persons with SCI. As described in our previous report of this sample, questionnaires were mailed to 339 persons with SCI in waves from September 5, 2002, to June 25, 2004, and surveys were returned from September 12, 2002, to August 13, 2004. A primary source of subjects for this study was the University of Washington MSCIS database, and therefore most surveys (320 [78%]) were sent to persons in Washington State. Other sources of data subjects included persons who learned about the study via word of mouth from study participants or from contact with one of the study investigators in their clinical practices. Each questionnaire was accompanied by a consent form and a cover letter inviting the potential study participants to participate in the study. Subjects were paid $25 for completing and returning the consent forms and survey. The study procedures were approved by the University of Washington Human Subjects Review Committee.

Measures

The survey assessed demographic information (age, education level, employment status, race and ethnicity, marital status) and descriptive information about the SCI (time since SCI, SCI level). Survey respondents were also asked about the presence, severity (on 0 [none] to 10 [very severe] numeric scales), and course over the long and short terms (whether each symptom had become worse, become better, or stayed the same since its onset; whether it had become worse, become better, or stayed the same in the past 6 months) of 7 symptoms that they might be experiencing: pain, weakness, fatigue, numbness, memory loss, vision loss, and shortness of breath. These symptoms were selected based on (1) our clinical experience that at least some persons with SCI report experiencing these and (2) previous research that indicates that these are indeed reported by patients with SCI.

All participants were asked to complete a measure of community integration and psychologic functioning. The 13-item Community Integration Questionnaire (CIQ) was used to assess community integration. The 3 integration domains of the CIQ include activities in the home (eg, meal preparation, housework), social activities (eg, leisure activities with others), and productive activity (eg, employment status). Evidence supports the reliability, discriminant validity, and construct validity of the CIQ scales.

We used the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36) mental health scale to assess psychologic functioning. This measure has shown reliability and validity. The mental health scale is scored so that it has a possible range of 0 to 100, with higher scores indicating better mental health.
Analyses

The response rate to the survey and basic demographic and descriptive information about the respondents were examined first to describe the sample. Next, the frequency (rates of respondents who provided a response of “1” or more when rating the severity of each symptom) of each symptom and the average severity (both for all participants who rated the symptom on the 0–10 scale and also only among those who endorsed the symptom—ie, rated the severity as being at least “1”) were computed for each of the 7 symptoms for descriptive purposes. The frequencies of each course type (worse, the same, better) associated with each symptom since onset and during the past 6 months were also computed to determine the extent to which each symptom is perceived to change over time.

Pearson correlation coefficients were computed between the severity ratings of each symptom to determine the extent of overlap among these ratings. A very high level of overlap (eg, correlation coefficients of .50 or higher among the ratings) would suggest a possible factor that might be contributing to symptom reporting, such as depression or somatization, whereas lower coefficients would suggest that there are unique factors that contribute to the experience and reporting of each symptom. Correlations were also computed between the symptom severity ratings and the criterion measures (3 subscales of the CIQ-SF-36 mental health scale) to estimate the associations between symptom severity and patient functioning. The association between symptom severity and patient functioning was initially examined by computing correlation coefficients between each severity rating and the measures of each functioning domain. To control for the increased risk of type I error rates associated with multiple tests, a Bonferroni adjustment was used (.05/28 tests = .0018) for determining whether an association was statistically significant.

Finally, 4 regression analyses were performed with the 4 measures of functioning as the criterion variables and the 7 symptom severity ratings as predictors entered as a block (ie, at the same time) to determine the extent to which each of the symptoms contributes independently to the prediction of the 3 community integration domains and psychologic functioning. Age, sex, and duration of SCI were controlled for in these analyses, because symptoms are likely to increase in frequency with age in the general population; both age and sex have been found to be associated with both community integration and mental health in SCI samples, and the duration of SCI could be related to both the development of symptoms (eg, shoulder pain) and the criterion variables.

RESULTS

Response Rate and Participant Characteristics

Of 339 surveys mailed, 27 were returned because the potential participant was no longer at the address on record, 2 were returned with information indicating that the subjects were deceased, 1 was returned with a note indicating that the patient was hospitalized and could not participate, and 2 potential participants wrote to say that they declined participation. Of the 309 possible surveys that might have been completed and returned (ie, excluding the 27 incorrect addresses, 2 deceased persons, and 1 unavailable person), we received 147 surveys with complete data (representing 48% of surveys that could have been returned and 43% of surveys mailed).

The mean age ± standard deviation (SD) of the study subjects was 48.8 ± 13.0 years (range, 21–88 y). As previously reported, there was a wide degree of variability in the number of years since SCI (mean, 16.6 ± 10.4 y; range, 3.2–57.4 y). The most frequent single cause of SCI was a motor vehicle collision (40.8%). Other causes included a fall (17.0%), a sports injury (7.5%), diving (7.5%), a gunshot wound (2.7%), or one of a variety of other causes (24.5%). The levels of injury reported by participants were C1–4 (15.6%), C5–8 (34.7%), T1–5 (10.2%), T6–12 (32.0%), and L1–S4/5 (7.5%). Thirty-nine percent of respondents reported that they had a complete injury, 50% reported that the injury was incomplete, and 11% reported that they did not know if their injury was complete or incomplete.

The majority (74.8%) of respondents were men, consistent with the higher frequency of males in the SCI population. Most (89.2%) of the survey respondents reported their ethnicity as white, with the remainder reporting their ethnicity as Native American (6.1%), Hispanic (3.4%), Asian or Pacific Islander (3.4%), or black (1.4%). The respondents reported having attended vocational or technical school (7.5%), having had some college (29.9%), being college graduates (27.9%), or having attended graduate school (19.0%). Relatively few (8.8%) had only a high school education or General Educational Development certificate, and only 6.8% reported that they did not graduate from high school.

Respondents to this survey were compared with nonrespondents on all demographic variables. The only significant difference to emerge was for education level; respondents to the current survey reported a higher frequency of having completed college or attended graduate school (46.9%) than persons who responded to the previous survey but not the current one (33.9%) (χ² test = 5.96, P < .05).

Table 1: Frequency and Course of 7 Symptoms in the Sample of Patients With SCI

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Frequency of Occurrence</th>
<th>Frequency of Severe Symptom</th>
<th>Average Severity</th>
<th>Course Since Onset (%)</th>
<th>Course Last 6 Months (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(% &gt; 1)</td>
<td>(% &gt; 6)</td>
<td>(mean ± SD)</td>
<td>Worse</td>
<td>Same</td>
</tr>
<tr>
<td>Pain</td>
<td>84</td>
<td>35</td>
<td>4.17 ± 3.00</td>
<td>40</td>
<td>39</td>
</tr>
<tr>
<td>Weakness</td>
<td>64</td>
<td>18</td>
<td>3.21 ± 3.10</td>
<td>35</td>
<td>30</td>
</tr>
<tr>
<td>Fatigue</td>
<td>67</td>
<td>18</td>
<td>3.23 ± 2.91</td>
<td>41</td>
<td>31</td>
</tr>
<tr>
<td>Numbness</td>
<td>66</td>
<td>38</td>
<td>4.41 ± 3.89</td>
<td>21</td>
<td>60</td>
</tr>
<tr>
<td>Memory loss</td>
<td>27</td>
<td>5</td>
<td>1.11 ± 2.14</td>
<td>49</td>
<td>39</td>
</tr>
<tr>
<td>Vision loss</td>
<td>27</td>
<td>4</td>
<td>0.99 ± 1.96</td>
<td>69</td>
<td>29</td>
</tr>
<tr>
<td>Shortness of breath</td>
<td>34</td>
<td>8</td>
<td>1.52 ± 2.01</td>
<td>28</td>
<td>53</td>
</tr>
</tbody>
</table>

*All symptoms rated by respondents on a 0 (none) to 10 (very severe) scale.
Frequency and Course of 7 Symptoms

The most common symptoms reported by the survey respondents were pain (84%), fatigue (67%), numbness (66%), and weakness (64%) (table 1). Shortness of breath (34%), vision loss (27%), and memory loss (27%) were reported less frequently. The most severe symptoms were pain (with 35% reporting pain levels of ≥7 on the 0–10 scale; mean severity rating, 4.71±3.14) and numbness (38% reporting numbness levels of ≥7 on the 0–10 scale; mean severity rating, 4.41±3.91). Weakness and fatigue were also reported as severe by a substantial subset of participants (both 18%; mean severity ratings, 3.21±3.08 and 3.23±2.89, respectively). Shortness of breath, memory loss, and vision loss were reported as less severe, on average.

All 7 of the symptoms were more often reported as staying the same (range, 24%–60%) or getting worse (range, 21%–69%) rather than getting better (range, 3%–35%) since the onset of the symptom. During the past 6 months, symptoms tended most often to be perceived as staying about the same (range, 53%–82%). When change did occur during the past 6 months, the symptoms were more likely to get worse (range, 14%–39%) than to improve (range, 0%–13%). In short, according to the respondents’ recollections, little change seems to have occurred in symptom severity once symptoms begin, but when change does occur, symptoms are more likely to get worse than to get better.

Associations Among the Symptom Severity Ratings

Strong correlations were found between fatigue and weakness (r=.63), fatigue and shortness of breath (r=.52), and shortness of breath and weakness (r=.52). All but 3 of the other associations between the symptom ratings were in the moderate range (r range, .20–.49), with the exceptions of the associations between memory loss and imbalance (r=.13), memory loss and numbness (r=.13), and memory loss and shortness of breath (r=.16). Overall, there was relatively little overlap among the symptom ratings, with the strongest association only indicating about 40% of variance shared between ratings of fatigue and weakness.

Associations Between Symptom Severity and Patient Functioning

The zero-order correlation coefficients between the symptom severity ratings and the 4 criterion measures of functioning used in this study are presented in table 2. These analyses indicate that the symptoms reported by this sample were more closely linked to psychologic functioning and social integration than they were to home competency or productive activity. In addition, the symptoms that were most closely associated with these domains of functioning were pain, weakness, fatigue, and memory loss. All statistically significant correlations were negative, indicating that the greater the symptom intensity, the lower psychologic functioning and social integration. Numbness, vision loss, and shortness of breath were only weakly and nonsignificantly associated with the measures of functioning.

Consistent with the correlation analyses reported earlier, the regression analyses indicated that symptom ratings did not contribute significantly to the prediction of the CIQ home competency scale or the CIQ productive activity. On the other hand, as a group, and controlling for age, sex, and SCI duration, the 7 symptoms were significantly associated with both the CIQ social integration and SF-36 mental health scales (table 3). As a group, the symptoms explained 23% of the variance of the CIQ social integration scale over and above statistically significant contributions of age, sex, and duration of SCI. Pain, weakness, memory loss, and vision loss all made independent and statistically significant contributions to the prediction of social integration. Fatigue showed a nonsignificant (P=.052) trend for making an independent contribution to the prediction of social integration. The regression analysis predicting the SF-36 mental health scale indicated that the symptoms explained 25% of the variance in this criterion measure over and above the contributions of age, sex, and duration of SCI. In this analysis, pain and memory loss made independent and statistically significant contributions, whereas the contributions of fatigue (β=-.19, P<.10) and vision loss (β=.17, P<.10) did not reach statistical significance (see table 3). In summary, the regression analyses indicated that when controlling for age, sex, duration of SCI, and each of the other symptoms, pain and memory loss made significant and independent contributions to the prediction of both social integration and psychologic functioning. Furthermore, weakness and vision loss made significant and independent contributions to the prediction of social integration, and fatigue showed a nonsignificant trend toward making an independent contribution to the prediction of both social integration and psychologic functioning.

DISCUSSION

This study provides new information concerning the frequency and severity of a number of symptoms and their association with functioning in persons with SCI. As predicted—and consistent with previous research—pain was very common (84%) and was reported to be more likely to stay the same or to get worse than to resolve or improve both since the onset of the symptom and within the past 6 months. We also found that each of the other 6 symptoms assessed were endorsed by at least some of the respondents. Of these, the most common were weakness, fatigue, and numbness (endorsed by about two thirds of the sample); memory loss, vision loss, and dyspnea were

---

**Table 2: Correlation Coefficients Between Patient-Rated Symptom Severity and Measures of Patient Functioning**

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Community Integration (CIQ)</th>
<th>Psychologic Functioning (SF-36 mental health scale)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Home Competency</td>
<td>Social Integration</td>
</tr>
<tr>
<td>Pain</td>
<td>−.05</td>
<td>−.33*</td>
</tr>
<tr>
<td>Weakness</td>
<td>−.15</td>
<td>−.33*</td>
</tr>
<tr>
<td>Fatigue</td>
<td>−.09</td>
<td>−.31*</td>
</tr>
<tr>
<td>Numbness</td>
<td>−.03</td>
<td>−.11</td>
</tr>
<tr>
<td>Memory loss</td>
<td>−.13</td>
<td>−.32*</td>
</tr>
<tr>
<td>Vision loss</td>
<td>.01</td>
<td>.00</td>
</tr>
<tr>
<td>Dyspnea (shortness of breath)</td>
<td>−.06</td>
<td>−.02</td>
</tr>
</tbody>
</table>

*P<.002.
endorsed by about one third of the sample. Like pain, once these other symptoms develop, they are reported to be more likely to stay the same or get worse than to resolve or get better. These findings have significant implications for the assessment and treatment of symptoms in persons with SCI.

Pain loss, in particular, is rarely reported as a direct consequence of SCI and yet, of the 27% of respondents who endorsed this symptom, 69% reported that it has become worse since onset, and 39% reported worsening vision over a relatively short time period, 6 months. The rate of vision difficulties in our sample is particularly notable given the low prevalence of vision problems reported in epidemiologic studies (usually <3%). Further study is needed to determine the extent to which the high rate of perceived vision problems in our sample is the result of medical conditions associated with aging (eg, presbyopia, cataracts, glaucoma) or associated medical conditions (eg, diabetes mellitus) or if there are additional conditions found in SCI, such as autonomic dysreflexia or postural hypotension, that could potentially be impairing vision in this population.

The frequency, severity, and refractory nature of pain reported by participants in this study provide additional support for the growing body of research indicating that pain is a serious problem among many persons with SCI. These findings support the need to develop more effective treatments for pain in persons with SCI. On a positive note, a review of the literature suggests that researchers may be now responding to this need. For example, recent pilot studies have reported on the potential efficacy of a number of innovative treatments, including electric oscillating field stimulation, magnet therapy, cognitive behavioral therapy, transcranial electric stimulation, exercise, hypnotic analgesia, and a number of medications such as gabapentin, lamotrigine, and amitriptyline. Much of this preliminary work is promising, although none of these treatments appears to produce a marked decrease in pain in all study participants. Our findings support the importance of this work to develop effective treatments for SCI-related pain or at least examine ways of combining existing treatments that have moderate efficacy, on average, into regimens and treatment protocols that could have a significant clinical impact on pain and suffering.

Although not as frequent or severe as pain in this sample, fatigue was also very common, was significantly linked to measures of patient dysfunction, and was reported to be very refractory among those who reported it. This finding underscores the potential negative impact that fatigue may have on the lives of many persons with SCI. Although there are some studies that have noted fatigue as a problem in SCI populations, very little is known about the nature, course, or treatment of fatigue in persons with SCI. The current findings, when considered in light of the limited research on SCI-related fatigue, suggest that such research is urgently needed.

Other symptoms and complaints, such as weakness, numbness, memory loss, vision loss, and shortness of breath, have not been previously systematically studied in samples of persons with SCI. The current findings indicate that, of these other symptoms, weakness and numbness are the most common, but that weakness and memory loss are most closely linked to important functioning domains. Thus, although numbness may be relatively common, it may not be particularly problematic for most patients with SCI. On the other hand, although memory loss was endorsed only by a little more than a quarter of participants, it showed a strong association with both mental health and social integration. Readers should keep in mind, however, that prevalence rates of cognitive problems in the general population, particularly memory loss, show a large variability in epidemiologic studies (probably because of differences in samples and assessment procedures), ranging from a low of 3.2% to as much as 53.8%. Thus, one cannot assume that the memory problems, or even the other symptoms, reported by participants in this sample are necessarily related to SCI.

Of course, patient-reported symptoms are not the only factors that contribute to decreased patient functioning. Previous research has identified a number of predictors of both psychological functioning (eg, recent urinary infections, spasticity, bowel problems, receiving family care, preinjury history of depression, marital status, employment status and community and social integration (eg, pain, neurologic classification and injury level, locus of control, family support) in subjects with SCI. As a group, this research is consistent with biopsychosocial models of functioning in persons with SCI, which argue that biologic, psychological, and social factors all can contribute to patient functioning. Patient care and functioning are likely maximized when clinicians assess and then treat as appropriate each of the biologic, psychological, and social problems or issues that may be contributing to dysfunction in any one patient. Perhaps pain, weakness, fatigue, and memory loss might be considered an important (but not the exclusive) part of addressing the biologic portion of the whole person with SCI.
Study Limitations

There are a number of limitations of this study that should be considered when interpreting the results. First, we only measured 2 primary domains of functioning (community integration, psychologic functioning). Although the measure of community integration used in this study, the CIQ, has evidence of good reliability and validity, it has been criticized given that scoring favors respondents who (1) have able-bodied friends over those whose friends are not able-bodied, (2) spend more time with friends than family, and (3) tend to do household activities alone rather than with others. It is possible that some of the symptoms that did not evidence significant associations with the criterion variables used would show significant and strong relationships with other functioning or QOL domains or with other measures of community integration or psychologic functioning. Future research on the nature, scope, and impact of symptoms in persons with SCI should therefore assess additional functioning domains, such as physical functioning, social role functioning, and participation, and examine the association between these symptoms and community integration and psychologic functioning assessed with different measures.

A second limitation of the current study is that only some secondary complications—those focusing on perceived symptoms—were assessed and not other secondary complications common to persons with SCI such as spasms, bladder and bowel dysfunction, pressure ulcers, insomnia, and obesity, among others. Thus, we were not able to determine the relative contribution of the symptoms studied to patient functioning over and above any effects of these other complications. Also, we did not examine any of the signs associated with the secondary complications to help verify and quantify the extent of these problems. Future researchers should include measures of additional secondary complications to examine their relative frequency, severity, and potential role in patient functioning.

It is also possible that some study participants may have been confused about the consequences of SCI (eg, paralysis) and the symptoms we assessed (eg, weakness or numbness). To the extent that such confusion occurred, the associations between the symptoms and functioning, in particular, may be overestimated. Future researchers could potentially control for extent of paralysis in the analyses when measures of this are available. Similarly, the domains of pain and other symptoms are multidimensional; they consist not only of magnitude, which was assessed in this study, but can include spatial, quality, and temporal components. Simple 0-to-10 magnitude estimates do not provide a thorough evaluation of these symptoms. On the other hand, the fact that significant effects emerged for pain and some of the other symptoms at all, despite this limitation, supports the importance of these symptoms as they relate to patient functioning. Nevertheless, future research could examine how the location(s) of these symptoms, their quality, and their frequency might moderate the associations between symptom severity and QOL or even themselves show significant direct associations with measures of patient functioning.

Another limitation concerns the cross-sectional nature of the data, which makes it impossible to draw causal conclusions regarding the potential impact of symptoms on patient functioning. It is possible, for example, that higher levels of psychologic dysfunction may make patients more aware, or more ready to endorse, symptoms such as pain or memory loss.

A number of factors limit the potential generalizability of the findings. For example, the respondents came from 1 geographic region and received initial care at the same center. Also, as a group, the respondents reported a higher level of education than the nonrespondents. Finally, although the response rate of 48% (of possible surveys that could have been returned) to our survey is standard for this type of study, it still represents only a subgroup of the population. Additional research in other SCI samples is necessary to determine the generalizability of the findings.

CONCLUSIONS

Despite the study’s limitations, the current findings: (1) replicate previous research concerning the frequency and refractory nature of pain in persons with SCI; (2) indicate that other symptoms, such as weakness, fatigue, and numbness, are very common; and (3) indicate that a number of these symptoms, including pain but also weakness, fatigue, memory loss, and vision loss are associated with measures of social integration and psychologic functioning. Research is needed to replicate these findings in other samples of persons with SCI. A better understanding of the types of symptoms persons with SCI experience and their frequency, severity and impact on day-to-day functioning is an important step toward identifying effective treatments and improving the QOL of those living with SCI.

Acknowledgments: We thank Amy Hoffman, MS, Emily Phelps, BS, Kristin McArthur, BS, Lindsay Washington, BA, Laura Nishimura, BS, Silvia Ammann, BS, and Kevin Gertz, BA, for assistance with data collection and management. We also thank Masuo Koyasu, PhD, and the Division of Cognitive Psychology in Education, Graduate School of Education, Kyoto University, for providing the physical resources that made the completion of this study possible, and 2 anonymous reviewers for their helpful comments on an earlier version of this study.

References


Reliability and Validity of the Incontinence Quality of Life Questionnaire in Patients With Neurogenic Urinary Incontinence

Brigitte Schurch, MD, Pierre Denys, MD, Chris M. Kozma, PhD, Pat Ray Reese, PhD, Terra Slaton, MS, Rich Barron, MS


Objective: To assess the reliability, validity, responsiveness, and minimally important difference (MID) of the Incontinence Quality of Life (I-QOL) questionnaire in patients with urinary incontinence due to neurogenic detrusor overactivity.

Design: Randomized, double-blind, multicenter, placebo-controlled study.

Setting: Eight centers across Belgium, France, and Switzerland.

Participants: Patients with urinary incontinence due to neurogenic detrusor overactivity inadequately managed on oral anticholinergics. Fifty-nine patients (spinal cord injury, n=53; multiple sclerosis, n=6) were enrolled.

Intervention: Single dose of botulinum toxin type A (Botox) (200 or 300U) or placebo.

Main Outcome Measures: I-QOL questionnaire completed at screening and over a 24-week post-treatment period.

Results: The Cronbach’s alpha ranged from .79 to .93, indicating that I-QOL is a reliable measure of QOL in neurogenic urinary incontinence patients. No item had more than 5.1% missing or out of range values. With the exception of 2 items, questions showed acceptable item-scale correlation and scaling success results varied by domain. Post-treatment correlations indicated acceptable construct validity. The I-QOL was responsive to improvements in symptoms. MID values ranged from 4 to 11 points.

Conclusions: Results suggest that I-QOL is a reliable, valid, and responsive measure of incontinence-related QOL in neurogenic patients.

Key Words: Bladder, neurogenic; Botulinum toxin type A; Quality of life; Rehabilitation; Reliability and validity; Urinary incontinence.

© 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

IT IS WELL RECOGNIZED that urinary incontinence can significantly decrease quality of life (QOL). This is not surprising, considering the impact poorly managed symptoms can have on an otherwise healthy person’s daily routine. Urinary incontinence sufferers report that they worry about going to new places and need to plan their activities around the requirement to have ready access to toilet facilities.1-2 For people with neurologic disease such as spinal cord injury (SCI) or multiple sclerosis (MS), however, whose lifestyle may well already be significantly restricted by their disease, urinary incontinence can further reduce their QOL.3-5

Assessment of the efficacy of treatments for urinary incontinence, especially pharmacologic interventions, is often based on measurement of clinical parameters, such as the frequency of incontinence episodes, and various urodynamic measures. Although these are relevant measures, particularly for the neurogenic patient where the physician’s priority is the management of urodynamic changes and reducing any associated risk of upper urinary tract complications,6 the importance of evaluating patient reported outcomes is becoming more widely recognized.7 Direct measurement of the effect of treatment on QOL is particularly important because the relationship between subjective and objective responses to treatment and QOL changes is complex.8-10

A variety of tools are available for assessing QOL changes as a consequence of urinary incontinence and its treatment.7 A number of generic questionnaires are available that measure general health-related quality of life (HRQOL), for example, the Medical Outcomes Study Short-Form 36-Item Health Survey (SF-36).11 When used in patients with urinary incontinence, these questionnaires provide a general assessment of the patient’s overall HRQOL. However, they tend to be relatively insensitive to treatment changes because they may not contain specific issues most pertinent to incontinence. In addition, generic measures would be assessing the effects of the underlying primary condition, such as SCI or MS, rather than specifically urinary incontinence. It is likely that the relatively small number of patients with this condition, compared with the overactive bladder population as a whole, has led to such people generally being excluded from the studies performed to develop and validate incontinence-specific questionnaires. Ideally, a measure used in the assessment of interventions should be responsive to changes in disease and assess factors that are clinically relevant to the specific patient group, as well as being reliable and valid for use in that population.12

A number of disease-specific questionnaires have been developed that specifically assess the effects of urinary incontinence on QOL.7 Only 2 such questionnaires are “highly recommended,” that is, there are published data providing evidence of their reliability, validity, and responsiveness to change, and are suitable for both men and women—the King’s Health Questionnaire and the Incontinence Quality of Life Questionnaire.
(I-QOL) questionnaire. At the time the current study was conducted, responsiveness data were only available for I-QOL.

The I-QOL questionnaire is an incontinence-specific assessment tool that has shown high levels of validity and reliability in stress incontinence and overactive bladder. This questionnaire was developed using information obtained from interviews with men and women suffering from urinary incontinence, and was evaluated in studies both in the United States and in Europe. Although the content validity of the questionnaire has been shown in neurogenic detrusor overactivity, these patients were not included in either the development or the evaluation of the questionnaire. It has, therefore, yet to be shown that the I-QOL is a valid tool for evaluating QOL changes in these patients. This is of particular importance given that various characteristics of urinary incontinence in people with neurogenic bladder disease are likely to differ from patients with non-neurogenic incontinence. For example, questions relating to limitations in lifestyle associated with urinary incontinence may be less relevant to people whose lifestyle is severely affected by their primary disorder (eg, patients confined to a wheelchair).

This study assessed the reliability, validity, responsiveness, and minimally important difference (MID) of the I-QOL in patients with neurogenic urinary incontinence. MID was evaluated to provide an understanding of how to interpret the I-QOL data generated in the neurogenic bladder population. Data were analyzed from a double-blind, randomized, multicenter, placebo-controlled study of botulinum toxin type A (BTX-A) in patients with SCI or MS.

METHODS

Patients and Treatment

This study analyzed data obtained in a 26-week, double-blind, multicenter, randomized, placebo-controlled, parallel-group study, which evaluated the effect of 2 dose levels of BTX-A in 59 patients with urinary incontinence due to neurogenic detrusor overactivity who were inadequately managed on oral anticholinergics. The study was done in accordance with independent ethics committee regulations in each country—prior approval was obtained prior to actual study initiation. This was in compliance with Good Clinical Practice and the Declaration of Helsinki (1996). Written informed consent was obtained from all study patients.

After a screening visit at week 2, eligible patients were randomized in a 1:1:1 ratio to receive single treatments of BTX-A (Botox) (19 subjects at 200U, 19 subjects at 300U) or placebo (n=21). Randomization block size was not divulged to maintain blinding. Doses were administered to the detrusor muscle as 30 injections, each of 1mL, using cystoscopic guidance. Number of daily episodes of urinary incontinence was monitored via patient-completed voiding diaries at screening and at weeks 2, 6, 12, 18, and 24 post-treatment. QOL was assessed at each follow-up visit using the I-QOL questionnaire. Full details of study design, treatment, and efficacy and safety measures are described elsewhere.

Clinical Efficacy Assessments

The primary efficacy variable was the change from baseline in the daily number of urinary incontinence episodes. The number of daily urinary incontinence episodes was recorded by patients in voiding diaries, completed for the week prior to each study visit. Baseline values were determined from data recorded by patients for the 2 weeks prior to receiving treatment. Key urodynamic parameters monitored throughout the study were maximum cystometric capacity, reflex detrusor volume, and maximum detrusor pressure during bladder contraction, defined using the standard International Continence Society classifications (table 1).

QOL Assessments

We assessed QOL using 2 measures: a disease-specific questionnaire, the I-QOL, and a generic HRQOL measure, the SF-36. The I-QOL was administered at screening and weeks 2, 6, 12, 18, and 24 and the SF-36 was administered at screening and week-24 post-treatment. Where I-QOL or SF-36 values were missing, a last observation carried forward approach was used.

The I-QOL consists of 22 items evaluating concerns specifically relating to incontinence (table 2). Subjects assign a value on a 5-point scale from 1 (extremely) to 5 (not at all) for each item. For all items, higher scores indicated better incontinence-related QOL. The 22 items are divided into 3 subscales: avoidance and limiting behavior; psychosocial impact; and social embarrassment. The I-QOL was scored according to the developer’s instructions (I-QOL user’s manual), with an additional requirement that at least 50% of the items in a scale had to be completed, similar to the SF-36 scoring requirements. A mean score for each subscale is calculated, as well as a total score for all 22 items. The scores are then transformed to a 0 to 100 scale for ease of interpretation using the following formula:

\[
\text{Scale score} = \frac{\text{The sum of the items} - \text{lowest possible score}}{\text{Possible raw score range}} \times 100
\]

The SF-36 consists of 36 items, scored according to the developer’s instructions. It measures 8 generic health concepts: physical functioning; role-physical; bodily pain; general health; vitality; social functioning; role-emotional; and mental health. Each concept is scored separately and can also be combined to give a physical component summary score and mental component summary score.

Psychometric Properties Assessed

Reliability. We determined the internal consistency reliability of the I-QOL by calculating values for the Cronbach \( \alpha \) for the I-QOL total score and the 3 subscales. This reliability calculation is derived from the Spearman-Brown prophecy formula, which is driven by the interitem correlations with the scale and the number of items. Values above 0.7 are generally considered acceptable.

Table 1: Definitions of Key Urodynamic Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum cystometric capacity</td>
<td>Volume at which involuntary voiding occurs and/or filling of bladder stops during cystometry (in milliliters)</td>
</tr>
<tr>
<td>Reflex detrusor volume</td>
<td>Infused volume required to induce the first hyperreflexia bladder contraction during cystometry (with or without voiding) (in milliliters)</td>
</tr>
<tr>
<td>Maximum detrusor pressure during bladder contraction</td>
<td>Maximum detrusor pressure during the bladder contraction (in cmH2O)</td>
</tr>
</tbody>
</table>
Validity. Because the I-QOL was developed in a population that excluded neurogenic urologic problems, it is important to evaluate whether items appear to be changing over the course of treatment. For example, on review, the I-QOL item focusing on concern about incontinence when standing or sitting may not be appropriate for this study population, particularly for those with SCI. Although large numbers of missing values would be one indicator of an item that was not meaningful to patients, we also evaluated the mean and median scores at screening and week 24 for both treated and untreated groups. Items that resulted in similar mean scores at the 2 time points across all groups would indicate that the item was either unresponsive to treatment or inappropriate for the study population. We selected an arbitrary 1-point difference in the mean and median scores at screening and week 24 in order to identify those items that were unchanged in patients who responded to therapy. There was an a priori assumption that on a 4-point scale a change of 1 point or more from baseline indicates a response and no change indicates no response.

We assessed the validity of summing the I-QOL items by determining: the mean and variance for items within each domain; whether each item contributed roughly equal portions of information to the total scale score; and the item-to-scale correlations. Item-to-scale correlations (corrected for overlap) were considered satisfactory if a correlation of 0.4 or more with the hypothesized scale was obtained.23,24 To achieve item discriminant validity (ie, scaling success), an item must have a higher (probable scaling success) or a statistically significant and higher (definite scaling success) correlation with its hypothesized scale than with other scales included in the measure.25

We assessed the construct validity of the I-QOL by analysis of the correlations between I-QOL scores (total and subscale scores) and SF-36 domain scores and clinical efficacy measures. Correlations were assessed using Pearson correlations. The strongest correlation was expected to be between the I-QOL and the number of involuntary losses of urine.

Responsiveness. The responsiveness, that is, the potential of the tool to detect relevant changes in patients, of the I-QOL was assessed. We assessed the potential for responsiveness by determining the floor and ceiling effects for each domain and the total score. A criterion of less than 10% was used to indicate that a domain had the potential to show both improvement and decline.26

The responsiveness of the I-QOL to differences between known groups (ie, discriminant validity) was also evaluated by comparing the mean I-QOL scores for patients grouped according to whether they had a 25% or greater decrease, no change (0%–24% increase or decrease), or a 25% increase in incontinence episodes from baseline. The statistical significance of differences between the means for each patient category was determined using Kruskal-Wallis tests for the omnibus test of differences between treatment groups; Wilcoxon-Mann-Whitney pairwise tests were used to compare active treatment with placebo when the omnibus test was significant. Sensitivity to change was also assessed by calculating the responsiveness statistic and the standardized response mean.27

Determination of the MID in I-QOL Score From Baseline

The MID is defined as “the smallest difference in score in the domain of interest which patients perceive as beneficial and
which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient’s management.28(p408) The MID is the threshold beyond which a response can be considered as “clinically meaningful” or important (ie, not random variation).

MID can be determined using distribution-based methods that rely on statistical characteristics of the measure. In this study, we used 2 distribution-based approaches. One was based on effect size and standard deviation (SD). Effect size is used to compare 2 subgroups or measure change over time to the SD at baseline.29,30 Conventional benchmarks can be used to interpret effect size. An effect size between .20 (0.2 SD) and .49 are considered to be a small effect size and between .50 (0.5 SD) and .79 to be medium effect size. Therefore, 0.2 and 0.5 SDs at baseline were calculated to determine the change in score that would be required to achieve a small and medium effect size.

The second method was based on standard error (SE) of measurement. The SE of measurement has been advocated as a distribution statistic largely unaffected by the distribution of the sample and relatively constant across the range of abilities in a given population.29,30

In this study, we calculated the MID in I-QOL score from baseline using 3 different definitions of MID: change corresponding to a small effect size (ie, 0.2 SD at baseline); change corresponding to a medium effect size (ie, 0.5 SD at baseline); and change corresponding to SE of measurement (1.0 SE criterion) at baseline.29-31

RESULTS

Study Population

Fifty-nine patients, with a mean age of 41.2 years (range, 20–72y), were enrolled in the study. The majority of subjects were white (93.2%) and male (61.0%). SCI was the cause of detrusor overactivity in 53 patients (90%), with the remaining cases arising as a result of MS (6 patients; 10.2%). The mean duration of detrusor overactivity was 63 months (range, 3mo to 24y).

Reliability

I-QOL total score and the 3 subscales showed high internal reliability, with values for the Cronbach α exceeding .70 for all 4 scores: total score, .93; avoidance and limiting behavior, .85; psychosocial impact, .89; and social embarrassment, .79.

Validity

Data for analysis included 53 subjects who provided complete I-QOL data and a further 3 subjects who completed at least half of the items. In addition, no item had more than 5.1% of missing or out of range values and the distribution of responses was acceptable. Additionally, the 2 BTX-A groups showed that most item means increased by at least 1 point over the course of the study, showing that these items of the questionnaire are responsive to changes experienced by patients as a result of treatment. Six items, however, failed to show a difference of at least 1 point for either treatment group: “I worry about coughing/sneezing because of my incontinence”; “I have to be careful standing up from sitting”; “I have difficulty getting a good night’s sleep”; “I don’t feel free to leave home for long periods”; “My incontinence limits my choice of clothing”; and “I worry about others smelling urine on me.”

Evaluating the validity of summing the I-QOL items, the range of item means for the domains was similar for the psychosocial impact and social embarrassment domains and somewhat larger than is usually desired for the avoidance and limiting behavior domain (avoidance and limiting behavior, 0.27–4.13; psychosocial impact, 2.32–3.20; social embarrassment, 1.80–2.99). The range of item-to-total score correlations was also similar for all items from all 3 domains (avoidance and limiting behavior, .45–.61; psychosocial impact, .48–.78; social embarrassment, .49–.73). Item-to-scale correlations were acceptable (ie, >.40) with the exception of the question “I worry about my incontinence getting worse as I get older,” each of the items showed acceptable item-to-scale correlations (see table 2). In addition, the values for the item-to-scale correlations within each domain were similar for each item, indicating that items are approximately equally weighted.

Two questions in each domain correlated equally or more highly with another domain (see table 2). These questions were: “It’s important to plan every detail in advance because of my incontinence” and “I have to watch how much I drink because of my incontinence” from the avoidance and limiting behavior domain; “Incontinence is always on my mind” and “My incontinence makes me feel unhealthy” from the psychosocial impact domain; and “I worry about others smelling urine on me” and “I worry about my incontinence getting worse as I get older” from the social embarrassment domain. The scaling success using the criterion of “2 standard errors” ranged from 10.0% to 18.8%, although probable scaling success (ie, items correlated higher with their hypothesized scales than with other scales) ranged from 60.0% to 77.8%.

Correlations between SF-36 scores and I-QOL scores at the end of the study (ie, week 24) were substantial for most SF-36 domains, and tended to be stronger and more likely to be significant than those at screening (table 3). Correlations with

<table>
<thead>
<tr>
<th>SF-36 Domains</th>
<th>Screening ALB</th>
<th>Screening PSI</th>
<th>Screening SE</th>
<th>Screening Total</th>
<th>Week 24 ALB</th>
<th>Week 24 PSI</th>
<th>Week 24 SE</th>
<th>Week 24 Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical functioning</td>
<td>-.07</td>
<td>-.03</td>
<td>.02</td>
<td>-.04</td>
<td>.22</td>
<td>.26</td>
<td>.26</td>
<td>.26</td>
</tr>
<tr>
<td>Role-physical</td>
<td>.15</td>
<td>.36*</td>
<td>.26</td>
<td>.30*</td>
<td>.22</td>
<td>.37*</td>
<td>.19</td>
<td>.29*</td>
</tr>
<tr>
<td>General health</td>
<td>.10</td>
<td>.20</td>
<td>.04</td>
<td>.14</td>
<td>.25</td>
<td>.44*</td>
<td>.29*</td>
<td>.36*</td>
</tr>
<tr>
<td>Vitality</td>
<td>-.04</td>
<td>.09</td>
<td>.02</td>
<td>.04</td>
<td>.36*</td>
<td>.54*</td>
<td>.39*</td>
<td>.47*</td>
</tr>
<tr>
<td>Social functioning</td>
<td>.27*</td>
<td>.44*</td>
<td>.27*</td>
<td>.38*</td>
<td>.43*</td>
<td>.54*</td>
<td>.52*</td>
<td>.52*</td>
</tr>
<tr>
<td>Role-emotional</td>
<td>.29*</td>
<td>.38*</td>
<td>.32*</td>
<td>.38*</td>
<td>.25</td>
<td>.42*</td>
<td>.25</td>
<td>.34*</td>
</tr>
<tr>
<td>Mental health</td>
<td>.20</td>
<td>.31*</td>
<td>.24</td>
<td>.29*</td>
<td>.45*</td>
<td>.59*</td>
<td>.45*</td>
<td>.54*</td>
</tr>
<tr>
<td>PCS</td>
<td>.13</td>
<td>.11</td>
<td>.07</td>
<td>.12</td>
<td>.24</td>
<td>.32*</td>
<td>.20</td>
<td>.27</td>
</tr>
<tr>
<td>MCS</td>
<td>.22</td>
<td>.39*</td>
<td>.29*</td>
<td>.34*</td>
<td>.39*</td>
<td>.57*</td>
<td>.41*</td>
<td>.50*</td>
</tr>
</tbody>
</table>

Abbreviations: ALB, avoidance and limiting behavior; MCS, mental composite scale; PCS, physical composite scale; PSI, psychosocial impact; SE, social embarrassment.

*Significant Pearson correlations (P < .05); last observation carried forward procedure applied for missing visits.
all I-QOL scores at week 24 were strongest for mental health (.45–.59), social functioning (.43–.54), and vitality (.36–.54) and were also moderate for I-QOL psychosocial impact score and role–physical, general health, and role–emotional (.37–.44). Correlations with the mental composite scale of the SF-36 were moderate to strong and statistically significant (.39–.57, \( P < .05 \)), whereas correlations for the physical composite scale were weaker and only statistically significant for psychosocial impact score. Correlations between physical functioning and I-QOL score, and between bodily pain and I-QOL score were not statistically significant (correlations \( r \leq .26 \)).

Significant correlations (\( P < .05 \)) were observed between I-QOL scores and various clinical efficacy measures during the course of the study (table 4). As expected, correlations were strongest for the change in daily number of incontinence episodes with values ranging from \(-.25\) to \(-.44\), and were negative because improvements in QOL (ie, increases in I-QOL score) were associated with decreases in incontinence episode frequency. Statistically significant correlations (\( P < .05 \)) were also observed at some time points for: the number of catheterizations, maximum cystometric capacity, and maximum detrusor pressure (see table 3). Correlations between I-QOL scores and other efficacy parameters, however, were generally weak (\( r < .30 \)). The poor correlation between reflex detrusor volume and I-QOL scores is likely to be due, at least in part, to 23 patients from the intention-to-treat population experiencing no reflex detrusor volume for at least 1 follow-up visit.

### Responsiveness

There were no ceiling effects for any of the domains or for total I-QOL score. Small floor effects were observed for the social embarrassment domain (8.9% of subjects had the lowest possible score) and the psychosocial impact domain (1.8% had the lowest possible score), both of which were less than the 10% criterion.

The responsiveness of I-QOL score to improvements in symptoms was assessed by comparing I-QOL scores at week 6 for subjects grouped by the change from baseline in number of daily urinary incontinence episodes (25% increase, 25% decrease or no change). For all 4 I-QOL scores, mean change from baseline in I-QOL score was greatest for the “decreased” group, and lowest for the “increased” group in avoidance and limiting behavior, social embarrassment, and total domains (fig 1). The difference between the increased group and the decreased group was statistically significant (\( P < .05 \)) for all domains except social embarrassment. The responsiveness statistic at week 6 was strong for both total score (1.4) and for individual domains (range, 0.9–2.0). The standardized response mean for the total score at week 6 was 1.0 for total score and ranged from 0.8 to 1.0 for the domains.

#### MID in I-QOL Score From Baseline

The change in I-QOL score corresponding to an MID was approximately 4 points when defined as that corresponding to a small effect size (ie, 0.2 SD at baseline), approximately 11 points when defined as corresponding to a medium effect size (ie, 0.5 SD at baseline), and ranged from 8 to 11 when defined as the SE of measurement.

### Table 4: Correlations Between I-QOL Domain Scores and Clinical Measures

<table>
<thead>
<tr>
<th>No. of Involuntary Losses of Urine</th>
<th>No. of Catheterizations</th>
<th>MCC</th>
<th>RDV</th>
<th>MDP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Week 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avoidance and limiting behavior</td>
<td>(-.40^*)</td>
<td>.19</td>
<td>.17</td>
<td>.03</td>
</tr>
<tr>
<td>Psychosocial impacts</td>
<td>(-.33^*)</td>
<td>.31</td>
<td>.16</td>
<td>(-.03)</td>
</tr>
<tr>
<td>Social embarrassment</td>
<td>(-.25)</td>
<td>.22</td>
<td>.17</td>
<td>.03</td>
</tr>
<tr>
<td>Total score</td>
<td>(-.36^*)</td>
<td>.27</td>
<td>.18</td>
<td>.01</td>
</tr>
<tr>
<td><strong>Week 6</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avoidance and limiting behavior</td>
<td>(-.44^*)</td>
<td>.10</td>
<td>.24</td>
<td>(-.01)</td>
</tr>
<tr>
<td>Psychosocial impacts</td>
<td>(-.38^*)</td>
<td>.15</td>
<td>.17</td>
<td>(-.09)</td>
</tr>
<tr>
<td>Social embarrassment</td>
<td>(-.34^*)</td>
<td>.25</td>
<td>.37*</td>
<td>(-.16)</td>
</tr>
<tr>
<td>Total score</td>
<td>(-.41^*)</td>
<td>.17</td>
<td>.26</td>
<td>.00</td>
</tr>
<tr>
<td><strong>Week 24</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avoidance and limiting behavior</td>
<td>(-.44^*)</td>
<td>.24</td>
<td>.22</td>
<td>.10</td>
</tr>
<tr>
<td>Psychosocial impacts</td>
<td>(-.30)</td>
<td>.21</td>
<td>.19</td>
<td>.08</td>
</tr>
<tr>
<td>Social embarrassment</td>
<td>(-.28^*)</td>
<td>.22</td>
<td>.36*</td>
<td>.27</td>
</tr>
<tr>
<td>Total score</td>
<td>(-.36^*)</td>
<td>.24</td>
<td>.25</td>
<td>.14</td>
</tr>
</tbody>
</table>

*Significant Pearson correlation at \( P < .05 \).

**Abbreviations:** MCC, maximum cystometric capacity; MDP, maximum detrusor pressure; RDV, reflex detrusor volume.
DISCUSSION

Limited data are available showing the validation of incontinence-specific QOL questionnaires in the neurogenic bladder population. Indeed, although the assessment of clinical efficacy is undoubtedly essential in the investigation of new treatments, it is increasingly being recognized that effects on QOL may be equally important, especially for conditions such as incontinence, which affect morbidity more than mortality.

This study provides the first psychometric assessment of a widely used and highly recommended questionnaire that assesses the impact of urinary incontinence on QOL, the I-QOL, in neurogenic detrusor overactivity. The questionnaire domains and total score showed good internal consistency and reliability. Analysis of correlations between I-QOL scores and clinical measures as well as with SF-36 domain scores was used to determine validity. Strong and statistically significant correlations between I-QOL and SF-36 scores were observed for all SF-36 scores with the exception of physical functioning and bodily pain. This corresponds to what would be expected because physical functioning and bodily pain in this patient population are likely to be dominated by the effects of underlying disease rather than urinary incontinence.

Correlations between I-QOL and SF-36 vitality and general health domains were found to be not significant at screening, but significant at study end. This is to be expected as general health and vitality are more “global” problems and, as such, affected to a far greater extent by the primary disease than urinary incontinence. Differences at screening may, therefore, reflect the differences in sensitivity between the tools for the effects of urinary incontinence symptoms. Because urinary incontinence symptoms improved throughout the course of the study, differences in sensitivity between the questionnaires would be expected to diminish. In addition, because of the entry criteria, the I-QOL response range at screening is narrow, whereas at post-treatment, there is a wider range of scores as the population becomes more diverse as a consequence on differences in response to therapy.

The validity of the I-QOL scores was also confirmed by the presence of reasonably strong and statistically significant correlations between I-QOL scores and the frequency of urinary incontinence episodes during the study. Correlations with urodynamic parameters were generally weak, which is in keeping with previously reported results, and confirms the importance of inclusion of patient reported outcome measures, such as I-QOL, in studies investigating the effectiveness of an intervention in the neurogenic bladder population.

A further analysis showed that the I-QOL is responsive to changes in QOL associated with changes in symptom severity (ie, number of daily urinary incontinence episodes). By week 6, approximately two thirds of subjects (in the total study population) had achieved a decrease in incontinence episodes of at least 25%. I-QOL scores were 10 to 15 points higher in these subjects compared with those who had experienced no change in number of incontinence episodes. This compares with a difference of 5 points reported for a study in non-neurogenic subjects. The difference between those with no change compared with those with increased symptoms was approximately 5 points in this study, comparable with that reported for non-neurogenic subjects. In addition, the responsiveness statistics and standardized response mean compared favorably with those reported in other validation studies.

The standard criteria for a definite scaling success are measured by the percentage of item-to-scale correlations that are at least 2 SEs greater with their own scale than with other scales. Scaling success results were suboptimal because some items were unable to discriminate between their hypothesized scale and other scales included in the measure. These results suggest overlap between the domains for some items of the I-QOL, but they are not definitive because this was a relatively small sample (n=53) of patients relative to items within the questionnaire (n=22). Combining of subscales into the total score appeared to overcome the suboptimal item-to-scale correlations, and may, therefore, prove to be a more valuable measure of QOL in the neurogenic bladder population.

Calculation of the MID using 3 different methods found that I-QOL is sensitive to changes in QOL in neurogenic detrusor overactivity. This is consistent with previous analyses, which have generally shown a 2% to 13% difference in I-QOL scores to be clinically relevant and are in agreement with other accepted measures of patient response (eg, number of urinary incontinence episodes).

The results of these analyses indicate that the I-QOL provides a valid, reliable and sensitive measure of the effect of treatment on incontinence-related QOL in patients with neurogenic disease, as has previously been shown for patients with non-neurogenic urinary incontinence. The 3 domains, in general, have acceptable properties, with the total score appearing to be more robust than the 3 domains. The robustness of the total score recommends its use as the major measurement domain in analyses of treatment effect.

Detailed analyses of the individual items within the I-QOL suggested that some modification of the scoring of the I-QOL for this population could further improve its performance in patients with neurogenic disease. Six of the items were found to be unchanged over the course of active treatment for 24 weeks. For some, the lack of change may relate to constraints of their primary condition, for example, “having to be careful about standing up” and may not be relevant for some patients with SCI who are wheelchair users. Additional research with specific subgroups of neurogenic patients may be appropriate to tailor the use of the questionnaire further to this population.

Further work on the arrangement of the individual items within the 3 domains (ie, avoidance and limiting behavior, psychosocial impact, social embarrassment) may also be of value to improve the scaling of the domains, because the scaling success of the 3 domains was not optimal (10%–12.5% using the criterion of 2 SDs). This reflected the fact that 6 of the items correlated equally or more strongly with a domain other than that to which they were assigned. The small sample size in this study may also have contributed to the suboptimal scaling success. Despite this, the results of this study clearly indicate that validity of the I-QOL was acceptable, but that I-QOL total score may be a more robust measure for evaluating QOL changes in this population.

Study Limitations

There are a number of limitations regarding this study that should be considered when interpreting the results. First, the study population was small and involved 2 distinct types of patients, 53 with SCI and 6 with MS. It is unknown to what extent the clinical characteristics of urinary incontinence and the aspects of the condition that are troublesome to patients differ for MS and SCI patients. The small sample size may also have limited our ability to evaluate the impact of different demographic variables, such as sex or age, on the performance of the I-QOL scale. Larger studies are, therefore, required to confirm the results of this study and determine specific differences between patient groups.

Second, analysis of the results for individual items within the I-QOL indicated that some of the items may be less relevant to this patient population, and several of the items showed an
equal or stronger correlation with a domain other than that to which it was assigned. These observations suggest that the sensitivity of the questionnaire could possibly be improved by modification of some of the items or scoring of the items within the domains. In addition to primary condition, however, the population also varied in degree of mobility. This would obviously affect the importance the individual places on some of the I-QOL questions, such as “I have to be careful standing up from sitting.” Further studies are required to determine the relevance of individual domains to populations on varying mobility and disability.

CONCLUSIONS

The results of this study indicate that the I-QOL is a valid, responsive, and reliable measure for evaluating the impact of treatment on QOL associated with urinary incontinence in patients with neurogenic bladder disease. Total score appears to be a more robust measure of change in QOL than individual subdomains.

Acknowledgments:

Allergan employees were involved (1) in the design and conduct of the study and (2) in the collection, management, analysis, and interpretation of the data. The authors had sole control over the preparation, review, and approval of this manuscript.

References

9. el Din KE, Kiemeney LA, de Wildt MJ, Rosier PF, Debruyne FM, de la Rosette JJ. The correlation between bladder outlet obstruction and lower urinary tract symptoms as measured by the international prostate symptom score. J Urol 1996;156:1020-5.
Effect of Orthoses on Postural Stability in Asymptomatic Subjects With Rearfoot Malalignment During a 6-Week Acclimation Period

Carl G. Mattacola, PhD, ATC, Maureen K. Dwyer, MS, ATC, Amy K. Miller, MS, ATC, Tim L. Uhl, PhD, ATC, Jean L. McCrory, PhD, Terry R. Malone, EdD, PT, ATC


Objective: To determine the effect of custom-fitted orthoses on postural sway over a 6-week acclimation period.

Design: Repeated-measures analysis of variance on postural sway measures with factors being group (control, malaligned), time (initial, 2wk, 4wk, 6wk postintervention), and condition (with orthoses, without orthoses). For single-limb stance, side (right, left) was analyzed to determine bilateral differences.

Setting: Biodynamics laboratory.

Participants: Twenty-one subjects, 11 asymptomatic with rearfoot malalignment and 10 asymptomatic with normal rearfoot alignment.

Interventions: Orthoses were prescribed and worn for 6 weeks. Balance testing was performed on 4 different dates with each subject tested in both orthotic conditions. Postural control was measured with three 10-second eyes-closed trials for single-limb stance, one 20-second eyes-closed bilateral stance with the platform moving, and one 20-second eyes-open bilateral stance with the platform and surroundings moving.

Main Outcome Measures: Sway velocity (in deg/s) for single-limb stance and equilibrium score for bilateral stance.

Results: Postural sway measures were significantly decreased during single-limb testing with orthoses versus without orthoses, regardless of group. The orthotic intervention significantly improved bilateral stance equilibrium score in the malaligned group at weeks 2, 4, and 6 when compared with measures at the initial week. Equilibrium score of the malaligned group with orthoses at initial week was significantly lower (worse) than the control group with orthoses at initial week; however, these results were not repeated during measurements taken at weeks 2, 4, or 6.

Conclusions: The application of orthoses decreased sway velocity for single-limb stance, improving postural stability regardless of group when visual feedback was removed. During bilateral stance, postural stability was initially worse for the malaligned group with and without orthoses when compared with the control group; however, improvements were seen by week 2 and continued throughout the remainder of testing. Clinically, the application of orthoses appears to improve postural control in people with rearfoot malalignment, particularly when vision is removed.

Key Words: Adaptation, physiological; Balance; Equilibrium; Musculoskeletal system; Orthotic devices; Posture; Rehabilitation.

© 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

IT HAS BEEN SUGGESTED that “structural and positional imbalances of the foot” may contribute to overuse injuries throughout the kinetic chain.1 Orthoses are designed to assist a malaligned foot in adapting to the external environment and, in theory, reduce the frequency of injury. Thus, orthoses are built to place the foot into a position of stability.1 This position is often termed the neutral position and is defined as the position of the foot in which the angle formed between the bisection of the distal third of the lower leg and the bisection of the calcaneus is 0°.1,2 For the majority of the population, however, subtalar joint neutral is generally a position of 2° to 3° of inversion, or rearfoot varus.3 Walker and Fan4 defined subtalar joint neutral position as a navicular angle between 130° and 150°, a normal medial longitudinal arch, and a calcaneal position of perpendicular to the ground.

The exact mechanism of effectiveness for an orthosis is unclear. Freeman et al5 have proposed that there are proprioceptive deficits after ankle injury due to pain, weakness, or limitation of movement. It has been suggested that orthoses provide support to the injured5,6 and fatigued7 foot. In turn, this may decrease the stress on the injured ligaments and allow the joint mechanoreceptors to provide appropriate feedback to the balance system.6,8

The efficacy of orthotic intervention has been limited to the immediate feedback from the initial application of orthoses to the foot. Nigg et al9 have suggested that orthoses act as a filter to the forces acting on the sole of the foot. These “filtered” forces are then transmitted to the central nervous system (CNS) to initiate an appropriate dynamic response. Nigg’s theory is based on research that has tested the initial reaction to orthoses, however, so it is possible that the “filter” is only an initial response to tactile stimulation of the foot. Without long-term study, it is difficult to clearly state the mechanism by which orthoses work, especially when much of that work has been done with nonpathologic subjects.10,11

Although traditional research has focused on the use of orthoses to alter the gait cycle, recent literature has begun to focus on the use of orthoses as an aid for proprioception and postural stability. Numerous researchers have examined the effects of orthoses on people who have suffered an acute ankle sprain. Orteza et al12 and Guskiewicz and Perrin13 reported, respectively, positive effects on both pain and postural sway.
measurements after orthotic intervention. Conversely, Hertel et al.\textsuperscript{12} reported orthotic intervention, regardless of type, to have no effect on improving postural sway measures after lateral ankle sprains. Hertel et al.\textsuperscript{10} also studied the effects of orthoses on postural control in healthy subjects. Their results showed an improvement in sway velocity measures in the frontal plane. Alternatively, Percy and Menz\textsuperscript{13} reported no significant effect of orthoses on postural stability in a group of healthy professional soccer players. Thus, although the evidence is conflicting, orthoses have been shown to have proprioceptive benefits after inversion ankle sprains\textsuperscript{8,12} and reducing mediolateral (ML) sway in nonimpaired subjects.\textsuperscript{10,13}

Few studies have actually evaluated the use of orthoses over a sustained time period. Stude and Brink\textsuperscript{11} reported an improvement in balance and a reduction in fatigue in a group of experienced golfers over a 6-week period. The subjects were tested only without orthoses initially, however, and then only with orthoses after the intervention, thereby raising some doubt as to the exact efficacy of the orthoses themselves. In another study following orthotic intervention over a course of 4 weeks, Rome and Brown\textsuperscript{14} reported significant improvement in ML sway measures in a group of subjects with rearfoot malalignment who wore orthoses compared with a group of subjects with rearfoot malalignment who did not wear orthoses. The device used to measure postural sway values did so only in a bilateral static condition, however.

The aforementioned studies have shown an immediate effect for the use of orthoses in both an injured\textsuperscript{8,12} and nonimpaired population.\textsuperscript{10,13} Although studies that examine the effects of orthoses over a sustained period do exist, they either incorporate subjects without rearfoot malalignment\textsuperscript{11} or use a measurement tool that only assesses static balance parameters.\textsuperscript{14} A study of effectiveness of orthoses over a sustained time period using subjects with structural abnormalities warranting orthotic intervention is needed. Therefore, the purpose of our study was to determine the effect of orthoses on postural sway over a 6-week acclimation period on subjects with a rearfoot malalignment. We hypothesize that the application of orthoses will significantly reduce postural sway measures over the 6-week time frame.

**METHODS**

**Participants**

Twenty-one subjects (10 men, 11 women) volunteered for this study. We recruited volunteers through a sample of convenience. Subjects were divided into 2 groups: malaligned (n=11) and control (n=10). Criteria for inclusion in the malaligned group included: bilateral calcaneal valgus or calcaneal varus of 5° or greater as measured with a standard goniometer, and an observational amount of standing pronation or supination.\textsuperscript{15} Subjects in the malaligned group were asymptomatic at the time of the study. Subjects not meeting these criteria were placed into the control group. Exclusion criteria for all subjects included: previous use of orthoses, history of severe ankle injury in the last 6 weeks, stress fracture within 1 year, use of any medications that affect the CNS, or the ability to balance, vestibular or neurologic disorders, history of severe head injury within the last 6 months, or history of unexplained falls. The institutional review board at the University of Kentucky, Lexington, KY, approved testing procedures and all subjects signed an informed consent prior to participation.

**Materials and Instrumentation**

Materials used to form foot impressions were provided by Foot Management Inc.\textsuperscript{4} After receipt of the foot impressions, Foot Management fabricated a pair of custom fitted semi-rigid orthoses (Ortho Arc sport model). A standard goniometer\textsuperscript{5} was used to measure calcaneal valgus and varus to the nearest degree.

We used the long forceplate of the NeuroCom Smart Balance Master\textsuperscript{6} to measure postural sway during single-limb stance. Unilateral postural sway was measured as sway velocity. Sway velocity is a ratio of distance to time (d/t). The center of gravity (COG) sway velocity is the ratio of the distance traveled by the COG (in degrees) to the time of the trial (in seconds).\textsuperscript{16}

We used the Smart Balance Master to measure postural stability. The Smart Balance Master is a balance-testing device that is composed of a dual forceplate, a visual surround around the forceplate, and a safety harness. It is integrated with a personal computer. The Sensory Organization Test (SOT) of the Smart Balance Master was used for evaluation. The main outcome measure for the SOT is equilibrium score. Equilibrium score is a nondimensional percentage that compares the patient’s peak amplitude of anteroposterior (AP) sway to the theoretical AP limits of stability. The subject’s theoretical limit of stability is the maximum forward and backward COG sway angles (θ) that can be achieved by a normal subject of similar height and weight, and is measured as the angular distance from vertical. It is based on the following formula:

\[
\text{Equilibrium score} = \left( \frac{12.5 - (\theta_{\text{max}} - \theta_{\text{min}})}{12.5} \right) \times 100
\]

Where θ\textsubscript{max} is the maximum AP sway angle range. Subjects that exhibit little sway will achieve an equilibrium score close to 100. Subjects that approach a fall will exhibit a score close to zero.

**Procedures**

During a familiarization session, we explained the testing protocol and the data collection procedures to the subjects and each subject was asked to sign an informed consent document. In this session, subjects completed a health and injury questionnaire that we developed. The questionnaire was used to determine the subject’s medical history and to determine in which group the subjects would be placed. If the subject met inclusion criteria, calcaneal alignment was then measured according to procedures outlined by Hunter et al.\textsuperscript{1}

We measured calcaneal valgus by placing the subject in a prone position on a plinth. The leg to be measured was positioned with knee extended and ankle approximately 15.2cm (6in) over the plinth. The opposite leg was placed in a position of flexion, abduction, and external rotation. A line was drawn along the midpoint of the posterior aspect of the calcaneus bisecting it into equal right and left portions. Similarly, a second line was drawn down the distal one third of the lower leg. The examiner then determined subtalar neutral by using the thumb and index finger to palpate the medial and lateral aspects of the talus.\textsuperscript{1} Next, the examiner inverted and everted the calcaneus until there was no pressure of the talus felt on either the thumb or index finger. The angle formed by the intersection of these 2 lines was measured using a standard goniometer.\textsuperscript{1}

Calcaneal alignment was also measured in a standing position. The subject stood in bilateral stance on a wooden box (38×46×21cm). Using the lines of bisection described above, the examiner used thumb and index finger to palpate the medial and lateral aspects of the talus. With the weight of the body on 1 limb, the subject was then asked to pronate and supinate the
foot while the examiner determined subtalar neutral. The angle formed by the intersection of the bisected lines was defined as standing neutral rearfoot motion (see fig 1A). These steps were then repeated on the opposite side. Next, the examiner measured resting standing foot posture (see fig 1B), as described by McPoil et al.17-19 Resting standing foot posture was defined as “the position of the subtalar and talocrural joints when the subject was standing relaxed with knees fully extended, the arms at the side, feet 6 inches apart, and a comfortable amount of toeing-out.”20(p368) If this angle was 5° or greater (varus or valgus) for both the left and right limbs for all 3 measurements, the subject was included in the malaligned group. If these measurements were less than 5° for either limb, they were placed in the control group. All rearfoot measurements were performed by the principal investigator.

Prior to testing, intratester reliability was established. Seven subjects were assessed in a single occasion. Intratester reliability measurements were separated by 1 hour. Rear foot motion was assessed by a single examiner, all marks were removed and after 1 hour, rear foot motion was reassessed. Pearson product-moment correlations were determined for prone neutral rearfoot motion and standing neutral rearfoot motion. Reliability for the left prone measurements was determined to be $r = .79$. Reliability for the right prone measurements was determined to be $r = .93$. Standing neutral subtalar motion for the left foot was calculated at $r = .90$; and standing subtalar joint motion for the right foot was calculated at $r = .94$. Resting standing foot posture for the left foot was calculated at $r = .91$; and resting standing foot posture for the right foot was calculated at $r = .86$.

Orthotic Construction

After determination of inclusion and exclusion criteria, we fitted subjects for orthoses. Subjects were seated with the ankles, hips, and knees each in a 90° position. Subtalar joint neutral position was found and maintained while the examiner guided the foot into the foam impression and pressed the heel approximately 5cm into the foam impression. The impressions were then sent to Foot Management for semi-rigid orthoses to be posted according to manufacturer recommendations (fig 2). The manufacturer was also instructed to correct for any malalignments or abnormalities revealed by the impression.

Testing

Subjects reported for testing on 4 separate occasions after orthosis construction (initial test and at 2-, 4-, and 6-wk intervals after receipt of the orthoses). The long forceplate of the Smart Balance Master was used to test the 2 single-limb stance conditions (eyes closed on the left and right foot) (fig 3A). Two test conditions of the SOT of the Smart Balance Master21 were used to test the 2 double-limb stance conditions (fig 3B). The SOT protocol assesses abnormalities in a subject’s somatosensory, visual, and vestibular systems, which help contribute to postural control.21 During the test, information from the various systems is altered through sway referencing. Sway referencing involves the tilting of the support surface and/or visual surround to directly follow the subject’s COG sway, such that the orientation of the surface remains constant in relation to the COG angle.22 The 2 conditions of the SOT that were used in our study were condition 5 (double-limb stance with eyes closed and the force platform moving in a sway-referenced position relative to the subject) and condition 6 (double-limb stance with eyes open and the visual surround and force platform moving with the subject’s AP sway). We chose to test conditions 5 and 6 because these conditions were found to exhibit the highest reliability between all testing sessions in our lab, and they challenged the 3 sensory systems to the greatest extent.

Fig 1. (A) Standing neutral rearfoot motion measurement position. (B) Resting standing foot posture measurement position.
Each subject participated in an introductory and practice session on both the long forceplate and the Smart Balance Master to minimize any type of learning effect. After this session, initial measurements of postural stability were obtained. Each subject completed each of the 4 different conditions on the appropriate instrument with and without orthoses. Subjects were instructed to stand in a comfortable stance with their arms down at their sides. The single-limb stance conditions were performed 2 times for 10 seconds each. Each bilateral stance condition was performed 2 times for 20 seconds each. There was a 10-second rest between each trial. All subjects performed the postural stability tests with and without orthoses in a counterbalanced fashion to prevent order biasing.

**Design and Data Analysis**

The research design consisted of a 1 between (group) and 2 within (time, condition) mixed-design analysis of variance (ANOVA). The dependent variables were group (control vs malaligned), time (initial evaluation, 2-wk, 4-wk, and 6-wk interval), and condition (orthosis, no orthosis). Separate ANOVAs were used for each limb. The dependent variable for bilateral stance was calculated as equilibrium score, and the dependent variable for single-limb stance was calculated as sway velocity (in degrees per second). An α level of P equal to or less than .05 was determined to be statistically significant. Any significant findings were further analyzed using Tukey post hoc testing with a Bonferroni adjustment. A statistical package for Windows was used to perform all statistical analyses.

**RESULTS**

There were no significant differences between the 2 groups for any of the subject demographic characteristics (P>.05). A total of 11 subjects met the criteria for inclusion into the malaligned group (age, 24.5±6.5y; mass, 75.5±4.6kg; height, 177.80±37.7cm), with 9 subjects classified as having rearfoot varus and 2 as having rearfoot valgus. Mean rearfoot measures in the prone position were 6.48°±1.2° (left foot) and 6.81°±2.1° (right foot) for the varus subjects and 5.25°±0.6° (left foot) and 6.00°±0.5° (right foot) for the valgus subjects. Mean rearfoot measures in the standing neutral position were 5.93°±1.1° (left foot) and 6.33°±1.8° (right foot) for the varus subjects and 5.33°±0.5° (left foot) and 5.17°±0.7° (right foot) for the valgus subjects. Mean rearfoot measures for the resting standing position were 8.17°±2.6° (left foot) and 9.04°±1.8° (right foot) for the varus subjects and 5.67°±3.3° (left foot) and 5.5°±0.2° (right foot) for the valgus subjects. Ten subjects did not exceed 5° for the rearfoot measures in any measurement and were placed in the control group (age, 22.3±2.2y; mass, 77.13±3.5kg; height, 169.9±25.9cm).

The means and standard deviations (SDs) for bilateral limb postural sway condition 6 (eyes open with forceplate and surround moving in a sway-referenced position) are presented in table 1.

**Single-Limb Postural Stability**

Analysis of the single-limb stance values revealed a significant main effect for orthoses regardless of group (P<.05). Left leg sway velocity was significantly lower for the orthotic condition versus the nonorthotic condition (1.79°±0.74°/s vs 1.92°±0.74°/s, P=.019). Right leg sway velocity was also found to be significantly lower for the orthotic condition than for the nonorthotic condition (1.74°±0.74°/s vs 1.86°±0.73°/s, P=.043). There were no significant differences between weeks for the single-limb conditions.

**Bilateral Postural Stability**

There were no significant differences for condition 6. There was a significant 3-way interaction (week by condition by group) for bilateral stance condition 5 (fig 4). Post hoc analysis revealed that equilibrium score for the malaligned group with orthoses at the initial week (mean, 60.27%±13.35%) was significantly lower than equilibrium score for the malaligned group with orthoses during week 2 (mean, 70.73%±10.54%), week 4 (mean, 71.23%±6.46%), and week 6 (mean, 70%±7.63%) (P<.01). Therefore, balance measures in the malaligned group improved over time with orthoses. At the initial week, equilibrium scores for the malaligned group with orthoses were significantly lower than the control group with orthoses (60.27%±13.35% vs 72.15%±9.73%, P<.01). Finally, when comparing the control group without orthoses, initial week measures were significantly lower than week 4 measures (69.15%±8.13% vs 76.81%±7.37%, P<.01).

**DISCUSSION**

The results of our study show a positive effect of orthotic intervention on certain measures of postural stability for individuals with rearfoot malalignment over time.

**Single-Limb Postural Sway**

During the initial test, we found that postural sway measures were reduced for single-limb stance with orthoses relative to without orthoses for both the right and left conditions. These results were similar to the results found by Ochsendorf et al, who reported that prior to a lower-extremity fatigue protocol, single-leg stance postural sway measurements for the orthotic condition were significantly less than the nonorthotic condition. Similarly, Hertel et al determined that rigid full foot orthoses posted at the medial side of the rearfoot provided a reduction in single-limb center of pressure (COP) length and velocity in healthy subjects. Hertel also reported that a Sprained Ankle Orthotic increased the movement of postural COP length and velocity, however. The Sprained Ankle Orthotic is a noncustom orthosis that is rigid and is equipped with a laterally posted heel wedge. Therefore, although these results differed from ours, the difference may have been because of the different materials used to construct the orthoses in the Hertel et al study and ours as well as the change in position of the foot due to the lateral wedge of the orthosis used by Hertel.

Although there are results to support the use of orthoses for improving single-limb postural sway, there are several studies...
that provide conflicting results.\textsuperscript{6,8} Guskiewicz and Perrin\textsuperscript{6} found no effects on postural sway measures for 12 unimpaired subjects in an orthotic and nonorthotic condition using the Chattecx Balance System. Likewise, Orteza et al\textsuperscript{8} showed that molded orthoses did not improve balance measures for 10 subjects with no history of ankle sprains. Hertel et al\textsuperscript{12} reported initial increases in single-limb postural sway measures after acute ankle sprain; however, an orthotic intervention did not

Table 1: Bilateral Postural Sway Values for Condition 6 (eyes open with the platform and surround moving)

<table>
<thead>
<tr>
<th>Week</th>
<th>Malaligned (orthotic)</th>
<th>Control (orthotic)</th>
<th>Malaligned (no orthotic)</th>
<th>Control (no orthotic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>67.77 ± 15.89</td>
<td>73.30 ± 11.72</td>
<td>67.00 ± 19.16</td>
<td>74.90 ± 10.20</td>
</tr>
<tr>
<td>2</td>
<td>69.14 ± 17.41</td>
<td>75.30 ± 11.37</td>
<td>71.36 ± 15.71</td>
<td>73.55 ± 13.56</td>
</tr>
<tr>
<td>4</td>
<td>73.95 ± 14.59</td>
<td>75.15 ± 13.85</td>
<td>72.18 ± 12.39</td>
<td>75.11 ± 16.88</td>
</tr>
<tr>
<td>6</td>
<td>75.86 ± 11.70</td>
<td>76.55 ± 16.56</td>
<td>76.41 ± 10.27</td>
<td>75.00 ± 12.55</td>
</tr>
</tbody>
</table>

NOTE. Values are mean ± SD and expressed as equilibrium score (%).
improve these measures. Conversely, our results suggest that balance was improved for both the unimpaired and malaligned groups.

Differences in our results may be attributed to methodology in testing, subject size, or subject inclusion criteria. One rationale for orthosis use is that it increases somatosensory afferent input. Therefore, we chose to test subjects with eyes closed during single-limb stance to minimize visual feedback so that subjects relied more on somatosensory feedback. Guskiewicz and Perrin, Hertel, and Orteza tested subjects with their eyes open. One possible difference in our results was the removal of visual feedback. Although athletes do not compete with their eyes closed, blind landings and changes in visual horizons during dynamic conditions result in theoretically more reliance on somatosensory feedback than visual feedback. Our study supports the rationale of Nigg et al that the orthosis increases somatosensory stimulus.

The Smart Balance Master allows for measurement of reaction forces with 4 transducers. Methodologically, Orteza tested subjects on a digital balance evaluator that only tests time out of balance. Because the Smart Balance Master may be a more sensitive measuring device, the difference in instrumentation sensitivity might explain variability of the results. Also, Orteza allowed for a recovery of 1 to 2 minutes between testing and subjects were tested with their eyes closed, whereas we allowed a resting period between trials of only 10 seconds. Therefore, subjects had less chance to adapt to the testing procedures.

Guskiewicz and Perrin and Orteza compared subjects with acute ankle sprains to noninjured subjects. Acute was defined as occurring within 21 days of testing and within 6 weeks of training. Therefore, the severity of injury and extent of dysfunction at the time of testing is unclear. Current literature suggests that injury to the ankle changes the joint mechanoreceptors and alters balance. Although their results showed that postural sway with orthotic intervention was reduced more in injured subjects than uninjured subjects, they also established that the injured group swayed more without orthoses.

**Bilateral Postural Sway**

For condition 5 at the initial week, the malaligned group with orthoses had lower equilibrium scores, indicating worse balance than the control group with and without orthoses. Therefore, the control group performed better initially with and without orthoses than the malaligned group with orthoses.

In 1999, Nurse and Nigg established that there was a relationship between tactile stimulation and vibration sensitivity of the human foot with plantar pressure distributions during gait. Thus, the higher the pressure at various areas of the foot (hallux, heel, lateral arch, first metatarsal head), the lower was the vibration threshold at these particular areas. Because of this, Nurse and Nigg suggested that the body is able to determine small biomechanical changes in the external environment. In comparison with the control group, the malaligned group had a greater biomechanical deficit. Similarly, McPoil et al determined that during the stance phase of gait, the COP migration was lower with orthoses for a population of women with forefoot varus. Therefore, the orthoses for the malaligned group were posted more than for the control group in order to place the foot into a neutral position. It is possible that, initially, it was harder for the malaligned group to adjust to the biomechanical change than for the control group; postural stability values improved from week 0 to week 2 for the malaligned group with orthoses. When bilateral postural stability scores between week 0 and week 2 were compared, there was no difference between the malaligned group without orthoses. Consequently, it is possible that by week 2, the malaligned group was able to adjust to the biomechanical change and the equilibrium scores increased, reflecting better balance. Interestingly, there was no difference between week 0 and week 2 when the control group was compared with and without orthoses. A possible theory for this lack of change was that the control group had already adjusted to the small biomechanical change during the initial test.

At week 4, the control group without orthoses was significantly better than week 2 and week 0 without orthoses. We are...
not certain why this occurred, and, because there were no significant differences between week 4 and week 6, it is possible that at week 4 subjects experienced a learning effect. It is also possible that this instance may be explained through some of the basic principles of neuromuscular education. First, all beings have potentials that are not fully developed.²⁷ Although the control group did not have a pathologic deficit in postural sway, it is possible that the orthoses provided a tactile stimulation to improve learning. Furthermore, motor learning is enhanced through the use of multisensory inputs. Tactile cues also provide direction, encouragement, and optimize learning opportunities.²⁷ Consequently, when the individuals were tested without orthoses in the shoe, some sort of altered sensitivity may have occurred to improve balance scores.

There were no significant differences from week 0 to week 6 for the control group. These results differ from previous studies on the effectiveness of orthoses over time.¹¹,¹⁴ Stude and Brink¹¹ found that orthoses improved balance for 12 unimpaired subjects over a 6-week time period; however, they only tested the individuals on 2 separate occasions: at the initial visit without orthoses and after 6 weeks of orthotic intervention with the orthoses. Consequently, one may question whether their results can be compared with ours, because evaluation at the 6-week intervention time frame is the only time that balance was assessed with orthoses. Rome and Brown¹⁴ reported improvements in ML sway measurements after orthotic intervention over a 4-week time period in a group of subjects with excessively pronated feet. Differences in results could be attributed to methodologic differences. Rome and Brown¹⁴ used a balance measurement tool that only measured static balance in a bipedal stance, whereas ours measured dynamic balance in a single-limb stance. It is difficult to make direct comparisons between studies when there are distinctly different measurement strategies.

We observed that subjects with rearfoot deformities tend to place pressure at a more focal point when standing. Conversely, individuals with a more neutral rearfoot alignment, however, place a more distributed pressure on the foot. One may theorize that it is possible that orthoses promote a more diffuse pressure under the foot and promote postural stability. According to our data, there was significant difference in the malaligned group between the orthotic and nonorthotic condition for the malaligned group during walking. Our inclusion criterion for the malaligned group of 5° or greater of rearfoot motion may be a relatively small number according to some of the literature.¹²-²⁰ Our choice of this definition of abnormal pronation of 5° or greater was based on the work of Tomaro and Burdett,⁻²⁻ however. Consequently, we might have seen different results with a less restrictive criterion for the malaligned group. Also, it was difficult to control the amount of time that subjects wore the orthoses each day. Although they were told to wear them a minimum of 6 to 8 hours a day, some subjects may have worn them the minimum amount of time, whereas other subjects may have worn them more. It might have been beneficial to have each subject record daily wearing using a calendar and report this usage to us at the end of the study to determine compliance.

**CONCLUSIONS**

The application of orthoses improved eyes-closed single-limb postural sway, and bilateral postural stability measures regardless of the rearfoot measure. The application of orthoses resulted in improved postural control after acclimation in a group of malaligned subjects. Further research focusing on the use of muscle activation patterns may better describe the neuromuscular role of orthoses. If the anterior musculature is enhanced through the use of multisensory inputs, it may theorize that it is possible that orthoses promote a more diffuse pressure under the foot and promote postural stability. However, place a more distributed pressure on the foot. One may theorize that it is possible that orthoses promote a more diffuse pressure under the foot and promote postural stability. According to our data, there was significant difference in the malaligned group between the orthotic and nonorthotic condition for the malaligned group during walking. Our inclusion criterion for the malaligned group of 5° or greater of rearfoot motion may be a relatively small number according to some of the literature.¹²-²⁰ Our choice of this definition of abnormal pronation of 5° or greater was based on the work of Tomaro and Burdett,⁻²⁻ however. Consequently, we might have seen different results with a less restrictive criterion for the malaligned group. Also, it was difficult to control the amount of time that subjects wore the orthoses each day. Although they were told to wear them a minimum of 6 to 8 hours a day, some subjects may have worn them the minimum amount of time, whereas other subjects may have worn them more. It might have been beneficial to have each subject record daily wearing using a calendar and report this usage to us at the end of the study to determine compliance.

**Acknowledgment:** We thank Foot Management Inc for providing the orthoses.

**References**


Suppliers
a. Foot Management Inc, 7201 Friendship Rd, Pittsville, MD 21850.
c. NeuroCom International Inc, 9570 SE Lawnfield Rd, Clackamas, OR 97015.
d. Version 13.0; SPSS Inc, 233 S Wacker Dr, 11th Fl, Chicago, IL 60606.
Fluid Collections in Amputee Stumps: A Common Phenomenon

Rajiv Singh, MRCP, John Hunter, FRCP, Alistair Philip, PhD


Objectives: To assess the incidence of fluid collections in postoperative amputee stumps and the impact on limb-fitting outcomes in patients with such collections.

Design: Cohort study.

Setting: Inpatient rehabilitation ward.

Participants: Successful patients with amputation examined with ultrasound over 1 year.

Interventions: Not applicable.

Main Outcome Measures: The presence of discrete fluid collections on admission and outcomes of successful limb fitting, length of inpatient stay, and presence of psychologic symptoms.

Results: In 105 consecutive admissions, we detected discrete fluid collections in 28 (27%) of stumps with a median volume of 38.5mL (range, 16–216mL). All collections diminished and disappeared by discharge with 81% undetectable within 30 days since surgery. A transfemoral amputee was more likely to develop a collection than a transtibial amputee (P<.01). Patients with collections took 9.5 days longer to achieve limb fitting (P=.04) and had a 10-day longer inpatient stay (P=.02). However, the overall success of limb fitting was similar as was the incidence of psychologic distress.

Conclusions: Discrete fluid collections are common in postoperative amputation stumps but regress by discharge. Although limb fitting may be delayed, the ultimate success of limb fitting is not reduced and patients can be reassured.

Key Words: Amputation stumps; Rehabilitation; Treatment outcome; Ultrasonography; doppler.

© 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

Ultrasonography is an established modality in the investigation and detection of fluid collections (eg, liver abscess or pleural effusion). It has also been used in amputees to assess socket fit, locate neuromas, and detect deep vein thrombosis.1-4 We had noted the presence of fluid collections in amputee stumps in which a failure to heal had led to ultrasound investigation and the finding of a discrete fluid collection as opposed to generalized edema which is common postoperatively. A literature search found only 1 case report with the use of ultrasound to monitor an infected collection.5 However, we found no other reports or any studies that measured the incidence of such collections or their natural history. There was concern that such collections may impair healing and possibly affect the outcome of rehabilitation, and we therefore undertook a study to explore these issues.

METHODS

All amputees admitted to a regional rehabilitation unit over a 12-month period were prospectively scanned with a handheld 7.5-MHz ultrasound scanner6 by the lead author to detect the presence of fluid collections in their stumps. This calculates collection volume automatically when used in B-mode (brightness). All patients were scanned within 24 hours of admission and were admitted from vascular and orthopedic surgical wards at another hospital usually after they were stabilized postoperatively. Any collections found were rescanned every 2 to 3 days to monitor their progression and continued until no longer detected on scan. All patients were treated with compressive elastic dressings to reduce swelling after the wound was healed. Demographic data such as sex, age, and time since surgery was also recorded. All patients had an assessment of anxiety and depression symptoms recorded during their stay by means of a Hospital Anxiety and Depression Scale (HADS) by using a cutoff between 8 and 9, which has been found to offer the best diagnostic cutoff.6,7 It was thought that if patients were less likely to proceed to limb fitting then psychologic symptoms may be more common. Comorbid medical conditions were those considered likely to affect amputation outcomes significantly (eg, ischemic heart disease or rheumatoid arthritis but not psoriasis or migraine).

Patients found to have a collection were not treated differently in terms of therapy or multidisciplinary team discussion when decision to limb fit was taken. Only the lead author was aware if collections were present. Time was recorded as days since surgery rather than admission to the rehabilitation ward.

The study was approved by the regional ethics committee. Outcome measures used were achievement of limb fitting, length of stay, and presence of psychologic symptoms.

RESULTS

Over 12 months, 105 consecutive admissions were scanned. Demographic data are presented in table 1. All patients were transfemoral or transtibial amputees, and 83% had dysvascular disease with or without diabetes.

Out of all limbs scanned, 28 had significant fluid collections present (27%) with a median of 38.5 mL (range, 16–216 mL). All of these collections diminished on subsequent scanning with 81% undetectable within 10 days. However, the largest took almost 5 weeks to disappear (fig 1). No new collections developed during patient stay. Median time from surgery to date of scan was 19 days (range, 9–62d). Delay in time to scan was because of a delay in the transfer of patients from vascular units to the rehabilitation ward, but 92% were scanned within a month of surgery and there was no difference in time to scan between those with or without collection.

When comparisons were made between those with or without fluid collections (tables 2, 3), there were no significant
differences in terms of age, sex, comorbidities, or proportion achieving limb fitting. However, there was a difference in time taken to limb fit with patients with collections taking 9.5 days longer before they were judged to be ready based on multidisciplinary team discussion of each case. This was also reflected in the inpatient stay, which was 10 days longer as a result of the delay. Fluid collections were more common in patients with transfemoral amputation (P<.01). We also found no difference in terms of HADS scores suggesting no difference in psychological symptoms. No collections were aspirated.

DISCUSSION

We found no previous report that has acknowledged the presence of discrete fluid collections in amputee stumps and measured the incidence. We have shown that discrete collections are common but subsequently shrink. It is well recognized that postoperative patients will have generalized stump edema as a response to surgery or indeed any insult or injury. However, we are describing the presence of a discrete fluid collection quite distinct from general tissue edema.

A literature search found a case report of a patient who had an infected collection and ultrasound was used to monitor progress. The authors suggest use of ultrasound in infected collections, although, to our knowledge, none of our patients had an infected collection and 81% of collections diminished and disappeared within 10 days, although the largest collection took 5 weeks. Limb fitting was delayed but not the eventual outcome.

We cannot explain why some patients develop these collections and others do not. Because they are more common in transfemoral amputees, it may be that the surgery that leaves a larger stump may be more likely to allow a collection. There may be differences in surgical technique that create pockets or allow a collection to develop in certain tissue planes. However, we have not explored this aspect. All patients were routinely treated with compressive dressing while on the ward so there should not be a difference.

The key issue that we have addressed apart from the incidence of collections was whether selected outcome would be affected. In most cases, the collections simply disappeared and, but for scanning, we would never have known that they were there. Patients with or without collections did not differ in terms of HADS scores, medical comorbidities, or limb fitting, although there was delay in time to limb fit presumably based on the stump swelling. Patients with collections can therefore be reassured that while limb fitting may be delayed, it will nevertheless, be just as likely. The inpatient stay was also prolonged by 10 days, matching exactly the delay in time to limb fitting.

We divided the patients according to whether fluid collection was present or not and did not make any effort to distinguish whether the size of the collection was significant. The presence of fluid, no matter what size, is abnormal, and we have therefore considered all such patients in 1 group. Furthermore, we had very few patients with very large collections, and any statistical analysis would be weak.

Ultrasound was a reasonable investigative modality because it is safe, cheap, well tolerated, noninvasive and, with a portable machine, can be done at the bedside. Furthermore, the lower limb is well suited for ultrasound because soft tissues are not shielded by bone, and there are no air interfaces to lessen the view. Magnetic resonance imaging may have been more accurate but is costly and could not be used for frequent monitoring.

Study Limitations

The main weakness of our study is perhaps the use of initial limb fitting as an outcome. Some patients may be judged suitable to limb fit but on discharge quickly stop using their prosthesis. It has been suggested that limb use at 1 or 2 years after discharge may be a better measure of success. In addition, we have only scanned limbs on average 19 days after surgery as patients were in another hospital. It would be interesting to scan immediately after surgery to gauge the exact time course of collections as well as incidence. It may be that collections are even more common than we have found.

CONCLUSIONS

Fluid collection in amputee stumps is a common phenomenon. It may delay limb fitting and lengthen rehabilitation inpatient stay
but, based on our study, does not alter the likelihood of fitting or of developing adverse psychologic symptoms. In cases in which there is concern over the possibility of fluid collection, a handheld ultrasound probe offers a simple means of detecting and monitoring such collections.

References

Supplier
Evidence of the Psychometric Qualities of a Simplified Version of the Activities-specific Balance Confidence Scale for Community-Dwelling Seniors

Johanne Filiatrault, OT, MSc, Lise Gauvin, PhD, Michel Fournier, MA, Manon Parisien, OT, MSc, Yvonne Robitaille, PhD, Sophie Laforest, PhD, Hélène Corriveau, PT, PhD, Lucie Richard, PhD


Objective: To evaluate the validity, reliability, and item hierarchy of a modified version of the Activities-specific Balance Confidence (ABC) scale using an item-response theory framework and integrating modifications aimed at increasing user-friendliness and promoting better congruence of the scale with public health falls prevention strategies.

Design: Cross-sectional study.

Setting: Community-based.

Participants: Two hundred community-dwelling seniors involved in an effectiveness study of a falls prevention program. Participants were recruited by community-based organizations.

Interventions: Not applicable.

Main Outcome Measure: Balance confidence.

Results: The modified ABC scale (called ABC-Simplified [ABC-S] scale) has high internal consistency (reliability index, .86) and good convergent validity (statistically significant associations with perceived balance; performances on the one-leg stance, tandem stance, tandem walking, functional reach, and lateral reach [on the right side] tests; fear of falling; and occurrence of falls in the previous 12 mo). Analyses also showed differing degrees of difficulty across items, allowing for a determination of the scale’s item hierarchy.

Conclusions: The ABC-S scale is a valid and reliable measure for the assessment of balance confidence among community-dwelling seniors. The fact that this measure was validated with high-functioning seniors makes it particularly well-suited for identifying community-dwelling seniors who are beginning to lose confidence in their balance and who could benefit from community falls prevention programs.

Key Words: Accidental falls; Accident prevention; Aged; Balance; Psychometrics; Rehabilitation.

SYSTEMATIC REVIEWS OF the literature1-2 show that multifactorial programs and exercises are successful in reducing falls incidence and in improving balance among community-dwelling seniors. Although falls and balance are important targets for intervention, other falls-related factors, such as fear of falling, self-efficacy for avoiding falls, and balance confidence, are also important ends and require increased attention in evaluative research, because these factors can have an impact on seniors’ quality of life.3,4 Evidence also indicates that these variables are precursors of falls, because they often lead to activity avoidance and subsequent physical deconditioning.3,5,6 Despite the recent development of measures to assess these concepts, more psychometric research is required.7,8 We aim to contribute to this development by further ascertaining the validity and reliability of a modified version of one of the most frequently used measures for assessing falls-related psychologic factors, namely the Activities-specific Balance Confidence (ABC) scale developed by Powell and Myers.8

The ABC scale is among the few scales that were developed jointly by clinicians and seniors.7,8 Although often described as a measure of self-efficacy, the ABC scale is framed as a balance confidence scale.7 That is, it requires respondents to self-rate their degree of confidence in their balance associated with the performance of a series of daily living tasks. The ABC scale is a questionnaire that includes 16 items representing basic daily living tasks (eg, walking around the house, going up and down stairs) and more difficult tasks performed in the community (eg, walking in a crowded shopping center, using an escalator). Respondents provide ratings on a 0%-to-100% continuous scale based on the following cue question: “How confident are you that you will not lose your balance or become unsteady when you [list of items].” With its wide range of item difficulties, the ABC scale seems well suited for assessing 1 psychologic falls-related construct in populations presenting a diversity of levels of functioning, including high-functioning community-dwelling seniors.6,8

From the Institut National de Santé Publique du Québec, Montreal, QC, Canada (Filiatrault, Robitaille); Programme de Doctorat en Santé Publique, Université de Montréal, Montreal, QC, Canada (Filiatrault); Groupe de Recherche Interdisciplinaire en Santé, Montreal, QC, Canada (Filiatrault, Gauvin, Laforest, Richard); École de Réadaptation, Université de Montréal, Montreal, QC, Canada (Filiatrault); Département de Médecine Sociale et Préventive, Université de Montréal, Montreal, QC, Canada (Gauvin, Robitaille); Centre de Recherche Léa-Roback sur les Inégalités Sociales de Santé de Montréal, Montreal, QC, Canada (Gauvin, Richard); Direction de Santé Publique de Montréal, Montreal, QC, Canada (Fournier); Department of Epidemiology and Biostatistics, McGill University, Montreal, QC, Canada (Robitaille); Département de Kinésiologie, Université de Montréal, Montreal, QC, Canada (Laforest); Centre de Santé et de Services Sociaux Cavendish-Centre affilié Universitaire en Gériontologie Sociale, Montreal, QC, Canada (Laforest); Département de Médecine de Famille, Université de Sherbrooke, Sherbrooke, QC, Canada (Corriveau); Centre de Recherche sur le Vieillissement, Institut Universitaire de Gériatrie de Sherbrooke, Sherbrooke, QC, Canada (Corriveau); Faculté des Sciences Information, Université de Montréal, Montreal, QC, Canada (Richard); and Centre de Recherche de l’Institut Universitaire de Gériatrie de Montréal, Montreal, QC, Canada (Richard).

Presented in part to the Canadian Occupational Therapy Association, May 27, 2005, Vancouver, BC, Canada.

Supported by the Canadian Institutes of Health Research (grant no. MOP 53123); the Regional Health and Social Services Boards of Montreal; Laval, and Montérégie, QC, Canada; and the Fonds de la Recherche en Santé du Québec (doctoral scholarship) and the Ordre des Ergothérapeutes du Québec (doctoral scholarship).

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the author(s) or upon any organization with which the author(s) is/are associated.

Reprint requests to Johanne Filiatrault, OT, MSc, Faculté de Médecine, École de Réadaptation, Université de Montréal, CP 6128, succursale Centre-ville, Montreal, QC H3C 3J7, Canada, e-mail: johanne.filiatrault@umontreal.ca.

0003-9993/07/$8805-11088$32.00 © 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

doi:10.1016/j.apmr.2007.02.003
Several studies support the psychometric qualities of the ABC scale. An initial study of the psychometric properties of the ABC scale with a sample of 60 community-dwelling seniors showed that the ABC scale has high internal consistency (Cronbach α = .96) and high test-retest reliability (r = .92). Results also showed that the ABC scale has strong convergent validity reflected by a moderate to large correlation with the physical abilities subscale of the Physical Self-Efficacy Scale (PSES) (r = .63) and a strong divergent validity based on low and nonsignificant correlations with an unrelated subscale (self-presentation) of the PSES (r = .03) and another unrelated scale called the Positive and Negative Affect Schedule (r = .12). The ABC scale also showed power to discriminate between high- versus low-mobility seniors. Several other studies have confirmed the scale’s reliability and validity with seniors (see Jorstad et al). In addition, the ABC scale was adapted for seniors living in the United Kingdom to accommodate for differences between American English and British English. A psychometric study of the British version also showed high internal consistency and good test-retest reliability of the scale. Recently, a Dutch version of the scale (the ABC-NL) showed satisfactory psychometric qualities among community-dwelling seniors. A Chinese version of the scale (the ABC-C), tested for reliability among a sample of Chinese immigrants aged over 50 years living in Vancouver, BC, Canada, was also found to be reliable.11

The ABC scale’s psychometric properties have also been studied among people presenting with specific medical conditions. A study conducted among people with lower-limb amputation supported the scale’s internal consistency, test-retest reliability, and convergent validity.12 There is also evidence that the ABC scale is a valid measure for use with people presenting with a vestibular dysfunction and with older adults who have sustained a stroke.13,14 More recently, similar findings were obtained among people who had sustained a stroke and who responded to either the original version or a French Canadian version of the ABC scale. Despite its psychometric qualities, the ABC scale could benefit from some improvements in terms of user-friendliness. Seniors have shown problems interpreting the cue question and using the response format of the scale. The ABC scale’s cue question is framed in an avoidance rather than action perspective (ie, losing balance as opposed to maintaining balance), which appears somewhat incongruent with the positive focus of balance confidence. Thus, 1 modification that could improve the scale consists of modifying the cue question from “How confident are you that you will not lose your balance or become unsteady when you do the following activities?” to “Up to what point are you confident that you will maintain your balance when you do the following activities?”19

In addition, the scale’s response format, which requires participants to rate their balance confidence on a continuous scale ranging from 0% to 100%, is sometimes problematic, because seniors show difficulty in grasping the full range of the scale. Others have reported similar problems with a 0%-to-100% response format, especially among seniors with limited educational background. It must be noted that in a recent publication, Myers replaced the 0%-to-100% continuous response scale with an 11-point response scale that includes 10% anchor increments (0%, 10%, . . . , 100%). Although this new response format is simpler than the original response format, the scale might remain problematic for selected subgroups of seniors who have reported difficulties with a 10-point scale. Therefore, another modification that could improve the user-friendliness of the ABC scale consists of modifying the 0%-to-100% response format to a 4-category response format with descriptive anchors (ie, 0, not at all confident; 1, slightly confident; 2, moderately confident; 3, very confident).

Furthermore, 1 item of the ABC scale (ie, walk on icy sidewalks) appears problematic from a public health perspective. Indeed, walking outside when sidewalks and roads are icy is a particularly hazardous task for seniors. This is supported by empirical data suggesting that slipping on ice is among the main factors contributing to the excess seasonal risk of hip, arm, and other fractures among seniors. Many seniors intuitively tend to avoid walking outside when weather conditions are bad. Therefore, questioning seniors about this hazardous task is somewhat irrelevant and conveys a message that is rather incongruent with public health falls prevention strategies. Indeed, avoiding walking on icy sidewalks is a protective behavior that should be reinforced by public health interventions directed at community-dwelling seniors. Therefore, removing this item from the scale is considered an improvement from a public health standpoint. In sum, the resulting version of the scale (designated as the ABC-Simplified [ABC-S]) contains 15 items rather than 16 and has a more user-friendly cue question and response format.

Finally, previous psychometric studies of the scale were based on classical test theory methods that are ill-suited for scales using categoric response formats. Given the proposed amendments to the response format of the ABC scale, item-response theory (IRT) methods were used to study the scale’s properties. More specifically, a polytomous IRT model known as the Samejima graded response model is appropriate in situations in which the objective is to measure a latent construct through responses to items that are scored on an ordered-category rating scale. Such a model is based on the assumption that the response to an item is not only a function of the psychologic characteristic of the person (or latent trait) but is also a function of the level of difficulty of the item.

Thus, the purpose of this study was to examine the psychometric properties of a simplified version of the ABC scale (the ABC-S scale) using an IRT framework. More specifically, we tested the scale’s internal consistency, item hierarchy, and convergent validity with perceived balance, balance performance, fear of falling, and occurrence of falls in the previous 12 months using a polytomous IRT model.

METHODS

Design and Participants

The initial sample included 200 community-dwelling seniors living in the Montreal area who were recruited by community-based organizations to participate in an effectiveness study of a falls prevention program (called Stand Up!) on the balance of participants. For inclusion in the intervention study, participants had to meet the following criteria: (1) be aged 60 years or over, (2) be community dwellers, (3) be exempt from disabling conditions (eg, Parkinson’s disease, multiple sclerosis) and have the required capacities to get involved in a group exercise program as assessed by a preselection grid developed for the program, (4) be exempt from cognitive deficits, and (5) be able to speak either English or French. Cognitive integrity was assessed by interviewers responsible for administering the
modified ABC scale. Interviewers were instructed to use an additional questionnaire, the Short Portable Mental Status Questionnaire, whenever they suspected cognitive deficits in participants. In the end, none of the participants required further assessment. In addition, recruitment was publicized as a search for people with 1 or several of the following characteristics: (1) having fallen once or several times in the previous 12 months, (2) being afraid of falling, or (3) expressing a concern about balance. Baseline data collected for the effectiveness study were used for the present investigation.

Data Collection Procedures

Each participant was scheduled for a 2-hour assessment session that took place in their residential neighborhoods. All variables included in this study, except balance performance, were assessed during face-to-face interviews. Balance performance was assessed by physical therapists using a battery of standardized tests. To minimize measurement errors, all interviewers (n = 7) and physical therapists (n = 16) received formal training before data collection. They were also blinded to participants’ group assignment in the context of the effectiveness study.

Measures

Demographic characteristics. Demographic data included age, sex, education level, and perception of personal economic conditions. Except for sex, demographic variables were transformed into categoric variables using a tertile split. There was 1 missing observation regarding 1 participant’s education level. Instead of discarding that participant from the present study sample, the average education level was imputed to this person.

Balance confidence. As mentioned previously, a simplified version of the ABC scale was administered to participants to assess their levels of balance confidence. Each participant was asked to rate his/her confidence in his/her ability to maintain balance (“Up to what point are you confident that you will maintain your balance when you do the following activities?”) in performing 15 tasks of daily living by choosing 1 of the following 4 response options: not at all confident, slightly confident, moderately confident, or very confident. Given that a majority of study participants listed French as their main language and the fact that there were no available French-Canadian versions of the ABC scale at the time of the study, the cue question and items were also translated into French. The modified ABC scale was translated by 5 native French-speaking experts (rehabilitation and physical activity professionals and researchers) involved in the field of falls prevention who were fluently bilingual (French and English). After translation of the measure by each expert, the 5 versions were compared and team consensus was used to resolve minor discrepancies. Because consensus was readily achieved and given that the items of the scale represent tasks of daily living rather than nuances about thoughts and feelings, further efforts at cross-validation of translations were deemed superfluous. Minor changes in word usage were also made to adapt to the cultural uniqueness of the use of the French and English languages in Quebec (eg, expression “shopping center” rather than “mall” in the English version). This procedure was analogous to that reported by other researchers using the ABC scale in the United Kingdom and more recently in Canada.15

Perceived balance. Each participant’s overall perception of his/her balance was assessed with a Likert-type rating scale with anchors ranging from 1 to 10 with descriptive labels (ie, poor balance, moderate balance, excellent balance) and a single question: “Using the following scale, show me how good you think your balance is.” Perceived balance scores were transformed into dummy variables through a tertile split for subsequent statistical analysis.

Balance performance. Balance performance tests included the one-leg stance,23 the tandem stance,24 the tandem walking,25 the functional reach,26 and the lateral reach27 tests. All these tests have shown good reliability and validity.28,29 The one-leg stance test measured the amount of time a participant could stand on 1 foot without losing balance. Even though a maximum of 30 seconds has been suggested when completing this test with seniors,30 we opted for a 60-second ceiling because of the high level of functioning of participants. The test was performed while standing on each foot, as well as with and without sight. The tandem stance test measured the amount of time each participant could stand in a tandem position (heel-to-toe position) without losing balance. The maximum time was set at 60 seconds. Tandem walking assessed participants’ dynamic balance with a narrow support base (heel-to-toe walking). Participants had to walk along a marked line on the floor as fast as possible for a distance of 3m. The amount of time taken to walk this distance was measured in seconds. The functional reach test provided a measure of stability during a self-initiated forward movement of the upper extremities. It consisted of measuring the difference between a person’s arm length and maximal forward reach when the shoulders were flexed at a 90° angle using a meter stick attached to the wall at shoulder height. Each person had to maintain a fixed base of support during this test. The procedure for the lateral reach test was similar except that the self-initiated movement was performed laterally. It consisted of measuring the distance between each participant’s arm length and maximal lateral reach. This test was performed on each side of the body. Each balance performance test was administered twice, and the best performance score was kept for analysis. Balance performance scores were also transformed into dummy variables through a tertile split.

Fear of falling and falls occurrence. Fear of falling was assessed using a single question—“Are you afraid of falling?”—and 4-level response format (very often, often, occasionally, never). Participants who declared that they were never or occasionally afraid of falling were grouped in the nonfearful category, whereas participants who reported that they were often or very often afraid of falling were grouped in the fearful category. The number of falls experienced by each participant in the previous 12 months was also recorded. Data were dichotomized for analysis (0, no fall; 1, 1 fall or more).

Health problems related to falls. Participants had to indicate whether they suffered from a series of health problems related to falls such as dizziness, visual problems (eg cataract, glaucoma), and urinary incontinence. A dummy variable was created for use as a control variable indicating the presence or absence of such problems (0, no problem; 1, 1 problem or more).

Data Analysis

Descriptive statistics for each variable were produced to create a profile of the study sample. Distributions of responses to each item on the ABC-S scale were also examined. Next, we performed a series of analyses to examine the scale’s psychometric properties, namely the scale’s internal consistency, item hierarchy, and convergent validity. These analyses were conducted with multilevel modeling methods for polytomous outcomes. Multilevel modeling methods are especially adapted to data that have a nested structure.31 In the present case, ordinal responses to the 15 items of the ABC-S scale were viewed as level 1 observations nested within people (level 2 observations).
The multilevel modeling analysis proceeded in 4 steps. First, a simple model (called the null model) that included only each person’s response to each item of the scale and only 1 random effect on the intercept was tested. This model allowed for determining whether there was significant between-person variability in overall balance confidence (ie, latent construct of balance confidence). Second, dummy variables were created to differentiate 14 of the 15 items and were entered simultaneously into the model as level 1 predictors. The second item on the scale (going up and down stairs) was chosen as the reference category, because it was thought to represent a standard for independent mobility among community-dwelling seniors. Coefficients associated with dummy variables representing the other 14 items allowed for computation of an estimate of a reliability index reflecting the scale’s internal consistency (for more details about the reliability index used in multilevel modeling, see description by Raudenbush and Bryk\textsuperscript{31}). These coefficients also allowed for establishing the item hierarchy in the scale.

As a third step, a multilevel model that included age, sex, education level, and perception of personal economic conditions as control variables was tested. This was performed by entering dummy variables associated with each control variable into the model as level 2 variables modifying the intercept. As a final step, multilevel models that included dummy variables related to either (1) perceived balance, (2) balance performance measures, (3) fear of falling, or (4) occurrence of falls in the previous 12 months were tested. Again, dummy variables representing each of these variables were entered as level 2 variables modifying the intercept. This last step allowed for an examination of the associations between each of these variables and the latent construct of balance confidence and therefore tested the convergent validity of the measure. The multilevel modeling method used for data analysis has the benefit of maximizing datasets. More precisely, when a level 1 observation is missing (eg, a person does not provide an answer on 1 item of the scale), no listwise deletion of the case occurs. Rather, data from all answers provided by that participant are used in estimating a final model through empirical Bayes estimation procedures. However, if a level 2 observation is missing (eg, score on a balance performance test), the case is discarded from the analysis. The issue of maximizing datasets justified the choice of building separate models to establish the scale’s convergent validity. The equations in appendix 1 provide an example of the final multilevel models tested. As can be seen, the equations are consistent with a polynormal IRT model and more specifically with the Samejima graded response model, which involves cumulative probabilities.\textsuperscript{20}

A final set of ancillary analyses were performed to examine the potentially confounding role of health problems related to falls (eg, dizziness, visual problems, urinary incontinence). We re-ran all final models adding a dummy variable contrasting participants with falls-related health problems to those without such problems. A parallel procedure was performed to control for language of the administration of the questionnaire. All multilevel analyses were preformed with HLM software.\textsuperscript{4}

**RESULTS**

**Participant Characteristics**

Among the 200 participants, 1 could not complete the ABC-S scale because of a communication problem (slight expressive aphasia). Data for 2 other participants were excluded from the analyses because the interviewers expressed serious concerns about the quality of responses. That is, 1 participant completed the interview in a rush and another participant was interviewed at the end of the day and appeared very tired. Most questionnaires (88%) were administered in French.

Table 1 provides information on the demographics, perceived balance, balance performance, and other falls-related characteristics of the participants. Overall, the sample was composed of people in their midseventies, mainly women, with varied education levels and with average financial means. On average, participants perceived their balance as somewhat higher than the midpoint on a scale from 1 to 10. Table 1 also highlights some heterogeneity in the sample in terms of balance performance reflected by large standard deviation values obtained for balance performance tests. More than one third of participants were considered fearful (ie, being often or very often afraid of falling), and 38% of participants reported having fallen once or on several occasions in the previous 12 months.

**Between-Subject Variability and Reliability Analysis**

The first step of the multilevel analysis (ie, testing the null model) indicated that there was significant variability between participants in terms of overall responses to items as reflected by a statistically significant variance component ($\chi^2$ = 1175.3, $P < .001$). The reliability index (equivalent to the Cronbach $\alpha$) computed after inclusion of dummy variables distinguishing scale items was .86.

**Distribution of Responses to Items of the Modified Scale and Item Hierarchy**

Table 2 outlines the observed responses for each item of the modified scale and the item hierarchy as a function of ascendent...
Values were rounded off to the nearest integer.
Abbreviation: SE, standard error.

The right column of the table provides the proportion of between-subject variance accounted for by the addition of each set of variables to the multilevel model after adjustment for age, sex, education level, and perception of personal economic conditions. With the exception of lateral reach performed on the left side, Table 3 indicates that all odds ratios associated with dummy variables representing high and moderate levels of perceived balance and balance performance were statistically significant. This means that people in the highest and middle tertiles of scores on perceived balance and balance performance had higher balance confidence scores compared with people in the lowest tertile of scores on perceived balance and balance performance.

In addition, results in Table 3 also indicate that fear of falling and occurrence of falls in the previous 12 months were significantly associated with balance confidence. Seniors presenting low levels of fear had higher balance confidence scores compared with those with high levels of fear. Similarly, seniors with no history of falls in the previous 12 months had higher balance confidence scores compared with those who had fallen during this period. Finally, examination of the percentage of between-subjects variance explained by each variable in the analysis showed that perceived balance and fear of falling were the variables that explained the largest proportion of level 2 variation in balance confidence.
variance in models (27%). As indicated in table 3, there were a few missing observations in some of the multilevel models. However, these missing data represented less than 5% of observations for all level 2 variables tested, with the exception of tandem walking performance.

Ancillary Analyses

Ancillary analyses taking into account health problems related to falls did not influence the findings. Similarly, findings remained unchanged when models included a dummy variable reflecting language of administration of the questionnaire (English or French). Therefore, we chose to present only the results of more parsimonious models.

Global Balance Confidence Score

In an effort to maximize the utility of the ABC-S scale for practitioners, we explored the relevance of a simple and readily accessible method of computing an overall balance confidence score. Thus, we examined the correlation between participants’ estimates of balance confidence as computed with the Samejima graded response model and scores obtained by summing item responses (0, 1, 2, or 3). A large and significant correlation (r=.94, P<.001) was obtained, thus supporting the cogency of using a simple summing of responses to the 15 items as an indicator of global balance confidence.

DISCUSSION

The purpose of this study was to examine the psychometric properties of a modified version of the ABC scale, the ABC-S scale, through application of IRT methods and multilevel modeling techniques. The proposed amendments to the scale include simplifying the cue question and response format to improve the scale’s user-friendliness with seniors and removing the last item of the original scale (walk on icy sidewalks) to increase congruence of the scale with public health falls prevention strategies.
The results of this study indicate that the ABC-S scale has good internal consistency with a reliability index (equivalent to the Cronbach $\alpha$) of .86. This value appears somewhat smaller than Cronbach $\alpha$ coefficients reported in previous psychometric studies. However, if these Cronbach $\alpha$ coefficients are recast into a 4-level scale instead of a 101-level scale using a variant of the Spearman-Brown prophecy formula, we obtain values that are similar to the reliability index obtained in this study (eg, a Cronbach $\alpha$ of .96 is equivalent to .83). Furthermore, it should be noted that compared with the traditional classical test theory framework, the IRT framework used in this study has the advantage of providing an estimate of a scale’s internal consistency in a given population that is independent of the sample.  

This study also allowed for a determination of the difficulty level of each item and for the establishment of an item hierarchy. “Sweeping the floor” is the item with the highest probability of being very confident (see table 2). The multilevel model also suggests that “using an escalator without being able to hold the ramp” is the most difficult item in terms of balance with the lowest probability of a very confident response. It is interesting to note that in their initial psychometric study, Powell and Myers ordered the items of the ABC scale according to mean level of confidence obtained across study participants. Although some similarities were observed between the present item hierarchy and the ranking obtained by Powell and Myers, there are some discrepancies. In their study, the “reach at eye level” item obtained the highest mean level of confidence, whereas in the present study, “sweeping the floor” was estimated to be the easiest item in terms of balance confidence. This highlights the fact that item hierarchy should be established by considering 2 separate sources of variability: variability related to people and variability related to items.

Although the analysis of the item hierarchy showed differing degrees of difficulty across items, it also showed that for only 2 of the 15 items (ie, items 6 and 15), less than half of the sample indicated that they were not very confident in their ability to maintain balance. This finding underscores the need for identification of other tasks of daily living that might be viewed as more difficult in terms of balance by high-functioning seniors and might allow for the early detection of balance risk in the context of preventive interventions.

Results of this study also provide evidence of the convergent validity of the scale. Indeed, results indicated that seniors with higher and moderate levels of perceived balance or with higher and moderate balance performance scores (except for left lateral reach performance) had significantly higher balance confidence compared with seniors with lower levels of perceived balance or with lower balance performance scores. These results support the convergent validity of the scale. Examination of associations between fear of falling or occurrence of falls in the previous 12 months and balance confidence also supports the convergent validity of the scale. The fact that perceived balance and fear of falling explained a larger proportion of variance in balance confidence than actual balance performance is not surprising. Indeed, perceived balance, fear of falling, and balance confidence are 3 psychologic falls-related concepts. It is likely that a person’s balance confidence related to task performance depends on the person’s overall subjective appreciation of his/her balance. Other data also show that balance confidence and fear of falling are closely related constructs. Finally, the results of this study support the cogency of using a global balance confidence score computed by the simple addition of scores obtained on the 15 items of the scale.

This study also opens the door to a new set of investigations into the impact of falls prevention interventions. That is, given that balance confidence and actual balance are so strongly associated, a relevant question is to examine how balance confidence is influenced by intervention efforts designed toward improving actual balance or by decline in balance due to the normal aging process or health problems. Identifying how and when balance confidence changes offers the possibility of improving interventions to include balance confidence either as an intermediary target of intervention or as a valued outcome of intervention for improving quality of life.

**Study Strengths and Limitations**

This study provides evidence in support of the validity and reliability of a balance confidence scale that could be useful to interventionists working toward falls prevention in community-dwelling seniors. Modifications brought to the cue question and response format of the original ABC scale aimed at rendering this measure more user-friendly to community-dwelling seniors. Also, the removal of 1 item of the original scale (walk on icy sidewalks) aimed at promoting a better congruence with public health falls prevention strategies. Finally, the IRT framework used in this investigation allowed for establishing the scale’s item hierarchy. This hierarchy could be helpful in guiding interventions. However, investigations of the sensitivity of this measure to change in balance confidence after the interventions are needed.

Despite these strengths, the results of this study must be interpreted in light of some limitations. Although the sample size was large enough to test the psychometric properties of this simplified version of the ABC scale and to detect relevant associations, it was relatively small in the context of population research. Future research aimed at replicating and extending the current findings is warranted. In addition, the sample of this investigation included high-functioning seniors who volunteered for a falls prevention intervention and who had experienced falls in the previous year or were concerned about their balance. Further investigations into the generalizability of the current findings with other subpopulations such as community-dwelling seniors receiving home care services or who are not informed volunteers for falls prevention interventions are certainly warranted.

**CONCLUSIONS**

This study responds to a need to better establish the psychometric properties of available measures for the assessment of psychologic constructs related to falls. This study provides evidence in support of the validity and internal consistency of a modified, more user-friendly version of the ABC scale. Its psychometric qualities, its simplicity, the fact that it can be rapidly administered, and its high correlation with overall scores produced through IRT and multilevel methods render this scale useful for interventionists who are interested in assessing psychologic factors related to falls without unduly increasing respondents burden. The fact that the ABC-S scale was validated with high-functioning seniors renders this measure particularly well suited for the detection of community-dwelling seniors who are beginning to present risks and who could benefit from community-based falls prevention interventions. The ABC-S scale also has high potential for assessing the effect of community-based falls prevention programs on 1 psychologic falls-related factor, namely balance confidence associated with daily living tasks. In keeping with other parallel initiatives regarding another psychologic falls-related measure, we believe that research and intervention in falls prevention will be accelerated through the availability of psychometrically valid yet user-friendly measurement instruments.
Acknowledgments: We thank Céline Gosselin, MA, and Francine Trickey, MSc, from the Direction de Santé Publique de Montréal for their contribution to the scale’s modification and translation into French.

APPENDIX 1: EQUATIONS FOR THE FINAL MULTILEVEL MODEL TESTING THE ASSOCIATION BETWEEN BALANCE CONFIDENCE AND PERFORMANCE ON THE TANDEM STANCE TEST

Level 1 equations:
Prob (R = 1 very confident|B) = P'(1) = P(1)
Prob (R ≤ 2 moderately confident|B) = P'(2) = P(1) + P(2)
Prob (R ≤ 3 slightly confident|B) = P'(3) = P(1) + P(2) + P(3)
Prob (R ≤ 4 not at all confident|B) = P'(4) = 1.0

Where:
P(1) = Prob[Y(1) = 1|B]
P(2) = Prob[Y(2) = 1|B]
P(3) = Prob[Y(3) = 1|B]
Ln[P(1)/(1 – P(1))] = β̂0 + Σβ̂nj (Item i)
Ln[P(2)/(1 – P(2))] = β̂0 + Σβ̂nj (Item i) + threshold (2)
Ln[P(3)/(1 – P(3))] = β̂0 + Σβ̂nj (Item i) + threshold (3)

Level 2 equations:
β̂0 = γ̂00 + γ̂01 (Age_Mid) + γ̂02 (Age_High) + γ̂03 (Male) + γ̂04 (Educ_Low) + γ̂05 (Educ_Med) + γ̂06 (Educ_High) + γ̂07 (Econ_Mid) + γ̂08 (Tandem_Mid) + γ̂09 (Tandem_High) + uj;
β̂nj = γ̂1j;
β̂nj = γ̂2j;
β̂nj = γ̂3j;
...;
β̂nj = γ̂9j;
For i = 1 to 15, where β̂nj = 0.

Other final models are analogous, although operational links to coefficients γ̂08 and γ̂09 reflect operationalized perception of balance, balance performance on other tests, fear of falling, or falls occurrence.

References


Supplier

The Effects of Tai Chi on Bone Mineral Density in Postmenopausal Women: A Systematic Review

Peter M. Wayne, PhD, Douglas P. Kiel, David E. Krebs, PhD, Roger B. Davis, ScD,
Jacqueline Savetsky-German, MPH, MAOM, Maureen Connelly, MD, Julie E. Buring, ScD


Objective: To evaluate the evidence for Tai Chi as an intervention to reduce rate of bone loss in postmenopausal women.

Data Sources: Literature search using Medline, Science Citation Index, Cochrane databases, China Biological Medicine Database, and additional manual reference searches of retrieved articles and personal libraries.

Study Selection: Randomized controlled trials (RCTs), prospective cohort studies, and cross-sectional studies that included Tai Chi as an intervention, and had at least 1 outcome related to measurement of bone mineral density (BMD).

Data Extraction: Authors critically reviewed studies, evaluated methodologic quality, and synthesized study results in a summary table.

Data Synthesis: Six controlled studies were identified by our search. There were 2 RCTs, 2 nonrandomized prospective parallel cohort studies, and 2 cross-sectional studies. The 2 RCTs and 1 of the prospective cohort studies suggested that Tai Chi-naive women who participated in Tai Chi training exhibited reduced rates of postmenopausal declines in BMD. Cross-sectional studies suggested that long-term Tai Chi practitioners had higher BMD than age-matched sedentary controls, and had slower rates of postmenopausal BMD decline. No adverse effects related to Tai Chi were reported in any trial.

Conclusions: Conclusions on the impact of Tai Chi on BMD are limited by the quantity and quality of research to date. This limited evidence suggests Tai Chi may be an effective, safe, and practical intervention for maintaining BMD in postmenopausal women. In combination with research that indicates Tai Chi can positively impact other risk factors associated with low BMD (eg, reduced fall frequency, increased musculoskeletal strength), further methodologically sound research is warranted to better evaluate the impact of Tai Chi practice on BMD and fracture risk in postmenopausal women.

Key Words: Bone mineral density; Exercise; Osteopenia; Osteoporosis; Rehabilitation; Tai Chi.

© 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

The 2004 Surgeon General’s report highlights that among U.S. women, osteoporosis and osteopenia—2 conditions characterized by lower than average bone mineral density (BMD)—are a serious and growing public health issue. In 2002, the number of osteoporotic and osteopenic women over the age of 50 in the United States was estimated at 44 million. Because of baby-boomer driven anticipated changes in demographics, this number is expected to increase substantially in coming years.

Tai Chi (also referred to as Tai Chi Chuan, Taijiquan) is a meditative, mind-body exercise that is growing in popularity in the United States. Over the past century, millions of Chinese have practiced Tai Chi’s flowing, meditative movements to cultivate and maintain health and well-being. Because of its reputed health benefits, apparent safety, low cost, and growing popularity, Tai Chi has become an increasingly recognized preventive and rehabilitative therapeutic tool by the conventional medical community. Recent studies have begun to address the safety and efficacy of Tai Chi as a therapeutic intervention for a variety of health concerns including: balance and postural stability, musculoskeletal strength and flexibility, coronary artery disease, hypertension, general cardiopulmonary fitness and functional status, multiple sclerosis, rheumatoid arthritis, osteoarthritis, microcirculation and endothelial function, immune function, dementia, and general stress management. Several Tai Chi review articles have recently been published. The 2004 Surgeon General’s report on osteoporosis specifically recommends Tai Chi as a good exercise for fall prevention, and Tai Chi is increasingly recommended to osteoporotic women as a safe and effective exercise for bone density maintenance. Although the fundamental principles of Tai Chi and some clinical research suggest that it may help to maintain bone density in postmenopausal women, to date there have been few attempts to systematically evaluate the evidence for this claim.

This review examines the use of Tai Chi as a potential intervention for postmenopausal women with low BMD. We begin by highlighting the growing prevalence and public health impact of osteoporosis and osteopenia, and the current standard of care for these conditions. We then review the fundamental principles of Tai Chi that may make it beneficial for women with low BMD, and critically review clinical studies that have evaluated the impact of Tai Chi on BMD. We also summarize research on the impact of Tai Chi on other risk factors associated with osteoporosis and osteopenia. Last, we offer suggestions for future research that will improve our ability to evaluate the benefits of Tai Chi for both prevention and treatment of low BMD.

Low BMD: Definitions and Prevalence

Osteoporosis is a skeletal disorder characterized by compromised bone strength that predisposes one to an increased risk of
fracture. Bone strength primarily reflects the integration of bone density and bone quality. Because bone density can be easily measured, people are often classified as having osteoporosis or osteopenia based on the value of their BMD. Osteoporosis is technically defined by the World Health Organization as a BMD T score of less than −2.5 (ie, 2.5 standard deviations [SDs] below a healthy, young white adult reference).\(^{46}\) whereas osteopenia is often used to characterize BMD T scores between −1.0 and −2.5. Because bone tends to be lost with aging, untreated osteopenic women are at risk of losing additional bone and becoming osteoporotic.\(^{1,2}\) Low BMD is a strong risk factor for future bone fractures in asymptomatic postmenopausal women.\(^{41,42}\) It is estimated that 4 in 10 white women 50 years or older in the United States will experience a hip, spine, or wrist fracture sometime during the remainder of their life.\(^{43}\) Models based on meta-analyses indicate a doubling of relative risk for fracture with each SD decline in BMD.\(^{44}\) A recent prospective study\(^{45}\) of more than 200,000 women over the age of 50 reported that relative to those with normative BMD, women with osteopenia and osteoporosis had 1.8-fold and 4.0-fold increases in fracture rates, respectively. Although the relative risk of fracture was higher in postmenopausal women, because of the far greater number of osteopenic women (40% vs 7\%), the absolute numbers of fractures were much higher among osteopenic women, making osteopenia a potentially more widespread public health issue.\(^{45}\)

Fractures associated with low bone density are a significant cause of disability.\(^{46-47}\) The downward spiral in health after osteoporosis-related hip fractures is associated with up to 20% higher mortality rates in the year after a fracture.\(^{48-50}\) Recent studies\(^{51,52}\) from Sweden have shown increased mortality after spine fractures as well as hip fractures. Medical costs associated with managing hip and other fractures are high, and were estimated to be $13.8 billion in the United States in 1995, increasing to $17.5 billion in 2002.\(^{53,55}\)

Despite the high prevalence of osteoporosis and osteopenia, and the substantial burden on the health care system, only limited progress has been made in developing effective, preventive, and sustainable interventions aimed at reducing rates of fractures associated with low BMD. For example, recent findings\(^{54}\) suggest that although calcium with vitamin D supplementation in relatively healthy women without known osteoporosis may result in modest improvement in hip bone density, it does not reduce the risk of hip fractures, and it may increase the risk of kidney stones. Pharmacologic treatment to “prevent” further bone loss and fractures in women with osteopenia has been shown not to be a cost-effective strategy primarily because of the expense of the drugs.\(^{55}\)

Current guidelines for the treatment of osteoporotic and osteopenic women generally include the recommendation of regular exercise.\(^{1,2}\) There is currently no consensus regarding the optimal types and regimens of exercise for treating low BMD, however, or for addressing other risk factors associated with osteoporosis and osteopenia (eg, poor balance, decreased muscle strength, diminished agility). Moreover, among postmenopausal women, compliance with conventional exercise regimens is often low, due to health factors that may limit certain types of exercise, lack of motivation, and inability to sustain long-term interest, among other reasons.\(^{56-57}\) The 2004 Surgeon General’s report\(^{1}\) on osteoporosis stresses the need for new, creative, sustainable exercise programs for women at risk for low BMD.

Tai Chi and Its Rationale as a Treatment for Women With Low BMD

Tai Chi has its roots in the martial arts; yet for the past century millions of Chinese have practiced its flowing, meditative movements to cultivate and maintain health. Considered one of the treasures of Chinese medicine, Tai Chi is based on the same basic principles that underlie acupuncture and Chinese herbal therapies. It employs detailed regimens of physical movement, breathing techniques, and cognitive tools (both visualization and focused internal awareness) to strengthen the body, calm the mind, and “balance the flow of Qi” (life force).\(^{58-61}\)

A number of characteristics of Tai Chi practice that might make it an effective therapy for maintaining bone density and improving postural control have been explored in recent reviews.\(^{53,56,58,62,63}\) These intended characteristics and their purported effects include: (1) a constant shifting of weight from 1 leg to the other, which facilitates improved lower-extremity strength and/or mechanical load and dynamic standing balance; (2) an emphasis on maintaining a vertical posture with an extended head and trunk position, which promotes a less flexed posture; (3) the use of different parts of the body taking turns playing the role of stabilizer and mover, which enables movements to be executed smoothly without compromising balance and stability; (4) a continuous, slow, even tempo that facilitates sensory awareness of the speed, force, trajectory, and execution of movements, as well as awareness of the external environment; (5) the symmetrical and diagonal arm movements of Tai Chi, which promote arm swing in gait and increase trunk rotation around the waist; (6) moderate knee flexion, which lowers the body’s center of gravity; and (7) flowing circular and spiraling movements, which promote joint flexibility. Although we are not aware of any studies that have been explicitly designed to examine Tai Chi’s impact on mechanical load and BMD in postmenopausal women, a handful of studies have shown that Tai Chi improves lower-extremity biomechanic efficiency during activities of daily living.\(^{9,64-66}\) Such changes are likely to translate into increased mechanical load on key regions of the skeleton including the femur, hip, and lower spine.

METHODS

Clinical Trials Examining Tai Chi’s Effect on BMD in Postmenopausal Women

To systematically review the evidence evaluating Tai Chi for reducing rates of postmenopausal BMD loss, we conducted a literature search using Medline, Science Citation Index, and Cochrane Database of Randomized Controlled Trials. Search strategies for each of these databases included using the following statements and key words: Tai Chi or Tai Chi Chuan or Taijiquan and bone or osteoporosis or menopause, and included the period 1966 through April 2006. We also conducted a separate literature search using China Biological Medicine Database for Chinese-language randomized trials using the key words Taijiquan, bone, and osteoporosis. Finally, we manually searched the bibliographies of retrieved articles and our personal libraries for additional relevant citations.

Because only a small number of the studies we retrieved were randomized controlled trials (RCTs), and because RCTs employing Tai Chi interventions are not amenable to double-blinding, we chose not to use a more traditional instrument (eg, Jadad score)\(^{67}\) to evaluate study methodologic quality. Rather, study quality was descriptively characterized with respect to reporting of the following criteria: randomization (yes or no); details of randomization methods; clear inclusion and exclusion
Chi intervention were poorly described. Tai Chi instructors, and for all the non-RCT studies, dropout adequate details of randomization methods, inclusion and exclusion criteria; use of appropriate statistical analyses; details of Tai Chi intervention (eg, style, training schedule); and experience of Tai Chi instructors (table 1).

Our database searches of Medline, Science Citation Index, and Cochrane identified a total of 191 citations. Titles and abstracts of these citations were manually reviewed and considered eligible only if they described a prospective or cross-sectional study that employed Tai Chi as an intervention, and had at least 1 outcome related to measurement of BMD. A total of 9 citations met these criteria. Six of these 9 citations were limited to abstracts of proceedings from scientific meetings and were thus excluded; the remaining 3 were included in this review. Two additional eligible citations were identified using the China Biological Medicine Database for Chinese-language randomized trials, and 1 was identified in the personal library of an author of this review.

The 6 eligible studies identified by our search are summarized in table 2. Two were RCTs,68,69 2 were nonrandomized prospective parallel cohort studies,70,71 and 2 were static cross-sectional comparisons.72,73 Tai Chi practitioners were compared with age-matched sedentary controls in 5 studies68-70,72; 1 compared Tai Chi with rope skipping and vigorous martial arts,69 and 1 compared Tai Chi with acupuncture and Chinese herbal medicine.71 The duration of the Tai Chi intervention in the 3 prospective studies with naive practitioners ranged from 8 to 12 months. Five of the 6 studies were conducted in China and included only Asian women.

The methodologic quality of most studies was poor, as summarized in table 1. Of the 2 RCTs, only 1 provided adequate details of randomization methods, inclusion and exclusion criteria, dropout rates, and justification for sample sizes. None of the 6 studies indicated that outcome assessors were blinded, none included any information on the experience of Tai Chi instructors, and for all the non-RCT studies, dropout rates, sample size justifications, and characteristics of the Tai Chi intervention were poorly described.

Results across the 6 studies summarized in table 2 suggest the following: First, long-term postmenopausal women with higher BMD than age-matched sedentary controls, and have slower rates of bone loss. In 1 cross-sectional study of postmenopausal women, Qin et al used dual-energy x-ray absorptiometry (DXA) to compare BMD of 48 long-term Tai Chi practitioners with 51 age-matched sedentary controls. Subjects in the Tai Chi group had significantly higher BMD in the lumbar spine (7.1%), the greater trochanter (7.2%), and Ward’s area (7.1%) of the proximal femur (P<.05). Similar magnitudes of BMD differences between Tai Chi and age-matched sedentary controls were observed in an earlier study conducted by the same research group in a similar population.70 This earlier study also tracked changes in BMD over a 12-month period and found that rates of both trabecular and cortical BMD loss in the distal tibia (assessed using peripheral quantitative computerized tomography [pQCT]) were approximately 50% lower in the Tai Chi group (P=.044, P=.031). The Tai Chi group also exhibited a nonsignificant trend toward lower BMD loss in the femur (measured with DXA). Another cross-sectional study also reported greater spine and femur BMD among long-term female Tai Chi practitioners (n=18) when compared with age- and sex-matched controls (n=22) (P=.01).72

Second, Tai Chi-naive women who undergo Tai Chi training exhibit reduced rates of postmenopausal BMD decline. One methodologically sound RCT of postmenopausal women observed that those randomized to 12 months of regular Tai Chi training (n=67) exhibited 3.6-fold (trabecular) to 2.3-fold (cortical) reductions in rates of BMD decline in the distal tibia as measured with pQCT (P<.005), as compared with a no-exercise control group (n=65). No significant differences between groups were reported for BMD of the spine or femur as measured with DXA.68 A second, less methodologically sound RCT observed that DXA measures of BMD at the lumbar spine significantly increased (1.81%) after 10 months of Tai Chi whereas sedentary controls decreased (1.83%), (P<.005). Another intervention arm in this study—Tai Chi pushing hands (a 2-person interactive exercise that involves a continuous issuing and receiving of gentle pushes)—exhibited even greater increases in lumbar BMD (3.4%). This study also reported significant BMD increases of the same magnitude in the distal ulna and radius. Finally, another methodologically weak, nonrandomized study reported that 4 months of Tai Chi training resulted in a 7.3% increase in bone density (skeletal location not indicated) as measured with broadband ultrasound attenuation (BUA).71 This study also reported that serum osteocalcin, a biomarker for bone formation, increased significantly in the Tai Chi group. No BMD or osteocalcin data for the control group were provided.

Third, 1 nonrandomized cross-over study provided qualitative data suggesting that Tai Chi improves perimenopausal symptoms including hot flashes and abdominal distention. Finally, Tai Chi appears to be safe for peri- and postmenopausal women. No significant adverse effects were reported in any of the 6 studies evaluated.

### Table 1: Quality of Design and Methodologic Features of Studies Evaluating Tai Chi for Low BMD

<table>
<thead>
<tr>
<th>Features</th>
<th>Qin et al72</th>
<th>Gong et al72</th>
<th>Chan et al68</th>
<th>Zhuo69</th>
<th>Xu et al71</th>
<th>Qin et al73</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomization employed</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Randomization methods</td>
<td>NA</td>
<td>NA</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Clear inclusion/exclusion criteria</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Outcome assessors blinded</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Withdrawal and dropouts reported</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Sample size justified/estimated</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Appropriate data analysis</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Tai Chi intervention described</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Qualifications of Tai Chi instructors</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Abbreviation: NA, not applicable.

Legend: ✓, design and methodology feature adequately reported; –, design and methodology feature not adequately reported.
<table>
<thead>
<tr>
<th>Study</th>
<th>Study Design (Duration)</th>
<th>Study Location (Language)</th>
<th>Study Population (Age)</th>
<th>Interventions and Sample Size</th>
<th>Outcomes Measured</th>
<th>Results</th>
</tr>
</thead>
</table>
| Qin et al⁷⁰ | Prospective cohort (12mo) | Hong Kong (English)       | Postmenopausal community-dwelling women (54±3.4y) | • Long-term Tai Chi practitioners (min 4y experience) (n=17)  
  • Age- and sex-matched sedentary controls (n=17) | BMD of lumbar spine and proximal femur (w/DXA), and distal tibia (w/pQCT)                              | Significantly greater BMD in lumbar spine, proximal femur, and tibia in Tai Chi vs control. Reduced rates of BMD loss in Tai Chi, but trend significant only w/pQCT |
| Gong et al⁷² | Cross-sectional          | Shanghai, PRC (Chinese)   | Community-dwelling men and women (67.0±1.3y) | • Long-term Tai Chi practitioners (min 5y experience) (n=28)  
  • Age-matched sedentary controls (n=32) | BMD of lumbar spine and proximal femur (w/DXA)                                                       | BMD significantly greater in L1 through L4 and femur for Tai Chi vs control. 5–10y experience not different from 10+ |
| Chan et al⁶⁸ | RCT (12mo)               | Hong Kong (English)       | Postmenopausal community-dwelling women (54.0±3.5y) | • Tai Chi: 5 sessions/wk, 45min (n=67)  
  • Sedentary control (n=65) | BMD of lumbar spine and proximal femur (w/DXA), and of distal tibia using (w/pQCT)                    | Reduced rate of tibial bone loss in Tai Chi group (pQCT); nonsignificant trends in reduced rates of bone loss w/DXA |
| Zhou⁶⁹      | RCT (10mo)               | Shanxi, PRC (Chinese)     | Postmenopausal school teachers (55.9±2.8y)    | • 5 groups:  
  • Rope skipping (n=12)  
  • Mulan boxing (n=12)  
  • Tai Chi solo form (n=12)  
  • Tai Chi push hands (n=12)  
  • Sedentary control (n=12) | BMD of L2-4, distal radius and ulna (w/DXA)                                                          | BMD decreased in nonexercise control and increased in all exercise groups Tai Chi pushing hands significantly higher increases in BMD |
| Xu et al⁷¹  | Paired crossover design (8mo) | Melbourne Australia (English) | Menopausal women (49.3y) | • Tai Chi (n=12)  
  • Acupuncture (n=14)  
  • Chinese herbs (n=14)  
  Half of each cohort initially allocated to sedentary control; then crossed over at 16wk | Broadband ultrasound attenuation; bone formation marker (osteocalcin)  
 Bone resorption markers (pyridinoline, deoxypyridinoline)  
 TCM diagnoses | Tai Chi reduced rate of decline in broadband ultrasound attenuation; Tai Chi increased rate of bone formation (osteocalcin), but no effect on resorption; Tai Chi improved a number of menopausal symptoms according to TCM theory |
| Qin et al⁷³  | Cross-sectional          | Hong Kong (English)       | Postmenopausal community-dwelling women (55.5±3.1y) | • Long-term Tai Chi practitioners (min 3y experience) (n=48)  
  • Age-matched sedentary controls (n=51) | BMD of lumbar spine and proximal femur (by DXA), quadriceps strength, flexibility, balance            | Significantly greater BMD in lumbar spine and some regions of femur (greater trochanter, Ward’s area) in Tai Chi vs control. Greater quad strength and balance in Tai Chi vs control |

Abbreviation: TCM, traditional Chinese medicine.
DISCUSSION

Limitations of Reviewed Studies Evaluating Tai Chi’s Impact on BMD

Although the 6 studies summarized in table 2 suggest Tai Chi may improve BMD of postmenopausal women, these results should be considered inconclusive and interpreted with caution for the following reasons.

First, the majority of these studies have design and methodologic limitations. Only 2 of the 6 studies were randomized trials. Although cross-sectional studies enable investigations of long-term effects of Tai Chi, absence of randomization and longitudinal monitoring introduces great potential for bias. Sample sizes across all studies were small, with an average of 31 participants in Tai Chi and control groups. Additionally, information on Tai Chi interventions, eligibility criteria, blinding methods, and qualification of instructors were not available or poorly described (see table 1). Finally, because of both the diversity of approaches used to characterize BMD (ie, DXA, pQCT, BUA) and the diversity of skeletal sites at which BMD was assessed, comparisons and synthesis across studies is difficult. Moreover, some of the specific methods used to characterize BMD, such as BUA, are known to have very low precision; the 1 study employing BUA that reported a 7.3% increase in bone density after only 4 months of Tai Chi most likely reflects a measurement artifact.

Second, the studies in table 2 include women with a range of baseline BMD scores, ranging from normative to severely osteoporotic. Results of conventional exercise studies suggest that the responsiveness of BMD to exercise may vary with osteoporotic. Results of conventional exercise studies suggest that the responsiveness of BMD to exercise may vary with osteoporotic. Furthermore, some of the specific methods used to characterize BMD, such as BUA, are known to have very low precision; the 1 study employing BUA that reported a 7.3% increase in bone density after only 4 months of Tai Chi most likely reflects a measurement artifact.

The Impact of Tai Chi Versus Conventional Exercise on BMD in Postmenopausal Women

The results of the most methodologically credible RCT we reviewed suggests that 12 months of Tai Chi training for Tai Chi-naive practitioners resulted in a clinically and statistically significant reduction in the magnitude of tibial BMD loss of approximately 1.0%, as measured with QCT. The magnitude of this retardation in bone loss is similar to that reported in a meta-analysis characterizing the effects of walking exercise on reductions in rate of BMD loss in postmenopausal women’s BMD (1.31% and 0.92% retardation in bone loss for hip and spine, respectively, as assessed using DXA).76 This comparison, however, must be made cautiously, because BMD estimates were made at different anatomic sites and using different instrumentation. Moreover, it is difficult to measure “exercise dosage” across varied types of exercise and studies. Because no studies have directly compared Tai Chi with walking, there is no exact comparability in terms of exercise, intensity, and compliance across these studies.

The use of QCT to assess BMD in Tai Chi studies is noteworthy.68 In contrast to DXA, QCT has the advantage of being able to quantify true volumetric density as well as partition the 2 types of bone, trabecular and cortical, which may respond differently to exercise. Moreover, it has the potential to have higher precision.77 In the study by Chan et al68 the magnitude of both cortical and trabecular BMD loss in the tibia decreased by approximately 1% in response to Tai Chi. In contrast, a recent conventional exercise study that employed QCT to monitor BMD dynamics observed that a 2-year intervention combining high- and low-impact training exhibited markedly greater impact on cortical versus trabecular BMD of the spine (=3% vs 1%, respectively).78 Again, direct comparison between studies is limited by the different locations that BMD was assessed. Nevertheless, because cortical and trabecular BMD are known to contribute differently to the mechanical strength of bone, future Tai Chi studies should consider using QCT to better understand the impact of this weight-bearing exercise on bone morphology and quality, and the relationship between BMD compartmentalization and fracture risk. Even though cost is greater and radiation dosage is higher with QCT (radiation dosage is comparable to the background radiation dose obtained over the course of a year), it is justifiable to use this technology when testing new therapies because it is essential to understand how potential improvements in the skeleton are manifest.

Tai Chi May Also Impact Other Risk Factors Associated With Low BMD

Independent of changes in BMD, Tai Chi may be of benefit to women with low bone density because of its positive effect on postural balance and fall risk. Systematic reviews,3,4,6,80 which include numerous randomized trials, suggest that Tai Chi practice can directly reduce risk of falls,3,4 and/or positively impact factors associated with postural control including fear of falling,3,4 static and/or dynamic balance,4,6,80,81 musculoskeletal strength3,4,6,80,81,86,91 flexibility,3,4,12 and performance of activities of daily living.82-85 Drawing on these data, a cost-benefit analysis concluded that Tai Chi could significantly reduce costs associated with fall-related hip fractures.86 Noteworthy across these studies is that the majority have focused on older people, including frail3 and deconditioned87 adults, and as in the trials reviewed above, few adverse effects have been reported. This suggests that the findings are relevant to postmenopausal women, and that Tai Chi can be safely practiced well into later stages of life.

Only 1 of the studies included in this systematic review reported data on fracture rates. During their 12-month prospective RCT, Chan68 observed 1 fracture in the Tai Chi group compared with 3 in the control group. Because this study was not designed and powered to compare fracture rates, and because so few fractures were observed, this data should not be overinterpreted. Surprisingly, none of the cross-sectional Tai Chi studies we reviewed included information on prevalence of falls or fractures. Future long-term prospective and cross-sectional studies evaluating Tai Chi for bone health should also include data on prevalence of fractures.

CONCLUSIONS

Conclusions on the efficacy of Tai Chi for reducing rates of BMD loss in postmenopausal women are limited by the small number and generally low quality of studies to date. Acknowledging these limitations, the totality of the available evidence suggests Tai Chi may be an effective, safe, and practical intervention for maintaining BMD in postmenopausal women. One methodologically sound prospective RCT suggests Tai Chi reduces rates of tibial BMD loss, 2 sound cross-sectional...
studies suggest long-term Tai Chi practitioners have higher lumbar and femur BMD than age-matched sedentary controls. Other sound research summarized above indicates that Tai Chi can be of benefit to women with low BMD by improving balance, reducing fall frequency, and increasing musculoskeletal strength. Finally, Tai Chi has been shown to be very safe in aging, frail, and deconditioned populations, does not require equipment, and is relatively inexpensive to administer. As such, Tai Chi may be a logical and practical response to the Surgeon General’s recent call for novel exercise programs for women with low bone density.

Further research is warranted to better characterize the potential of Tai Chi, both as a therapy for women with low BMD and as a preventive intervention for women at risk for osteoporosis. This research should include appropriately powered randomized trials that include women representing a variety of races and ethnicities. This research should also explore the use of alternatives to DXA for the measure of BMD, including QCT, as has been suggested by the Surgeon General. Long-term prospective studies and cross-sectional studies should also include data on incidence of fractures. Combining the use of sensitive markers of BMD dynamics with objective measures of Tai Chi’s impact on biomechanic (eg, mechanical load at specific skeletal sites), physiologic, and psychosocial outcomes could provide important insight into the mechanisms by which Tai Chi impacts bone health. Finally, because low BMD is not a condition limited to women, future trials specifically evaluating the impact of Tai Chi on BMD in osteopenic and osteoporotic men may also be warranted.

Acknowledgments: We thank Weidong Lu for his assistance reviewing the Chinese language literature, and Monica Shields and Ellen Connors for administrative support. The contents of this review are solely the responsibility of the authors and do not necessarily represent the official views of the Center for Complementary and Alternative Medicine or the National Institutes of Health.

References

TAI CHI AND BONE MINERAL DENSITY, Wayne 679


60. Helm B. Gateways to health: T'ai jiquan and traditional Chinese medicine. T'ai jiquan J 2002;8-12.


Prevalence of Female Athlete Triad Characteristics in a Club Triathlon Team

Anne Z. Hoch, DO, John E. Stavrakos, MD, Jane E. Schimke, AAS


Objective: To determine the prevalence of the female athlete triad in club triathletes.

Design: Cross-sectional.

Setting: Academic medical center in the midwestern United States.

Participants: Fifteen women (mean age, 35±6y).

Interventions: Not applicable.

Main Outcome Measures: Disordered eating and menstrual status were determined by questionnaires. Energy status was determined by a 3-day food record, resting energy expenditure, and exercise energy output. Bone mineral density (BMD) was measured in the total left hip and lumbar spine (L2-4) by dual-energy x-ray absorptiometry.

Results: Sixty percent of the triathletes were found to be in calorie deficit, 53% had a carbohydrate deficit, 47% had a fat deficit, 40% had a protein deficit, and 33% had a calcium deficit. Forty percent of triathletes reported a history of amenorrhea. BMD was normal in the lumbar spine (L2-4) (1.3±0.1g/cm²) and total left hip (1.1±0.1g/cm²).

Conclusions: Triathletes are at risk for components of the female athlete triad. Continued efforts need to be directed at prevention through education of athletes, coaches, parents, and health care professionals.

Key Words: Amenorrhea; Athletics; Eating disorders; Female; Rehabilitation.

© 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

METHODS

Fifteen women from the same triathlon team volunteered to participate in this study and signed an informed consent in accordance with our institutional review board. All of the triathletes competed at a recreational level except for 2 who were professionals. Subjects trained on an average of 5 days per week for at least 90 minutes. Subjects reported starting athletics at a mean age of 12.6 years. None had a history of anorexia or bulimia. Subjects completed menstrual and health questionnaires, which were reviewed by the primary investigator (AZH). Eating habits and attitudes were assessed by the Eating Attitude Test (EAT-26), which has been validated for anorexia. Each subject completed a body image questionnaire.

We calculated energy status on each subject from a prospective 3-day food diary using a Nutrient Analysis program. Resting energy expenditure (REE) was calculated by the Harris-Benedict equation and multiplied by an activity factor of 1.1, which was used to estimate energy expended from daily activities of living. Energy burned through exercise was estimated using known calculations for time of exercise for running, swimming, and biking. These calculations were based on factors associated with the participants’ weight, age, sex, and intensity of exercise. The total calories expended from exercise was then added to the adjusted energy expenditure (REE × 1.1). This value of total energy expended was compared with the subjects’ caloric intake to determine if an energy deficit was present. The sports dietitian was able to calculate the participants’ daily carbohydrate, protein, and fat requirements and compare these values with the amount of macronutrients consumed. Diet records were recorded during the training season and consisted of 2 weekdays and 1 weekend day. Finally, calcium intake was recorded based on the 3-day food diary and compared with daily requirements for age.

We measured bone mineral density (BMD) (in g/cm²) of the lumbar spine (L2-4) and total left hip by dual energy x-ray absorptiometry (DXA) to assess for evidence of reduced BMD for age.

RESULTS

Mean age ± standard deviation (SD) of the subjects was 35±6 years. Nutritional analysis revealed that 60% of the triathletes were in a calorie deficit. Fifty-three percent had a carbohydrate deficit, 47% had a fat deficit, 40% had a protein deficit, and 33% had a calcium deficit. The average calcium deficit was −307±283mg, average carbohydrate deficit of the affected group was −245±187kcal, and the average fat deficit was −15±11g. Average carbohydrate and protein deficits were −70±30g and −15±7g, respectively. Calcium deficit is defined as caloric intake not commensurate with the athlete’s caloric expenditure. Calcium deficit was defined as an average daily intake of less than 1300mg/d. The average score on the EAT-26 was 4.3±4.4 (>15 is abnormal). This self-reported questionnaire consists of 26 questions and is scored on a 0-to-6 Likert scale. A score of greater than 30 on the EAT-26 represents a high likelihood of anorexia and a score between 15 and 30 represents a subclinical group with disordered eating habits and anorectic attitudes. The body image silhouette was 3.2±0.2. Subjects were asked to circle the body type they

From the Women’s Sports Medicine Program/Sports Medicine Center, Department of Orthopaedic Surgery (Hoch, Stavrakos, Schimke) and Physical Medicine and Rehabilitation (Hoch, Stavrakos), Medical College of Wisconsin, Milwaukee, WI.

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the authors or upon any organization with which the authors are associated.

Reprint requests to Anne Z. Hoch, DO, Women’s Sports Medicine Program/Sports Medicine Center, Depts of Orthopaedic Surgery/Physical Medicine and Rehabilitation, 9200 W Wisconsin Ave, Medical College of Wisconsin, Milwaukee, WI 53226, e-mail: azhoch@mcw.edu.

0003-9993/07/$15.00 © 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation.
considered to be the “ideal” triathlete from a validated 9-body silhouette scale,\(^6\) which ranges from an extremely thin or nearly anorexic figure at position 1 to a very overweight figure at position 9. A calorie deficit (>200kcal) was considered consistent with disordered eating for the purposes of this study.

Forty percent of the participants admitted to a history of primary or secondary amenorrhea. Average length of amenorrhea was 14±2 months. None were currently amenorrheic.

DXA revealed a mean BMD ± SD of 1.3±0.1g/cm\(^2\) in the lumbar spine (L2-4), which corresponds to a T score of 0.7 and z score of 0.8. The average BMD of the total hip was 1.1±0.1g/cm\(^2\), which corresponds to a T score of 0.7 and z score of 0.8.

Both T score and z score values correspond to values that are within normative limits based on the World Health Organization (WHO)\(^7\) classification system and criteria published by the International Society for Clinical Densitometry (ISCD).\(^8\) In 1994, WHO established criteria for making the diagnosis of osteoporosis, as well as determining levels that predict higher chances of fractures for postmenopausal white women. These criteria are based on comparing BMD in a particular patient to the average 20-year-old woman. BMD values, referred to as T score, that fall well below the average for the 20-year-old (stated statistically as 2.5 SDs below the average) are diagnosed as “osteoporotic.” If a patient has a BMD T-score value between −1.0 and −2.5 SDs below average, they are considered to be “osteopenic.” The z score is the number of SDs below average for a person of the same age, sex, and race. ISCD recommends using z scores for premenopausal adolescents and children. If a z score is less than −2.0 using a pediatric database of age-matched controls, then ISCD recommends using the term “low bone mineral density for age.”\(^9\) Subjects with a history of amenorrhea had the lowest BMD. This was not statically significant given the small sample size, however.

Overall, 60% of the subjects had at least 1 component of the triad, with 27% of these having 2 components of the triad (disordered eating and history of amenorrhea). Finally, 53% of the subjects were not able to identify all 3 components of the triad and 60% were not aware of daily calcium requirements for age.

**DISCUSSION**

This was a well-educated group of athletes that showed relatively normal body image. They showed several triad characteristics, however. Of the 3 components of the triad, disordered eating, specifically caloric restriction, was the most prevalent followed by a history of amenorrhea. None of the subjects had an abnormal BMD value. Over half of the team showed 1 or more triad characteristics.

The term disordered eating encompasses a spectrum of abnormal behaviors that may range from a mild preoccupation with food and exercise to a diagnosis of anorexia or bulimia according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition.\(^7\) These disorders often stem from internal or external pressures to be thin, or they may be the result of the misconception that thinness improves athletic performance. It is estimated that anywhere from 15% to 65% of female athletes suffer from some form of disordered eating pattern.\(^10\) In this study, 60% of the participants were found to be in a calorie deficit.

Athletic associated amenorrhea is a diagnosis of exclusion. It was formerly believed that the stress of exercise and low body fat were the causes of exercise associated amenorrhea. Recent studies by Loucks\(^11\) have found that diminished energy intake, rather than exercise, stress, or body fat depletion, was the regulating factor in the cessation of menses in active women.

BMD values in premenopausal women can be calculated by T score according to WHO.\(^7\) A T score of −1.0 or less is classified as osteopenia and −2.5 or less is considered osteoporosis. It should be noted that these labels may not be entirely applicable to this subject population, because the WHO osteoporosis criteria were designed for postmenopausal women in whom the mechanism of low BMD is the result of premature bone loss, and the T score was designed to predict fracture risk. ISCD\(^5\) recommends using the term “low BMD” in premenopausal women who do not have risk factors and have a z score less than 2.0. ISCD reserves the term osteoporosis for premenopausal women with secondary risk factors. Fortunately, none of the athletes in this study showed evidence of abnormal BMD in the hip or lumbar spine. In addition, BMD was obtained at 1 point in time in this study. Therefore, we were unable to determine who had a loss, gain, or no change over time.

**CONCLUSIONS**

Women triathletes are at risk for components of the female athlete triad, especially disordered eating. The participants in this study were college-educated women living in a major metropolitan area with normal body image, yet over half of these athletes were unaware of the triad and its potential dangers. This underscores the need for continued effort directed at prevention of the triad through education of athletes, coaches, parents, and other health professionals and for comprehensive clinical programs to treat the triad.

**References**


**Suppliers**

The Work That Remains at the Intersection of Gender and Career Development

Janet Bickel


Building on the study by Wagner et al, this commentary opens with a discussion of the persistence of gender disparities in career development and the challenge of interpreting those disparities. Given the multifaceted challenges facing rehabilitation medicine, facilitation of the career and leadership development of women physiatrists is critical. I suggest 3 areas of targeted action to facilitate the realization of women physiatrists’ intellectual capital: (1) updated approaches to faculty and leadership development and mentoring, (2) more flexible faculty structures, and (3) chair support and accountability. Each member of the rehabilitation medicine community who cares about the future of the discipline is challenged to contribute to the dialogues that are necessary to carry these recommendations forward.

Key Words: Career mobility; Rehabilitation; Women.

© 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

THE STUDY BY Wagner et al1 on the impact of gender on career development and leadership provides a timely wake-up call for rehabilitation medicine. After controlling for all major critical confounding factors, this investigation found that gender remained an independent and significant factor in salary and scholarly manuscript production. This study also showed that fewer women than men are achieving their academic and leadership potential. Corroborating these results, virtually all the other specialties and individual medical centers that have conducted such studies have drawn similar conclusions.

One critical variable, however, is missing from this analysis: career aspirations. Although this study found that women lagged significantly behind men in terms of leadership roles at both the institutional and national levels, it is unclear whether this trend is because women: (1) did not apply because they either lacked the time given their other responsibilities, did not imagine themselves capable of leadership, or were not effectively mentored; (2) were not offered the chance or did not successfully compete for the roles; or (3) declined an offer. As discussed later in this commentary, these possibilities are exceedingly difficult to tease apart. It is worthwhile to note that 1 institutional study2 found that, among medical faculty, there were no differences between men and women in self-reported importance of career advancement or aspirations to hold leadership positions but that women were much less likely to be asked to serve in leadership positions.

WHY SO SLOW?

Now that substantial numbers of women have been in the professional pipeline for 3 decades, why do such gender disparities in career development persist? Why has change been so difficult to accomplish?

To begin with, impetus for change is lacking. Women who are not realizing their potential tend to leave quietly or to remain invisible, and the costs associated with their wasted potential remain hidden as well. Second, the many young women now entering medicine, surrounded by women peers and unaware of their predecessors’ struggles, are assuming that all the heavy lifting to achieve gender equity has been completed and that they themselves will not have to settle for less. Most significant perhaps is the observation that many men—alert to the number of women entering medicine—are concluding that gender equity has been achieved. This stand is a “pipeline dream”—increases in the number of women are not reducing gender disparities in advancement.3,4

The Wagner1 study has shown, as have all other studies of these phenomena, that women are much more aware than men that the playing field is not level and women face continuing career disadvantages. Hard at work on their own career development and lacking a forum in which to comprehensively examine these issues, men tend to assume that women’s lack of progress stems from a lack of self-confidence, perhaps coupled with lack of appetite for competition or from a preference to devote more of themselves to their families.

THE CHALLENGE OF INTERPRETATION

Interpreting the results of even the best-designed studies on these subjects remains a challenge, with abundant ambivalence about the meaning of gender disparities in career development. Women are just as different from each other as men are and separating the influences of “nature” from “nurture” is almost impossible, especially given the common socialization that physicians experience. I thus conclude that, although this study draws much-needed attention to a critical challenge, the complexity of the multifarious variables affecting career and leadership development and gender differences is such that more studies will not provide better answers.

Take, for example, the finding that, after controlling for all major confounding variables, women physiatrists still earn only 75% of men. Clearly, the dynamics here are not nearly as simple as discrimination or family responsibilities, although women often do have priorities in addition to level of salary (eg, flexible schedule). Women are less likely to negotiate for the necessary resources at the start of their career, resulting not only in lower salary, but also in less space, less support staff, and fewer other resources that can be key to their success.5

Women do not negotiate as frequently or as successfully as

See article p 560

From Janet Bickel & Associates, Falls Church, VA.

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the authors or upon any organization with which the authors are associated.

Correspondence to Janet Bickel, Career Development and Executive Coach and Faculty Career & Diversity Consultant, 7407 Venice St, Falls Church, VA 22043, e-mail: Janebebickel@cos.net.

0003-9907/07/8805-11574$32.00/0

doi:10.1016/j.apmr.2007.02.038
men: both men and women negotiators take a harder line against women, making poorer first offers, pressuring women to concede more, and conceding much less themselves. This reflects different social conditioning and expectations of women’s worth—complex phenomena in themselves.

Women’s lower frequency and success in negotiating also reflect the strong women’s “likeability double-bind”: a woman’s influence increases the more she’s liked, but nonassertive women are better liked than assertive ones. Restricted to a narrower range of assertive behaviors than men are, aggressive women are penalized. That is, while men are expected to negotiate assertively, strong women negotiators may be perceived as unlikable, which actually interferes with their ability to build relationships key to their work.

Having a mentor and powerful career advocate is a clear advantage in building professional relationships in the highly political and competitive world of academic medicine. While women faculty are as likely as men to report access to a mentoring relationship, most studies show that women gain less benefit from these relationships in terms of career planning and less encouragement to participate in professional activities outside the institution, and that women are more likely than men to report that their mentor takes credit for their work and that their mentor is a negative role model. Successful men who from boyhood have had role models reflecting their aspirations often take this advantage for granted, thereby discounting the extra challenges women face in finding role models and in building developmental relationships.

Men are often not as forthcoming or comfortable with women as with men mentees, which can impede the value of the mentoring women receive. Also at the critical point when their women protégés begin spreading their wings and seeking more independence, some men withdraw their support. Some men seem more comfortable in paternalistic relationships with women (ie, father-daughter) than as equals and some feel threatened by the growing power of the protégé. These are some reasons why women’s informal networks are less extensive and less likely to include superordinates or colleagues from previous institutions. Because influence and credibility are tightly linked to visibility, lack of visibility within their institution or their network of colleagues means that it is harder to judge if a woman has the “right stuff” to lead.

Being underpaid and undermentored seems to translate into a personal “glass ceiling” —that is, women underestimating their own abilities and internalizing the difficulties they face. And as Stephen Jay Gould has remarked: “Few tragedies can be more extensive than the stunting of life, few injustices deeper than the denial of an opportunity to strive, by a limit imposed from without, but falsely identified as lying within.” All these factors reinforce each other, resulting in many women losing their ambitions and confidence and hence becoming less likely to successfully compete for raises, publication, and grants.

Even though the interactions of these influences are exceedingly difficult to tease out, this study is timely and important for the following reasons. One, given that 50% of medical students are women, the future of any specialty is inextricably linked to its development of women professionals. Two, women achieving more of their leadership potential will strengthen research, medical education, and patient care. Three, the paucity of women achieving senior ranks and leadership positions is becoming more of a liability. Diverse teams are more extensive than the stunting of life, few injustices deeper than the denial of an opportunity to strive, by a limit imposed from without, but falsely identified as lying within.”

The issues that Wagner et al delineate are not “women’s issues” —they are “our” issues as a profession because gender differences in career preferences are decreasing and because rehabilitation medicine needs all its members to realize their potential.

Based on my work at over 100 academic health centers (AHCs), I recommend focusing energies in the following directions: (1) updating approaches to faculty and leadership development and mentoring, (2) more flexible faculty structures, and (3) chair support and accountability.

Updated approaches to faculty and leadership development and mentoring

The professional development needs of women are best addressed within the context of assisting all faculty to realize their potential and to make the most of their faculty appointments. For example, as Wagner suggests: “Development of a mentorship program for residents, students, entering faculty, as well as chairs and senior faculty, may very well have a positive impact on the rehabilitation research enterprise. When implementing these mentorship programs, the specific needs of women within the current culture . . . should be emphasized.”

Mentoring is a professional activity that medical schools should formalize and recognize as a core academic responsibility. Faculty can be taught to improve their techniques of active listening, avoiding assumptions, and combining an optimal balance of support and challenge, thereby maximizing their impact in the limited time available for this activity.

The way in which individual leaders might be able to make the biggest impact is to improve their advocacy and mentoring.
of women. Men who make the extra effort to help women see their own potential, to connect them to key people, and to prepare them for leadership roles will have the satisfaction of witnessing a great impact on their development.

A primary goal of efforts described above is building a supportive ecology in which collegial relationships develop naturally. In addition to one-on-one mentoring programs, an emerging model is collaborative and peer mentoring programs, for instance, facilitated group mentoring that provides a framework for professional development, emotional support, and career planning. This updating of mentoring practices also responds to medicine’s need for new models of mutuality and “facilitative” leadership based on shared authority.

With regard to the development of leaders, with the generational differences mentioned above and with leadership challenges accelerating, it is risky to assume that the “cream” will continue to “rise to the top,” ready and skilled for tomorrow’s demanding leadership roles. Many AHCs are now creating internal leadership development programs. (A compilation of extant programs is available from Association of American Medical Colleges’ Faculty Development and Leadership Program.) While women should take advantage of such programs, they also greatly benefit from programs designed to equip them to manage the extra challenges they face. The Executive Leadership in Academic Medicine Program on Women and those offered by the Association of American Medical Colleges are worthy examples.

More Flexible Faculty Structures

As Wagner et al. note: sex differences in obtaining funding and in research productivity are probably linked to substantially differing personal responsibilities and to women’s being much less likely to have full-time support at home. The model of success in academic medicine is still based on the career trajectories of men not primarily responsible for home life, who completely devote themselves to their careers, and who tend to value independent work over team work. While this model no longer fits many young physicians, this “either advancement or family” thinking continues to interfere with academic medicine’s exploration of alternatives. Given that both men and women increasingly seek opportunities that permit robust personal lives and ways to integrate family with professional life, AHCs need to reduce where possible the most disadvantageous features of promotion policies, especially during the decade following residency training when most physicians have young children.

Part-time pathways that can expand and contract as personal issues emerge are vital to making academic competitive with other medical career paths. Adaptive structures include non-punitive less-than-full-time alternatives and adding off- and on-ramps. While such options may incur some up front costs, they are less expensive than recruiting and on-boarding replacements. Offering less-than-full-time options and temporal flexibility is effective stewardship of a department’s investment in young faculty in that the availability of such options also builds commitment and loyalty in people who have many decades of active professional contributions ahead of them.

Accountability

Department heads hold the keys to faculty vitality and to improving faculty-related practices. To be sure, department chairs already have a lot on their plates. But they need to include on their list of priorities ways to increase the percentage of women faculty at each rank. During annual performance reviews, some deans now include chairs’ progress in achieving greater faculty diversity. In particular, administrators and faculty should be accountable for their competencies in mentoring “across differences,” that is, people of a different sex, ethnicity, and career stage. If junior faculty are given the opportunity to evaluate their chairs and mentors on such indicators as “provides timely feedback that both challenges and supports me,” “advocates effectively for my development,” and “inspires me as a role model,” this will build a database that can be used for both summative and formative purposes. Linking effectiveness in these areas to a consequence of value, such as approval of new positions or access to faculty development resources, would add weight to this evaluation.

To accomplish these improvements, department chairs deserve support from their dean’s offices. The most forward-looking schools offer substantial faculty development programs and resources, building partnerships with chairs in becoming better developers of their human resources. They also offer educational sessions on improving mentoring and connection to coaches. A staple of leadership development in the corporate world, one-on-one executive coaching has been shown to increase the capabilities of motivated professionals, particularly in the areas of accomplishing objectives and improving relationships. Career and executive coaching can also greatly assist women in achieving their potential, especially for those who have lacked effective mentoring and for those just stepping into their first administrative role.

As Wagner recommends, accountability for compensation equity is also key, with departmental leaders checking regularly to ensure that women are fairly paid. For the reasons discussed above, sometimes “fair” can be difficult to determine; but regular examination assures that serious unexplained differences do not go unaddressed. Although paying women less may seem innocuous, such discrepancies in compensation can take down morale, increase attrition, and open a department to legal action; all of these undermine from the attractiveness of academic physiatry to young women.

CONCLUSIONS

Women are a burgeoning source of intellectual capital, but without targeted action to facilitate its realization, this talent will not reach fruition. As women comprise ever-increasing percentages of the talent pool and as the challenges facing physiatry multiply, access to and realization of this talent becomes more critical.

As Wagner recommends, the rehabilitation community must actively engage in a meaningful dialogue about how to effect necessary changes. I have suggested a number of credible directions. Each member of the community who cares about the future of the discipline is challenged to contribute to these dialogues and actions.

References

Letters to the Editor

Do We Really Want to Suggest 4 Transforaminal Epidural Steroid Injections Prior to Surgery?

In their review of the evidence for transforaminal epidural steroid injections (TFESIs) in the treatment of lumbosacral radiculopathy, DePalma et al.1 emphasized the need to repeat these injections prior to referral for surgical intervention. They referenced in particular a 2000 study by Riew et al.,2 which states that 13 of 19 patients who underwent more than 1 injection were able to avoid surgery.

A close review of the Riew article brings up several important points in terms of DePalma’s conclusions. Fifty-five patients in the Riew study received TFESIs—27 with bupivacaine only and 28 with bupivacaine and betamethasone (the experimental group). Patients were allowed to choose to receive as many as 4 injections at any time during the follow-up period. This was consistent with standard treatment in this setting.

The 19 patients who received multiple injections came from the overall sample of 55 patients, not from the group that only received bupivacaine and betamethasone. While there were 13 patients in the multiple-injection group who avoided surgery, the study did not allow for differentiation between the 2 study cohorts in terms of success.

Making a statement as strong as suggesting at least 4 TFESIs prior to a surgical procedure based on the Riew study does not follow the basic tenants of evidence-based medicine. The protocol in the Riew study appears to allow the patient to choose whether to repeat an injection, regardless of effect. This type of protocol does not allow for any definition of when a repeat injection should be made based on clinical evidence. The lack of breakdown between the treatment groups in terms of repeat injections allows for no conclusions to be made in terms of suggestions for frequency of TFESIs.

Of even greater interest is the 5-year follow-up of the original Riew study published in 2006.3 Of the patients available for follow-up who avoided surgery (21/29, with 9 from the group that received bupivacaine only and 12 from the bupivacaine and steroid group), there was no statistical difference between the groups with regard to the avoidance of surgery. These results emphasize the statement made by DePalma that definitive conclusions on the role of TFESIs in treating lumbosacral radiculopathy cannot be made due to the lack of a true placebo-control group.4 These findings also appear to be consistent with the statement that “The true utility of TFESIs may lie in their value to lessen lumbosacral radicular pain earlier than its natural progression without changing the chronology of complete resolution.”5(p1481)

Additional research is needed to determine the appropriate number of epidural steroid injections (regardless of approach), efficacy guidelines to help to determine when a repeat injection should be performed, and the role of injectate in overall treatment. In addition, we must continue to investigate the most appropriate placebo treatment to use as a control in these studies.

Suzanne Novak, MD, PhD
College of Pharmacy, University of Texas
Austin, TX
snovak@mail.utexas.edu

References

The author responds

Drs. Novak and Nemeth identify 2 important issues that were discussed in our review article. They are concerned about a “strong” recommendation that a patient should undergo multiple injection procedures prior to surgical referral. Their contention is that there is no evidence substantiating such a position. We concur with that view. It is exactly for that reason that we stated that “[p]erhaps only after 4” (emphasis added) injections should a patient be referred for surgery.1(p1481) Perhaps we should have used different language to ensure that we were not making a “strong” statement. The evidence we used is our experience at The Penn Spine Center. Observing the responses of thousands of patients who have presented with painful radiculopathy reveals that more than 1 injection is sometimes necessary if success is to be achieved employing medical rehabilitation and interventional spine treatment. As well, the randomized trial by Karppinen et al.2 revealed no incremental benefit when only 1 injection was used. We fully recognize that our conclusions may not be borne out by rigorous scientific inquiry, though we doubt it. Until such a study is conducted, we will use conditional statements about how many injections are needed.

Their second concern stems from an article4 that appeared after our critical view was published. Novak and Nemeth query whether a transforaminal selective nerve root injection is a technique that diminishes pain without changing the natural history of a herniated disk rather than a technique that provides a cure. We agree with their assessment. Understanding whether such an injection alters the natural history would certainly be valuable information. Further study is required before an answer is forthcoming.

More importantly, the question whether this procedure should be performed does not hinge on the answer. Treatment of radicular pain caused by a herniated disk is meant to diminish pain. If a selective transforaminal epidural provides enough relief that allows the patient to wait for the pathologic process to eventually resolve, then we have provided a wonderful service. The patient will have increased function and an improved quality of life as the inflammatory process subsides. Moreover, the probability the patient will seek surgery will be clinically and statistically significantly decreased as demonstr-
Strated in the study by Riew et al. This is certainly the worst-case scenario. If a selective nerve root block accelerates the healing process then that would be an even more compelling reason to perform this therapeutic technique. In summary, we categorically concur with Novak and Nemeth and appreciate their re-emphasis of these 2 issues.

Curtis Slipman, MD
Penn Spine Center
Hospital of the University of Pennsylvania
Philadelphia, PA
cslpman@mac.com

Michael J. DePalma, MD
Sheltering Arms Spine and Sports Center
Richmond, VA

References

doi:10.1016/j.apmr.2007.03.018

Corrections

In the Results section of the abstract (p1964), in table 2 (p1965), and in the Results of the study (p1966) of Wang YH, Huang TS, Liang HW, Su TC, Chen SY, Wang TD. Fasting serum levels of adiponectin, ghrelin, and leptin in men with spinal cord injury. Arch Phys Med Rehabil 2005;86:1964-8, all reporting of serum adiponectin and adiponectin levels should be in micrograms per milliliter (not nanograms per milliliter).

On page 1646 of Richter WM, Rodriguez R, Woods KR, Karpinski AP, Axelson PW. Reduced finger and wrist flexor activity during propulsion with a new flexible handrim. Arch Phys Med Rehabil 2006;87:1643-7, we misidentified the handrim, used and reported in a proceedings abstract by Baldwin et al (reference 13 in our study), as a beta prototype of the Natural-Fit handrim by Three Rivers Holdings. We have subsequently learned that the ergonomic handrim used by Baldwin was heavier and constructed of different materials than the prototype we purchased. Additionally and more importantly, the design of the handrim used in Baldwin differs substantially from the current Natural-Fit handrim that is now on the market. Consequently, the results of the Baldwin study do not apply to either the early version of the Natural-Fit handrim that we have or to the current version of the product. Also, the comparisons we drew between the flexible handrim design and the Natural-Fit handrim are irrelevant and should be disregarded. We apologize for any damage this may have done to the reputation of the Natural-Fit handrim or Three Rivers Holdings. The misrepresentation was unintentional.

We also selectively interpreted Baldwin’s discussion of results and chose only to include those results we felt were directly relevant to our study. That was short sighted; we should have discussed the complete set of results. As we noted, Baldwin did not find significant differences in electromyographic measurements between the handrim designs. However, Baldwin did find increased resultant forces on the handrim with the ergonomic handrim design. Based on these data, Baldwin suggested that the ergonomic rim allowed for greater force and moment application to the pushrim while not increasing muscle activity.
ARTICLE OBJECTIVES:

Article One: How Gender Impacts Career Development and Leadership in Rehabilitation Medicine: A Report from the AAPM&R Research Committee

Learning Objective: After completing this article and with appropriate self-study, the participant will be able to:

a) Compare men’s and women’s perceptions about the input of female gender on career development, access to mentorship, and leadership and promotion.

b) List 3 factors that were associated with achieving the rank of associate or full professor or achieving tenure.

c) Discuss the relationship between gender and rank and gender and leadership positions.

Article Two: Exercise Testing and Training in a Cancer Rehabilitation Program: The Advantage of the Steep Ramp Test

Learning Objective: After completing this article and with appropriate self-study, the participant will be able to:

a) Describe the training program given to patients 6 weeks after completing chemotherapy.

b) Recall the change in maximum oxygen consumption after an 18-week training program.

c) Discuss the value of interval fitness testing in patients with cancer.

Article Three: Recovery of Function in Skeletal Muscle Following 2 Different Contraction-Induced Injuries

Learning Objective: After completing this article and with appropriate self-study, the participant will be able to:

a) Describe 2 types of experimental protocols to induce muscle damage and compare them to 2 etiologies of muscle injury.

b) Compare the influence of myogenic cells on the recovery of muscles injured by multiple small contractions and those injured by a single lengthening contraction.

c) List 2 differences in the recovery after single lengthening contraction and multiple small strain contractions.

INDICATE THE DEGREE TO WHICH YOU AGREE OR DISAGREE WITH EACH STATEMENT

<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Disagree</th>
<th>Not Certain</th>
<th>Agree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Objectives for article one were met.</td>
<td>SD</td>
<td>D</td>
<td>NC</td>
<td>A</td>
</tr>
<tr>
<td>2. Objectives for article two were met.</td>
<td>SD</td>
<td>D</td>
<td>NC</td>
<td>A</td>
</tr>
<tr>
<td>3. Objectives for article three were met.</td>
<td>SD</td>
<td>D</td>
<td>NC</td>
<td>A</td>
</tr>
<tr>
<td>4. I learned a new skill or patient management approach.</td>
<td>SD</td>
<td>D</td>
<td>NC</td>
<td>A</td>
</tr>
<tr>
<td>5. This material will enhance my professional effectiveness.</td>
<td>SD</td>
<td>D</td>
<td>NC</td>
<td>A</td>
</tr>
<tr>
<td>6. I plan to implement a change(s) to my practice as a result of this material.</td>
<td>SD</td>
<td>D</td>
<td>NC</td>
<td>A</td>
</tr>
</tbody>
</table>

If you circled “Agree” or “Strongly Agree,” please give one example:

___________________________________________________________________________
___________________________________________________________________________

7. In what ways did/will you utilize the information from these articles in your medical practice? I have used/will use it to:

☐ Confirm previous knowledge and reinforce clinical practice
☐ Study for recertification examination
☐ Serve as initial resource for clinical problems
☐ Apply new techniques/procedures to the care of my patients
☐ Use the information to train staff
☐ Share the information with colleagues
☐ Help develop new policy and procedures
☐ Other (please specify):

8. The material was fair, objective, and unbiased toward any product or program. Yes .... No

9. Please share any general comments, recommendations, or an elaboration of any item on this form:

___________________________________________________________________________
___________________________________________________________________________

Evaluation data collected from this form will be processed confidentially by a third party and will only be reviewed by Academy staff in the aggregate.
CME PROCESSING FEES
AND APPLICATION

CME Certificates will not be processed without the completion of the Evaluation

<table>
<thead>
<tr>
<th>AAPM&amp;R/ACRM member</th>
<th>$20.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-member</td>
<td>$25.00</td>
</tr>
</tbody>
</table>

TO RECEIVE CME CREDIT: YOU MUST COMPLETE THE FOLLOWING 5 ITEMS.

The American Academy of Physical Medicine and Rehabilitation is accredited by the Accreditation Council for Continuing Medical Education to provide continuing medical education for physicians. The AAPM&R designates this educational activity for a maximum of 3 AMA PRA Category 1 Credits™. Physicians should only claim credit commensurate with the extent of their participation in the activity.

1. Membership/ID Number: [ ]

2. AAPM&R Member _____ ACRM Member _____ Non Member____

3. Enclose relevant Program Evaluation(s).

4. Number of hours:
Credit may be obtained for up to 3 years from date of publication.

☐ I studied the articles identified as eligible for CME in this issue of the Archives of Physical Medicine and Rehabilitation, consulted the reference materials as needed, and completed the corresponding evaluation.

Therefore, I claim _____ (fill in blank) AMA PRA Category 1 Credits™
(up to 3 hours).

5. To ensure that we accurately process your CME certificate, please print/type your full name, degree(s), and address as they should appear on your certificate.

Name and Degrees: ____________________________________________________________

Address: ___________________________________________________________________
                  ___________________________________________________________________
                  ___________________________________________________________________
                  ___________________________________________________________________

Signature: ___________________________ Date: _______________________

Your CME Certificate will be mailed to you following receipt of your application, evaluation, and check in the Academy Office.

PLEASE COMPLETE AND RETURN THIS CME APPLICATION, EVALUATION AND PAYMENT TO:
AAPM&R
PO Box 95528
Chicago, IL 60694-5528