American Journal Of Physical Medicine & Rehabilitation

Issue Table of Contents
June 2005, Volume 84, Issue 6

Commentary
Brain Injury

Soldiers with Occult Traumatic Brain Injury.
Lew, Henry L. MD, PhD; Poole, John H. PhD; Alvarez, Stephanie RN; Moore, William MD

Research Article
Electrodiagnosis

Under-Recognition of Polyneuropathy in Persons with Diabetes by Nonphysician Electrodiagnostic Services Providers.
Dillingham, Timothy R. MD, MS; Pezzin, Liliana E. PhD

Two Exercise Interventions for the Management of Patients with Ankylosing Spondylitis: A Randomized Controlled Trial.
Fernandez-de-las-Penas, Cesar PT; Alonso-Blanco, Cristina PT; Morales-Cabezas, Matilde MD; Miangolarra-Page, Juan Carlos MD, PhD

Green, David MD, PhD; Sullivan, Susan RN; Simpson, Janet RN; Soltysik, Robert C. MS; Yarnold, Paul R. PhD

Efficacy of Computer-Aided Dosing of Warfarin Among Patients in a Rehabilitation Hospital.
Mitra, Raj MD; Marciello, Michael A. MD; Brain, Carlos SM; Ahangar, Brian BS; Burke, David T. MD, MA

Training of Somatosensory Discrimination After Stroke: Facilitation of Stimulus Generalization.
Carev, Lee Anne M. PhD; Matvas, Thomas A. PhD

Diagnostic Precision of Ultrasonography in Patients with Carpal Tunnel Syndrome.
Keles, Isk MD [latin dotless i]; Karagulle Kendi, A Tuba MD; Aydn, Gulumser MD [latin dotless i]; Zog, S Gulfer MD; Orkun, Sevim MD

CME Article * 2005 Series * Number 4
Outcomes

Association of Hemoglobin Levels, Acute Hemoglobin Decrease, Age, and Co-Morbidities with Rehabilitation Outcomes After Total Knee Replacement.
Wang, Xin MD; Rintala, Diana H. PhD; Garber, Susan L. MA, OTR; Henson, Helene K. MD

CME Self-Assessment Exam

Study & Analysis
Outcomes

Assembling a Toolkit to Measure Geriatric Rehabilitation Outcomes.
Demers, Louise PhD; Desrosiers, Johanne PhD; Ska, Bernadette PhD; Wolfson, Christina PhD; Nikolova, Rossitza PhD; Pervieux, Isabelle MSc; Auger, Claudine BScOT

Answers for CME Self-Assessment Exam

CME Self-Assessment Exam.
Invited Review

Spine

Barr, Karen P. MD; Griggs, Miriam MD, FAAPMR, PT; Cadby, Todd MS, PT, ATC

Visual Vignette

Pseudo-Ulnar Neuropathy.
Robinson, Lawrence R. MD
ABSTRACT

Key Words: Brain Concussion, Electrophysiology, Neuropsychology, War

Recent advances in emergency medicine and protective armor technology have increased the survivability of soldiers on the battlefield and during troop transport. An unforeseen consequence of these advances has been an apparent shift in the distribution of various injuries. For example, the routine use of body armor has reduced the frequency and severity of injuries to vital organs within the torso region, while leaving the face and limbs exposed. Likewise, Kevlar helmets have greatly reduced penetrating head injuries from projectiles. However, helmets give limited protection against nonpenetrating forces from impacts, falls, and explosive blasts, which cause coup-contrecoup, torsion, and pressure-wave trauma to the brain. Thus, although more troops now survive injuries, the head remains susceptible to concussive forces. For example, among wounded soldiers admitted to Walter Reed Army Medical Center due to blast injuries, motor vehicle accidents, or falls, 60% have sustained a TBI (D. L. Warden, L. M. Ryan, Defense and Veterans Brain Injury Center, Washington, DC, 2004, personal communication, not an official statement of the United States Department of Defense). At our regional VA medical center, the number of TBI admissions has almost doubled in the past 2 yrs.

The special vulnerability of the brain is primarily due to two factors: the presence of bony protuberances on the inferior cranial vault, and the delicate composition the cerebral cortex, brainstem, and axonal fibers. Unlike the clinical evaluation of significant injuries to the body, which are usually accompanied by obvious physical signs and symptoms, the clinical assessment of closed-head injuries must rely on objective evaluations of altered consciousness, cognition, behavior, and where available, neuroimaging. For these reasons, it is not surprising that the diagnosis of TBI is often delayed or missed, resulting in tardy or inadequate treatment. This underscores the potential value of objective diagnostic procedures that may be more sensitive to occult brain injury.

Previous research has suggested that neuropsychologic and electrophysiologic evaluations can play an important role in verifying the diagnosis and severity of TBI. For example, it is now known that mild TBI is most likely to impair specific neuropsychologic functions: the speed and capacity of information processing, working memory, and complex problem-solving skills. Likewise, we have previously demonstrated that TBI is associated with altered cognitive event-related potentials (ERPs), especially the P300 component, and...
that these changes are related to the severity of injury and prognosis. Recent reviews of the P300 literature have indicated that this ERP component can provide clinically useful information on the efficiency of patients’ attention and short-term memory processes. Thus, the existing literature on mild TBI, and our recent experience with armored military personnel, exemplifies the role of both neuropsychologic and electrophysiologic measures in verifying and describing patients’ cognitive impairments.

**SUBJECTS**

We summarize two consecutive cases of injured soldiers whose initial mental status and treatment were in many respects typical of a growing number of patients referred to our care. The first patient is a 26-yr-old man with 15 yrs of education, previously employed as a computer programmer. The second is a 45-yr-old man with 13 yrs of education, previously employed as an engineering contractor. As passengers in a heavy-duty military vehicle rollover at high speed, both were wearing body armor. Fellow passengers received fatal injuries. The vehicle rolled onto the second patient, but his armor saved his life. Both patients lost consciousness for less than half an hour and had Glasgow Coma Scores of 14/15 on awakening (moderate confusion). They were transported to military hospitals, where they were treated for contusions and fractures below the cervical level. After magnetic resonance imaging of the brain, which were within normal limits, they were diagnosed with concussion due to their initially altered mental status.

After treatment of his physical injuries, the first patient was returned to light duty. He was reported to spend much of his time alone in his bunk and was considered for disciplinary action when he had persistent problems following through with his daily orders. He was evaluated further by the base neurologist, who observed decreased memory and processing speed and recommended a complete neuropsychologic evaluation. Five months after his injury, this patient was transferred to our regional Veterans Affairs medical center for comprehensive evaluation and inpatient rehabilitation.

The second patient was discharged from the hospital 4 mos after his injury and given 2 wks of leave at home with his family. His wife observed that he became easily confused about their daily routine and had difficulty remaining focused on conversations and tasks. She noted word-finding problems, frequent memory lapses, slow performance of familiar tasks, and frequent fatigue. She was especially concerned at changes in his personality, which included uncharacteristic irritability and argumentation, daily anxious rumination, and periodic depressed mood. She reported these signs to his commanding officer, and the patient was transferred to our medical center for comprehensive evaluation and inpatient rehabilitation.

**Neurology**

During their neurologic exams, both patients reported daily headaches. The first patient showed normal sensory, motor, cranial nerve, and other neurologic functions, with the exception of ataxic movements of his face and neck that he said were an attempt to reduce headache. The second patient had mild neurologic signs limited to the left side of his body: reduced 2-point discrimination in the hand, clumsiness during rapid finger movements, and diminished hearing. On the Mini-Mental Status Examination, both patients scored ≥24/30, within the conventional criterion of normal limits. However, using what we consider more appropriate, education-based norms, the first patient’s Mini-Mental Status Examination was clearly impaired (26/30 for orientation and recall errors) and the second patient’s score was low-average (28/30 for concentration and recall errors).

**Neuropsychology**

Figure 1 shows the patients’ cognitive profiles in four broad areas of ability: language, visual-spatial, attention-concentration, and short-term memory. In each area, both patients showed relatively preserved capacity for many basic cognitive tasks but had difficulty when greater speed and executive control were required (tests graphed to the right within each domain). In the verbal domain, for example, they showed average to high-average intellectual abilities (e.g., vocabulary, general information, verbal concept formation). However, their language skills were markedly impaired when demands for speed and mental flexibility were added (e.g., oral reading speed, color naming speed, verbal fluency). In the visual-spatial domain, both patients struggled with an executive problem-solving task (card sorting) that requires interpretation of ambiguous stimuli, making perseverative errors commonly seen in frontal injuries. Similarly, their attention skills were average to low-average on simple tasks (auditory digit span, sequential visual search), but they showed marked impairment on more complex speeded tasks (mental reversal of digits, serial mental addition, visual-motor symbol encoding, figural fluency). Across ten tasks in different cognitive domains, the first patient’s mental processing speed was about 40% slower than average for his age (severely impaired), and the second patient was about 30% slower than average (moderately impaired). On memory testing, the first patient was able to recall 3/3 words;
however, his recall of longer lists of objects, short stories, and the location of objects was impaired. This did not improve when he was given prompts or simply asked to recognize previously rehearsed information—a pattern that indicates impaired storage and retention of new information, not simply inefficient retrieval. The second patient showed impaired ability to recall the location of objects but was within normal limits on the other memory tests.

Neurobehavioral signs consistent with these test results were evident for both patients: reduced initiative, visibly slowed responses, frequent lapses of memory for recent events, word-finding problems, distractibility, and need for frequent redirection to tasks. Both reported ongoing head pain, fatigue, irritability, and difficulty sleeping. The first patient also showed extremely hypophonic speech, constricted affect, social avoidance, and complained of hypersensitivity to sound. The second patient reported multiple alterations in sensation: hand numbness, reduced sense of taste, loss of appetite, bilateral tinnitus, and mild but frequent visual hallucinations in his periphery that he understood to be illusory.

The first patient’s insight into his deficits was inconsistent. Although he acknowledged his need for recovery and rehabilitation, he stated voca-
tional goals that reflected his preinjury functioning rather than his current capabilities. He denied any depressed or anxious reactions that would be considered normal consequences of his injury and impairments. In contrast, the second patient showed greater awareness of his physical and cognitive declines in function, and he experienced periodic depressive and anxious reactions to these losses. He was able to state near-term goals that realistically included his need for rehabilitation and intermediate-term plans for switching to employment that would make fewer demands on his reduced processing speed. His insight into these issues was consistent with his relatively mild executive and memory deficits.

Electrophysiology

Following published procedures, both patients’ psychophysiological responses were measured as they attempted to detect infrequent “oddball” auditory targets. Figure 2 shows these patients’ ERPs compared with those of healthy individuals. Consistent with our previous findings for patients with moderate to severe TBI, both patients’ P300 responses were smaller and slower than normal. Overall, the first patient’s P300 waveform (70% lower amplitude and delayed latency of 56 msecs) was well below the 0.1 percentile of normalcy, which indicated profoundly impaired information processing. His manual reaction time (signaling his detection of the target stimulus) was more than half a second slower than normal, also in the profoundly impaired range. The second patient’s P300 waveform (60% lower amplitude, delayed latency of 40 msecs) was at the first percentile, indicating moderately impaired information processing. His manual reaction time was 124 msecs slower than normal, also moderately impaired.

DISCUSSION

It is noteworthy that both patients, like other cases that we have seen in our center, scored within normal limits on conventional cognitive screening criteria, including Mini-Mental Status Examination of ≥24/30, recall of three words, naming common objects, copying designs, digit repetition, and trail making. Although these kinds of tests readily identify patients who are severely impaired functionally, individuals with mild to moderate impairments can often do surprisingly well on them. This may reflect the reliance of these tests on skills that are familiar and overlearned in the general population (e.g., reciting telephone numbers, children’s connect-the-dots games, reciting the alphabet). Although the tests do include some requirements for mental flexibility and speed (e.g., serial-7 subtraction, trail making B), such demands are relatively minimal. In contrast, our patients showed clearly impaired functioning when they attempted tasks that make greater demands for mental speed, flexible problem solving, or incorporating large amounts of information. These clinical observations are consistent with research that has identified impaired speed and capacity of information processing, poor problem solving, and impaired complex working memory skills as the most frequent neuropsychologic signs after mild TBI.

Both of these patients’ P300 ERP waveforms indicated that their cognitive responses to stimuli were diminished in intensity and slower than normal. In research in moderate to severe TBI cases, we have reported similar changes in cognitive ERP responses to stimuli, which are related to the severity of injury and prognosis. Our growing experience with occult brain injuries suggests that these same ERP methods may also have clinical utility in assessing apparently mild TBI patients.

In addition to direct neurologic damage, traumatic injuries frequently have emotional sequelae.
that can also impair central nervous system functions. These include post-traumatic stress disorder (PTSD) and depression. When these emotional effects are severe, as in recurrent major depression or highly symptomatic PTSD, cognitive measures are typically depressed about 1 standard deviation. Therefore, when significant emotional symptoms are evident, clinicians must determine whether the extent and severity of cognitive impairment exceeds the expected effects of emotional sequelae. For example, case 2 complained of continuing anxiomorphic rumination and periodic depressed mood after his injury. These disturbances of thinking and mood may have mildly reduced some of his test scores, but they cannot account for his severe decrements on at least six measures ($Z < -1.7$ in Fig. 1).

Patients who have had combat injuries with loss of consciousness require efficient neuropsychologic screening. Such screening must balance the need for sensitive, reliable measures with the requirement to evaluate many individuals quickly. Based on the reviewed research and our own clinical observations, four types of tests would seem to be efficient additions to existing screening procedures: rapid oral naming and reading, timed visual symbol encoding, timed serial mental addition, and recall of long word lists. Examples of these tests in common clinical use are, respectively, the Stroop (color naming and reading subtests), Symbol-Digit Modalities, Paced Auditory Serial Addition (single modality), and the Hopkins Verbal Learning Test. This small battery adds $< 10$ mins to an evaluation but yields important feedback regarding patients’ capacity to process relatively high information loads normally. Three of the tests require very rapid mental, visual, auditory, and motor processing. Unlike most brief memory screens (using 3–4 items), the word memory test requires patients to learn and retrieve more information (12 items). All four tests are well normed by age and education, and each has demonstrated utility in efficiently discriminating individuals with normal vs. impaired cognition. Thus, prospective studies are warranted to verify whether these and other brief tests have sufficient sensitivity and specificity to identify those patients with apparently mild injuries who will manifest significant functional problems later. The goal, of course, is to efficiently identify at-risk individuals and direct them to appropriate clinical care before significant problems emerge.

A recent prospective study identified the following seven predictors of radiologically verified brain injury after mild TBI: trauma above the clavicle, head pain, vomiting, seizure, anterograde memory deficits, intoxication when injured, or age of $> 60$ yrs when injured. However, even structural radiologic evaluations can miss clinically significant brain injuries. Thus, although we agree with the guideline that patients be referred for neuroimaging if they meet any of these criteria after a head injury, our experience indicates that other warning signs are also very important when deciding whether additional treatment is warranted.

It has been suggested that the best way to identify occult TBI is a comprehensive assessment of physical symptoms, cognition, and mood. A recent review of the literature has identified the following constellation of signs and symptoms as the most common manifestations of significant postconcussive impairment in patients with mild TBI: headache, dizziness, decreased concentration, memory problems, irritability, fatigue, visual disturbances, noise intolerance, judgment problems, anxiety, and depression. With rare exceptions, the mild TBI cases referred to our center, despite their lack of visible brain lesions on magnetic resonance imaging, display all or most of these classic features. Ultimately, it is the sensitivity of family members, coworkers, and clinicians to these neurobehavioral signs that bring these patients to comprehensive clinical evaluation and treatment.

**CONCLUSIONS**

Advances in military body armor and helmets have reduced abdominal, thoracic, and penetrating head injuries for the modern soldier. Although this has increased survival on the battlefield, it seems to have raised the relative number of returning soldiers with TBI, many of whom have been referred to our acute rehabilitation center. Many of these patients showed relatively minor evidence of bodily injury in the field but met criteria for mild TBI: brief loss of consciousness, initial Glasgow Coma Scores of $> 12$, absence of external injury above the clavicle, and apparently normal neuroradiologic findings. They were initially judged fit for return to light duty or independent living at home. However, dramatic changes in the patients’ functioning and personality were noted by 3–5 mos postinjury. When these cases were referred to our center, we conducted comprehensive neuropsychologic, electrophysiologic, and functional evaluations that revealed significant impairments on tasks that require sustained attention, mental speed, multitasking, problem solving, or incorporating large amounts of information.

We conclude that patients with a history of head concussion should be referred for comprehensive assessment if they present with the following key signs and symptoms: (1) visual or auditory disturbances; (2) reduced attention, response speed, memory, or judgment; and (3) altered interpersonal behavior, personality, or mood. Also, the
potential overlap in the symptomatology of TBI with PTSD and depression cannot be ignored. Relying solely on initial Glasgow Coma Scores, observable external injury, neuroimaging, and brief cognitive screening procedures as diagnostic standards can result in underdetection and delayed treatment of individuals with significant brain injuries.

REFERENCES

Under-Recognition of Polyneuropathy in Persons with Diabetes by Nonphysician Electrodiagnostic Services Providers


Objective: Healthcare providers commonly refer patients to physiatrists and neurologists for electrodiagnostic testing when they have symptoms suggestive of a peripheral nerve disorder. Published practice guidelines specify that electrodiagnostic medicine consultants should possess special neurologic and procedural training in this area. We recently found that despite these practice guidelines, physical therapists, chiropractors, and podiatrists perform 17% of electrodiagnostic studies in the United States. These findings prompted the current investigation examining electrodiagnostic care across different providers for an important target population—persons with diabetes.

Design: A retrospective cohort of patients with diabetes who underwent electrodiagnostic testing in 1998 was identified in the MarketScan Commercial Claims & Encounters Database (The MEDSTAT Group) using CPT and ICD9CM codes. This database represents the healthcare claims for 16 million Americans in private and employer-based health plans. The outcome of interest was the rate of polyneuropathy identification across different providers, controlling for patient characteristics.

Results: There were 6381 electrodiagnostic encounters for persons with diabetes in 1998. Polyneuropathy identification rates were highest for physiatrists, osteopathic physicians, and neurologists (12.5%, 12.2%, and 11.9%, respectively). Podiatrists and physical therapists identified 2.4% and 2.1%, respectively, as having polyneuropathy—rates about one sixth that of physiatrists and neurologists despite controlling for casemix differences. Nonphysician providers who did not recognize polyneuropathy performed almost exclusively EMG testing (~90%) at the expense of nerve conduction studies.

Conclusions: This study raises concerns about the quality of electrodiagnostic testing by nonphysician providers for persons with diabetes. These results should prove useful for physicians, third-party payers, and health policy makers when confronting issues related to provision of electrodiagnostic services.

Key Words: Electrodiagnostic Services, Electromyography, Nerve Conduction, Physical Therapy, Chiropractor, Podiatrist, Polyneuropathy, Diabetes
Electrodiagnostic testing is a common diagnostic procedure used to evaluate patients with a wide variety of symptoms including pain, weakness, and numbness in a limb, and is generally provided by physiatrists and neurologists. This testing generally encompasses two components—nerve conduction studies and electromyography.

Evidence from a recent study suggests that, despite published guidelines for electrodiagnostic medicine consultants,1–3 a substantial proportion of electrodiagnostic consultations in the United States are conducted by nonphysician providers.4 Using data from a large and nationally diverse sample of privately insured persons from the 1998 MarketScan Commercial Claims & Encounters Database, the authors found that nonphysician providers accounted for nearly one fifth of all electrodiagnostic encounters in that population. Among nonphysician providers, physical therapists were the dominant providers of electrodiagnostic services (9.3%), followed by podiatrists (5.5%).4 The study also uncovered significant differences in the extent of testing across provider types, with nonphysicians generally performing less extensive testing than their physician counterparts.4 The relatively large proportion of studies conducted by nonphysicians, combined with their relatively limited extent of testing, raised concerns about quality of care for these patients and prompted the current investigation. Extrapolating the MarketScan rates for electrodiagnostic testing to the United States population reveals that such services account for approximately a half-billion dollars in direct costs to the healthcare system. This does not include indirect costs for care resulting from the electrodiagnostic consultation such as surgical interventions for median nerve decompression when carpal tunnel syndrome is identified.

Diabetes mellitus is an important medical condition that is dramatically increasing in incidence and prevalence in the United States.5–10 The current epidemic has been fueled by earlier onset of diabetes in obese youth as well as an increasingly aged population that is living with diabetes. Diabetes is a leading cause of renal failure, retinopathy, and blindness, as well as peripheral vascular disease and atherosclerosis resulting in lower limb amputations, especially in minority and Native American populations.11–15 A particularly prevalent comorbid condition for persons with diabetes, and one that often prompts electrodiagnostic consultation, is diabetic polyneuropathy.16–18 Painful diabetic polyneuropathy is a frequently disabling condition that requires accurate diagnosis through electrodiagnostic testing. Many medications successfully alleviate these symptoms, underscoring the importance of high-quality testing.16–18

The purpose of this investigation was to examine quality of electrodiagnostic care by different providers in persons with diabetes, using polyneuropathy as the indicator condition of interest. The importance of diabetic polyneuropathy, coupled with the current epidemic of diabetes in the United States, makes polyneuropathy suitable for use as a tracer condition with which to examine clinical care and outcomes for persons with diabetes undergoing electrodiagnostic testing.

METHODS

Data Source

Data for this analysis are drawn from the 1998 MarketScan Commercial Claims & Encounters Database (The MEDSTAT Group). The MarketScan database represents the inpatient and outpatient health care service use for over 16 million individuals nationwide who are covered by the benefit plans of large United States employers, health plans, and government and public organizations. The database links claims and encounter data to patient information across sites and types of providers, and over time. These data represent the medical experience of insured employees, early retirees, COBRA insureds, and Medicare-eligible retirees with employer-provided Medicare Supplemental plans, and their dependents. The annual medical database is constructed from data collected from over 50 large, generally self-insured individual employers, and includes private-sector health data from over 100 different insurance companies, including Blue Shield/Blue Cross plans, and third-party administrators. Both commercial claims and managed care encounters are included in the 1998 MarketScan database, which covers employees located in all 50 states, the District of Columbia, and Puerto Rico.

In addition to comprehensive utilization by provider type, service, and setting, the database includes demographic information (e.g., age, gender, state of residence) for all persons in the sample. The study was approved by the Johns Hopkins University and the Medical College of Wisconsin Institutional Review Boards.

Study Population and Definitions

Electrodiagnostic-related claims were identified by searching inpatient and outpatient claims files with the following CPT codes: 1) motor nerve conductions, CPT 95900 and 95903; 2) sensory nerve conductions, CPT 95904; 3) H-reflex testing, CPT 95934; 4) somatosensory-evoked potential testing, CPT 95925 and 95926; 5) repetitive nerve stimulation testing, CPT 95937; and 6) EMG testing, CPT 95860, 95861, 95863, 95864, 95867, 95868, 95869, 95870, and 95872.
The data structure of the MarketScan data was such that each individual component of the study (such as nerve conduction and EMG) was reflected by a separate claim. For instance, a provider might have submitted claims for motor and sensory nerve conduction studies and for electromyography, all performed during the same consultation. Because of this, all electrodiagnostic-related individual claims submitted by a single provider for the same patient on a given service date and with similar claim types (i.e., inpatient or outpatient) were combined into a single electrodiagnostic encounter or “episode” of care that reflected a complete electrodiagnostic consultation.

The sample was further restricted to electrodiagnostic studies among persons with diabetes. Patients were coded as having diabetes if any claim (electrodiagnostic-related or otherwise; inpatient or outpatient) during 1998 contained a diagnosis code for diabetes (ICD9CM; 250.0–250.9). It was assumed that if a person’s claim contained the diagnosis of diabetes at anytime during the year, then that person was diabetic at the time of the electrodiagnostic encounter.

Data extracted from the MarketScan database captured patient sociodemographic status and electrodiagnostic encounter characteristics. Patient characteristics included age (yrs), gender, employment status (employed full time, part time, other), and geographic region of residence (Northeast, North Central, South, West). Electrodiagnostic characteristics included the number of studies per encounter and provider specialty. Provider specialty was mapped from carrier specific coding to MarketScan’s standard values and, as a first step, classified into one of ten groups: Physical Medicine and Rehabilitation (physiatry); Neurology; Orthopedic Surgery; Family Practice and Internal Medicine; Podiatry; Osteopathic Care; Chiropractic; Physical Therapy; Physician, specialty unidentified; and Facility, provider unspecified (e.g., “acute care hospitals,” “outpatient centers,” “rehabilitation facilities”). Claims grouped in the category of “physician, specialty unidentified” were then examined in depth and, whenever appropriate, were reclassified into one of the five main physician provider groups.

To assess and control for casemix across providers, comorbidity information was collected from healthcare claims for an individual over the full year. The MarketScan database contained unique patient identifiers that allowed us to link all 1998 claims for persons across all episodes of service use during that year, including inpatient and outpatient visits and admissions. For persons selected as having an electrodiagnostic study and an ICD9 code for diabetes, we examined all inpatient and outpatient claims for the presence of other comorbidities. The type of comorbidity (e.g., peripheral vascular disease, heart disease, cancer) as well as the number of comorbidities not including diabetes, were used in multivariate analyses to determine their influences on identification of polyneuropathy across different providers. The goal was to casemix-adjust the diagnostic identification rates for polyneuropathy across all providers.

Polyneuropathy at each electrodiagnostic encounter was identified by examining the two diagnostic codes available for each study in an episode of care. A diabetic patient was coded as having been diagnosed with polyneuropathy during an electrodiagnostic encounter if any studies (claims) associated with that episode had an ICD9CM code of 356.9, 356.4, 357.0, or 357.2. These codes included persons with diabetic polyneuropathy as well as those with idiopathic polyneuropathy, acute infectious polyneuropathy, and unspecified polyneuropathy. The latter ICD9CM diagnosis codes were included to capture those persons with diabetes who had a polyneuropathy, recognizing that other diagnostic tests are required to precisely identify the etiology of an underlying polyneuropathy.

**Statistical Analyses**

Electrodiagnostic encounters of persons with diabetes were contrasted across provider specialty groups by patient’s health and sociodemographic characteristics using univariate (t and χ²) test statistics. Nonparametric (Mann Whitney) test statistics were used to examine variation in non-normal characteristics of electrodiagnostic encounters, such as number of claims, across providers. To determine the independent effect of provider specialty on the probability of making a diagnosis of polyneuropathy during electrodiagnostic consultations for diabetic persons, we relied on multivariate techniques. Specifically, a probit specification was used to examine factors affecting the likelihood that a diagnosis was made. All data analyses were conducted using Stata 7.0 statistical software. Unless otherwise noted, only differences that were statistically significant at a P level less than 0.05 are discussed in the text.

**RESULTS**

There were 6381 electrodiagnostic encounters (consultations) for persons with diabetes in the MarketScan database, corresponding to 13.2% of all electrodiagnostic encounters among this privately insured population in 1998. Table 1 presents the distribution of electrodiagnostic encounters for persons with diabetes by provider specialty, according to sociodemographic and health characteristics of the patients. The mean age of diabetic persons receiving electrodiagnostic services was 51 yrs old. About two fifths of all encounters served male
<table>
<thead>
<tr>
<th>Provider Specialty</th>
<th>Percent Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Physician Providers</strong></td>
<td></td>
</tr>
<tr>
<td>Neurologist (n = 1587)</td>
<td>39.7</td>
</tr>
<tr>
<td>Physiatrist (n = 1626)</td>
<td>43.1</td>
</tr>
<tr>
<td>Physician, Unspecified (n = 598)</td>
<td>39.9</td>
</tr>
<tr>
<td>Family/Int. Medicine (n = 360)</td>
<td>36.0</td>
</tr>
<tr>
<td>Orthopedic Surgeon (n = 79)</td>
<td>40.0</td>
</tr>
<tr>
<td>Osteopathic Physician (n = 49)</td>
<td>42.8</td>
</tr>
<tr>
<td><strong>Nonphysician Providers</strong></td>
<td></td>
</tr>
<tr>
<td>Physical Therapist (n = 423)</td>
<td>38.8</td>
</tr>
<tr>
<td>Chiropractor (n = 51)</td>
<td>40.7</td>
</tr>
<tr>
<td>Podiatrist (n = 246)</td>
<td>39.0</td>
</tr>
<tr>
<td>Facility (n = 1362)</td>
<td>43.8</td>
</tr>
<tr>
<td>Total (n = 6381)</td>
<td>39.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patient Characteristics</th>
<th>Neurologist (n = 1587)</th>
<th>Physiatrist (n = 1626)</th>
<th>Physician, Unspecified (n = 598)</th>
<th>Family/Int. Medicine (n = 360)</th>
<th>Orthopedic Surgeon (n = 79)</th>
<th>Osteopathic Physician (n = 49)</th>
<th>Physical Therapist (n = 423)</th>
<th>Chiropractor (n = 51)</th>
<th>Podiatrist (n = 246)</th>
<th>Facility (n = 1362)</th>
<th>Total (n = 6381)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>39.7</td>
<td>38.1</td>
<td>37.8</td>
<td>43.1</td>
<td>53.2</td>
<td>30.6</td>
<td>38.8</td>
<td>49.0</td>
<td>43.5</td>
<td>40.3</td>
<td>39.7</td>
</tr>
<tr>
<td>Age group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–34</td>
<td>5.0</td>
<td>4.9</td>
<td>5.4</td>
<td>5.3</td>
<td>3.8</td>
<td>8.2</td>
<td>7.1</td>
<td>5.9</td>
<td>10.5</td>
<td>8.2</td>
<td>6.1</td>
</tr>
<tr>
<td>35–44</td>
<td>12.3</td>
<td>11.2</td>
<td>17.4</td>
<td>8.3</td>
<td>21.5</td>
<td>14.3</td>
<td>15.6</td>
<td>3.9</td>
<td>13.4</td>
<td>14.8</td>
<td>13.1</td>
</tr>
<tr>
<td>45–54</td>
<td>39.9</td>
<td>36.5</td>
<td>36.0</td>
<td>32.8</td>
<td>35.4</td>
<td>42.8</td>
<td>40.7</td>
<td>43.1</td>
<td>39.0</td>
<td>33.1</td>
<td>36.8</td>
</tr>
<tr>
<td>55–64</td>
<td>42.4</td>
<td>47.2</td>
<td>40.0</td>
<td>53.6</td>
<td>38.0</td>
<td>34.7</td>
<td>36.4</td>
<td>47.1</td>
<td>36.2</td>
<td>43.0</td>
<td>43.5</td>
</tr>
<tr>
<td>65 and older</td>
<td>0.4</td>
<td>0.1</td>
<td>1.3</td>
<td>0.0</td>
<td>1.3</td>
<td>0.0</td>
<td>0.2</td>
<td>0.0</td>
<td>0.8</td>
<td>0.8</td>
<td>0.5</td>
</tr>
<tr>
<td>Comorbidities</td>
<td>1.5</td>
<td>1.7</td>
<td>1.4</td>
<td>1.7</td>
<td>1.5</td>
<td>1.1</td>
<td>1.2</td>
<td>1.1</td>
<td>1.1</td>
<td>1.4</td>
<td>1.5</td>
</tr>
<tr>
<td>Employment status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full time</td>
<td>71.8</td>
<td>60.3</td>
<td>69.7</td>
<td>56.7</td>
<td>70.9</td>
<td>71.4</td>
<td>75.2</td>
<td>80.4</td>
<td>76.4</td>
<td>67.5</td>
<td>67.4</td>
</tr>
<tr>
<td>Part time</td>
<td>23.2</td>
<td>36.0</td>
<td>23.8</td>
<td>40.3</td>
<td>25.3</td>
<td>16.3</td>
<td>20.8</td>
<td>15.7</td>
<td>21.1</td>
<td>26.4</td>
<td>27.8</td>
</tr>
<tr>
<td>Not in labor market</td>
<td>5.0</td>
<td>3.7</td>
<td>6.5</td>
<td>3.1</td>
<td>3.8</td>
<td>12.2</td>
<td>4.0</td>
<td>3.9</td>
<td>2.4</td>
<td>6.1</td>
<td>4.8</td>
</tr>
<tr>
<td>Managed care</td>
<td>41.9</td>
<td>36.8</td>
<td>30.6</td>
<td>43.6</td>
<td>30.4</td>
<td>16.3</td>
<td>33.8</td>
<td>37.2</td>
<td>27.2</td>
<td>35.5</td>
<td>36.8</td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>13.3</td>
<td>11.4</td>
<td>26.9</td>
<td>23.9</td>
<td>13.9</td>
<td>2.0</td>
<td>15.4</td>
<td>21.6</td>
<td>9.3</td>
<td>17.1</td>
<td>15.5</td>
</tr>
<tr>
<td>North Central</td>
<td>11.1</td>
<td>60.8</td>
<td>7.4</td>
<td>17.8</td>
<td>12.7</td>
<td>46.9</td>
<td>22.5</td>
<td>17.6</td>
<td>6.5</td>
<td>38.9</td>
<td>30.7</td>
</tr>
<tr>
<td>South</td>
<td>61.8</td>
<td>21.0</td>
<td>52.3</td>
<td>47.0</td>
<td>57.0</td>
<td>20.4</td>
<td>53.2</td>
<td>32.9</td>
<td>28.4</td>
<td>33.9</td>
<td>41.4</td>
</tr>
<tr>
<td>West</td>
<td>11.2</td>
<td>6.0</td>
<td>7.2</td>
<td>9.2</td>
<td>12.7</td>
<td>16.3</td>
<td>7.3</td>
<td>7.8</td>
<td>48.4</td>
<td>6.6</td>
<td>9.6</td>
</tr>
</tbody>
</table>
patients (39.7%), and slightly over two thirds of them were for services to patients working full time. Overall, 36.8% of all encounters were covered by managed care plans.

Relative to physiatrists and neurologists, nonphysician providers—most notably, podiatrists and physical therapists—tended to serve a slightly younger population. On average, patients served by these providers were 3 yrs younger than those treated by neurologists and physiatrists (49 vs. 52 yrs old, respectively, P < 0.05). It is important to remember that elderly, Medicare eligible persons are not included in the MarketScan database, only those in private health plans. The mean number of comorbidities (exclusive of diabetes) among patients receiving electrodiagnostic services from nonphysician providers was 1.1 compared with 1.6 (P < 0.05) among physician providers.

Table 2 presents unadjusted and adjusted rates of identification of polyneuropathy by provider specialty. Unadjusted polyneuropathy identification rates were highest among physiatrists, osteopathic physicians, and neurologists (12.5%, 12.2%, and 11.9%, respectively). Podiatrists and physical therapists identified 2.4% and 2.1%, respectively, of all diabetic patients undergoing electrodiagnostic testing as having polyneuropathy—a rate about one sixth of that of physiatrists and neurologists. Chiropractors and orthopedic surgeons provided electrodiagnostic services to relatively small numbers of diabetic persons, yet did not identify any of those patients as having polyneuropathy.

The adjusted identification rates, shown in the last column of Table 2, are essentially identical to the raw identification percentages discussed above. These probabilities, which have been adjusted for a wide array of factors capturing variation across patients in dimensions that might have confounded the relationship between provider specialty and the probability of being diagnosed with polyneuropathy, reveal that the marginal effects of provider specialty on the probability of making the diagnosis are essentially unaffected by the inclusion of controls for patient’s age, gender, presence of comorbid conditions, work status, and geographic region of residence. Among physician providers, there were no differences in the probability of diagnosing a polyneuropathy among neurologists, physiatrists or osteopathic physicians. Type of provider was the only factor that influenced polyneuropathy identification.

Table 3 presents mean and standard deviations for individual tests (claims) per electrodiagnostic consultation. The mean number of studies per encounter performed by physician providers was 3.7, higher than that for nonphysician providers (2.9, P < 0.001) in the group with polyneuropathy. Physiatrists and neurologists exhibited similar patterns of testing, each performing about 3.5 studies per encounter. Physical therapists performed the lowest number of studies per consultation overall (2.9). There were no statistically significant differences among neurologists and physiatrists in the extent of testing across patients ultimately found to

### TABLE 2 Provider-specific polyneuropathy identification rates and adjusted probabilities among diabetic patients

<table>
<thead>
<tr>
<th>Provider Specialty</th>
<th>Number of Diabetic Patients</th>
<th>Percent with Polyneuropathy</th>
<th>Adjusted Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Physician providers</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neurologist (reference group)</td>
<td>1587</td>
<td>11.9</td>
<td>11.6</td>
</tr>
<tr>
<td>Physiatrist</td>
<td>1626</td>
<td>12.5</td>
<td>12.7 [P=0.82]</td>
</tr>
<tr>
<td>MD/physician unspecified</td>
<td>598</td>
<td>11.7</td>
<td>10.9 [P=0.67]</td>
</tr>
<tr>
<td>Family practice/ internal medicine</td>
<td>360</td>
<td>8.1</td>
<td>7.1 [P=0.01]</td>
</tr>
<tr>
<td>Orthopedic surgeon</td>
<td>79</td>
<td>0.0</td>
<td>a</td>
</tr>
<tr>
<td>Osteopathic physician</td>
<td>49</td>
<td>12.2</td>
<td>12.3 [P=0.72]</td>
</tr>
<tr>
<td><strong>Nonphysician providers</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical therapist</td>
<td>423</td>
<td>2.1</td>
<td>2.2 [P=0.00]</td>
</tr>
<tr>
<td>Chiropractor</td>
<td>51</td>
<td>0.0</td>
<td>a</td>
</tr>
<tr>
<td>Podiatrist</td>
<td>246</td>
<td>2.4</td>
<td>2.8 [P=0.00]</td>
</tr>
<tr>
<td>Facility</td>
<td>1362</td>
<td>7.2</td>
<td>7.1 [P=0.00]</td>
</tr>
</tbody>
</table>

Note: Numbers in brackets are P values. The sample consists of 6381 electrodiagnostic encounters for persons with diabetes. Adjusted probabilities are calculated at the individual level (then averaged over the entire sample) by applying parameter estimates obtained from multivariate Probit regressions that control for provider specialty and patient’s gender, age, employment status, union status, beneficiary status (self vs. dependent), health plan type (managed care vs. traditional indemnity), number of comorbidities, and region of residence. Individual-level predicted probabilities are obtained by assuming all persons in the sample were seen by a given provider type (e.g., Neurologist) while maintaining all other factors constant at their original values.

* = Provider specialty contributed no variation to the dependent variable, i.e., provider specialty with zero diagnoses of polyneuropathy.

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.
have polyneuropathy (3.5 and 3.3, respectively) and patients without polyneuropathy (3.1 and 3.0, respectively), indicating that similar examination and testing procedures were followed for both groups of patients. In contrast, physical therapists and podiatrists performed nearly twice the number of studies on patients they diagnosed with polyneuropathy relative to testing conducted on patients with diabetes whom they did not diagnose with polyneuropathy (2.9 vs. 1.4 for physical therapists and 3.2 vs. 1.8 for podiatrists, respectively).

In addition to differences in the extent of testing, we also observed marked differences across providers in the type of studies performed. Table 3 shows the percentage of studies that were motor nerve conductions, sensory nerve conductions, and needle EMG tests. The figures represent the average number and proportion of studies by type of tests performed for each provider group. Among physicians, the patterns of motor, sensory, and EMG testing were roughly similar across the polyneuropathy and nonpolyneuropathy patient groups, with all patients receiving approximately the same comprehensive testing of the motor and sensory peripheral nerves. In contrast, types of testing across patient groups differed substantially among nonphysician providers. Among podiatrists, for example, 22.2% of studies performed in persons diagnosed with a polyneuropathy were motor nerve conduction studies, another 22.2% were sensory nerve conduction studies, and the remaining 55.5% were needle EMGs. For persons not diagnosed with a polyneuropathy, however, 90.8% of the studies performed by physical therapists involved only EMG testing. This failure to examine motor and sensory nerves by nerve conduction likely contributed to the under-recognition of polyneuropathy by these providers. Among podiatrists, two thirds of all EDX studies performed in persons ultimately diagnosed with polyneuropathy were sensory nerve conduction studies and one third were needle EMGs, and there were no motor conduction studies. As with physical therapists, however, needle EMGs comprised most (93.3%) of all the studies performed in persons without a polyneuropathy diagnosis.

**DISCUSSION**

This study examined patterns of electrodiagnostic testing across different provider specialties. In particular, it focused on physician and nonphysician providers’ ability to identify a complex condition—polyneuropathy—among persons with diabetes. Using claims data available for a large and diverse sample of employees and their dependents, the probability of recognizing polyneuropathy among physiatrists and neurologists was nearly 6-fold that of nonphysician providers and orthopedists. Demographic and health differences across groups of patients seen by these different providers

<table>
<thead>
<tr>
<th>Provider Specialty</th>
<th>With Polyneuropathy</th>
<th>Without Polyneuropathy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Number of Studies (SD)</td>
<td>Percentage of Studies by Provider Group</td>
</tr>
<tr>
<td>Physician Providers</td>
<td>3.7* (2.3)</td>
<td>Motor 44.7</td>
</tr>
<tr>
<td>Neurologist</td>
<td>3.5 (1.7)</td>
<td>Motor 39.7</td>
</tr>
<tr>
<td>Physiatrist</td>
<td>3.3 (2.1)</td>
<td>Motor 36.7</td>
</tr>
<tr>
<td>MD/physician unspecified</td>
<td>5.8*† (5.1)</td>
<td>Motor 39.2</td>
</tr>
<tr>
<td>Family/internal medicine</td>
<td>3.2* (2.3)</td>
<td>Motor 2.2</td>
</tr>
<tr>
<td>Orthopedic surgeon</td>
<td>a</td>
<td>Motor a</td>
</tr>
<tr>
<td>Osteopathic physician</td>
<td>3.8* (1.2)</td>
<td>Motor 16.7</td>
</tr>
<tr>
<td>Nonphysician providers</td>
<td>2.9†* (1.6)</td>
<td>Motor 22.2</td>
</tr>
<tr>
<td>Physical therapist</td>
<td>2.9† (1.0)</td>
<td>Motor a</td>
</tr>
<tr>
<td>Chiropractor</td>
<td>a</td>
<td>Motor a</td>
</tr>
<tr>
<td>Podiatrist</td>
<td>3.2*† (1.3)</td>
<td>Motor 0</td>
</tr>
<tr>
<td>Facility</td>
<td>3.4* (2.2)</td>
<td>Motor 2.4† (1.8)</td>
</tr>
</tbody>
</table>

* Differences in mean claims per encounter between those with and without a diagnosis of polyneuropathy are statistically significant at the $P < 0.05$ level.
† Differences in mean claims per encounter between specific provider and neurologist (reference category) are statistically significant at the $P < 0.05$ level.

Provider specialty with zero diagnoses of polyneuropathy.
¶ These percentages refer to proportions of the total number of individual tests (nerve conduction studies and EMG) done by the entire group of providers across the two groups of patients—those with polyneuropathy and those without polyneuropathy.
could not explain the 6-fold differences in recognizing polyneuropathy. Relative to physician providers, nonphysician providers tended to conduct significantly less extensive testing, a practice pattern that, in combination with less awareness of polyneuropathy as an important comorbidity in patients with diabetes, likely contributed to their observed under-recognition of diabetic polyneuropathy. Furthermore, nonphysician providers rarely performed motor or sensory nerve conduction studies, a finding that likely contributed to the low rate of recognition of polyneuropathy among their diabetic patients. Adjusted identification rates indicated that the observed differences across provider specialties were not attributable to casemix differences.

A growing body of literature has begun to examine the division of labor between physicians and nonphysician providers. Physicians are viewed as the highest skilled practitioners, able to handle the most complex patients. They delegate to other healthcare professionals care tasks that are less complex or less demanding of knowledge and skill. State licensing authorities often recognize this distinction, but have come under increasing pressure from professional advocacy groups to expand the provision of medical services and scope of practice for nonphysician professionals to perform care traditionally reserved for physicians. Although physicians and patients can benefit from appropriate use of such physician extenders, quality information is necessary to fully inform the decisions regarding who is qualified to practice in specific areas of medicine.

Our findings raise concerns about nonphysician providers (physical therapists, podiatrists, and chiropractors), as well as physician providers with less neurologic and electrodiagnostic training than physiatrists and neurologists (most notably, orthopedists, family practice physicians, and internists) engaging in the practice of electrodiagnostic medicine. The steep rise in the incidence of diabetes in the past two decades, combined with recent evidence of earlier onset of the disease, especially among minority populations, highlight the importance of providing timely, high-quality, electrodiagnostic services to persons at risk for diabetes-related secondary conditions such as polyneuropathy. Patients with diabetes with painful polyneuropathy, who are not accurately diagnosed with this condition through electrodiagnostic testing, not only fail to receive appropriate medications for their pain, but run the risk of inappropriate surgical interventions for misdiagnosed entrapment neuropathies and radiculopathies.

Physician specialties for whom electrodiagnostic medicine constitutes only a small component of professional practice—most notably, orthopedists, family practitioners, and internists—were also significantly less likely than physiatrists and neurologists to identify polyneuropathy in diabetic patients. Osteopathic physicians, on the other hand, exhibited similar identification rates to those of neurologists and physiatrists. It is likely that the professional specialty code designation of osteopathic physician was used for claims purposes instead of physiatry or neurology codes reflecting their specialty areas of practice. We are unable to determine the specialty of providers in facilities, yet because a large number of studies were performed in this setting, it was appropriate to include such designation to provide a complete picture of electrodiagnostic services provision.

The concordance of physiatrists and neurologists in identifying polyneuropathy in this large national sample with demonstrated similar patterns of testing across groups suggests that these specialists, with substantial education and training in peripheral neurologic diseases, are rendering similar care. Both physiatrists and neurologists demonstrated comprehensive testing that assessed both motor and sensory peripheral nerves, and reached similar study conclusions. Because of the large number of patients seen by physiatrists and neurologists, as well as their identical rates of recognition, it is reasonable to consider these rates of polyneuropathy identification the best estimates for the true prevalence of electrodiagnostically confirmed polyneuropathy among nonelderly diabetic patients receiving electrodiagnostic services in the United States.

The MarketScan database provided a unique opportunity to examine electrodiagnostic services in a large national sample of persons with diabetes. An important limitation of this study, however, was the focus on privately insured persons employed by large firms that contribute data to the MarketScan database. The experiences of uninsured persons, as well as those of persons insured through public programs such as Medicare, Medicaid, and the Veterans Administration, are not represented in the data. A prospective study would be necessary to fully confirm whether there were any systematic differences in patient symptoms, physical examination signs, or severity and duration of diabetes that might explain the differences in diagnostic rates between physiatrists and neurologists, and other nonphysician and physician practitioners with different study outcomes (diagnoses). It is unlikely that such differences in clinical presentations would, however, fully explain 6-fold differences in diagnostic rates for a group of nonelderly persons with diabetes.

An important limitation regarding the generalizability of our findings relates to the fact that the MarketScan data reflected a younger population of privately insured persons. These findings do not directly apply to older patients. Studies focus-
The AAEM guidelines in electrodiagnostic medicine, published in 1999, present the best practice recommendations, derived from the scientific literature as well as expert opinion, for assessing persons suspected of polyneuropathy. According to those guidelines, adequate diagnosis of polyneuropathy requires that motor and sensory nerve conductions be performed in at least two limbs (e.g., four studies), and that EMG studies be performed in at least one distal muscle in both legs and an upper limb. Findings from examinations of electrodiagnostic testing by provider (shown in Table 3) indicate that nonphysicians fall short in their performance of sensory and motor nerve conduction testing, particularly among diabetic patients in whom they did not diagnose polyneuropathy. Further, the scope of testing by nonphysician providers differed substantially from the testing performed by physicians. In fact, among patients without polyneuropathy, nonphysicians performed needle EMG almost exclusively at the expense of motor or sensory nerve testing. Such a reliance on EMG, without the requisite nerve conduction studies, without polyneuropathy, nonphysicians performed by physicians. In fact, among patients from examinations of electrodiagnostic testing by provider (shown in Table 3) indicate that nonphysicians fall short in their performance of sensory and motor nerve conduction testing, particularly among diabetic patients in whom they did not diagnose polyneuropathy. Further, the scope of testing by nonphysician providers differed substantially from the testing performed by physicians. In fact, among patients without polyneuropathy, nonphysicians performed needle EMG almost exclusively at the expense of motor or sensory nerve testing. Such a reliance on EMG, without the requisite nerve conduction studies, may have contributed to the low rate of polyneuropathy recognition among nonphysician providers.

CONCLUSION

Recognition of polyneuropathy in nonelderly diabetic persons referred for electrodiagnostic testing is an important aspect of high-quality healthcare, particularly in view of rising rates of diabetes, and the disabling nature of painful diabetic polyneuropathy. Physical therapists, chiropractors, and podiatrists who perform electrodiagnostic testing, identified polyneuropathy in persons with diabetes at a rate about one sixth that of physiatrists and neurologists despite casemix adjustment. Other physician groups—orthopedists, family practitioners, and internists—demonstrated significantly lower rates of polyneuropathy identification as well. Underutilization of nerve conduction testing by nonphysicians likely contributed to these differences recognizing polyneuropathy.

Future investigations are necessary to fully interpret and confirm these findings. However, the results of this study should prove useful for physicians, insurers, and health policy makers when confronting issues related to electrodiagnostic services provision in the United States.

REFERENCES


5. Harris MI: Diabetes in America: Epidemiology and scope of the problem. Diabetes Care 1998;21 (suppl 3):C11–C14


Two Exercise Interventions for the Management of Patients with Ankylosing Spondylitis
A Randomized Controlled Trial


Objective: The purpose of this clinical trial was to evaluate the impact of a 4-month comprehensive protocol of strengthening and flexibility exercises developed by our research group versus conventional exercises for patients with Ankylosing Spondylitis (AS) on functional and mobility outcomes.

Design: Randomized controlled trial. Forty-five patients diagnosed with AS according to the modified criteria of New York were allocated to control or experimental groups using a random numbers table. The control group was treated with a conventional protocol of physical therapy in AS, whereas the experimental group was treated with the protocol suggested by our research group. The conventional intervention consisted of 20 exercises: motion and flexibility exercises of the cervical, thoracic, and lumbar spine; stretching of the shortened muscles; and chest expansion exercises. The experimental protocol is based on the postural affectation of the AS and the treatment of the shortened muscle chains in these patients according to the Global Posture Reeducation (GPR) method. This intervention employs specific strengthening and flexibility exercises in which the shortened muscle chains are stretched and strengthened. The study lasted 4 mos. During this period, patients received a weekly group session managed by an experienced physiotherapist. Each session lasted an hour, and there were 15 total sessions. Changes in activity, mobility, and functional capacity were evaluated by an assessor blinded to the intervention, using the following previously validated scores from the Bath group: BASMI (tragus to wall distance, modified Schöber test, cervical rotation, lumbar side flexion, and intermalleolar distance), BASDAI (The Bath Ankylosing Spondylitis Disease Activity Index), and BASFI (The Bath Ankylosing Spondylitis Functional Index).

Results: Both groups showed an improvement (prepost scores) in all the outcome measures, mobility measures of the BASMI index, as well as in BASFI and BASDAI indexes. In the control group, the improvement in tragus to wall distance ($P = 0.009$) and in lumbar side flexion ($P = 0.02$) was statistically significant. Although the rest of the outcomes also improved, they did not reach a significant level ($P > 0.05$). In the experimental group, the improvement in all the clinical measures of the BASMI index ($P < 0.01$) and in the BASFI index ($P = 0.003$) was statistically significant. The intergroup comparison between the improvement (prepost scores) in both groups showed that the experimental group obtained a greater improvement than the control group in all the clinical measures of the BASMI index, except in tragus to wall distance, as well as in the BASFI index.

Conclusions: The experimental protocol developed by our research group, based on the GPR method and specific strengthening and flexibility exercises of the muscle chains, offers promising results in the management of patients suffering from AS. Further trials on this topic are required.

Key Words: Ankylosing Spondylitis, Physical Therapy, Randomized Controlled Trial, Functional Index, Activity Index, Metrology Index
Ankylosing Spondylitis (AS) is a chronic rheumatic disorder affecting mainly the axial skeleton, which progressively limits spinal and thoracic mobility throughout the course of the disease. Pain and structural lesions within the evolutionary frame of the disease force the patient to adopt antialgical postures, leading to the typical distortions frequently observed in these patients: protrusion of the jaw, thoracic kyphosis, loss of lumbar curve, protraction of the scapular girdle, and flexion and internal rotation of the pelvic girdle. According to published series the prevalence of AS is 0.1% to 1.4%. The male to female ratio is about 2:1 to 3:1. Based on available data, AS patients have about a 50% increased risk of mortality. There is some evidence that the progression is stronger in the first 10 yrs of the disease, but it is also clear that the disease keeps on being active for further decades. Early limitation of spinal mobility has been identified as one of the most important prognostic factors.

There is no definitive treatment of AS, but good control of the disease can be achieved. Physical therapy is highly recommended. The aim of the physiotherapy treatment in AS is to maintain or improve general functioning and quality of life. The long-term goal for the patient is to try to maintain a good posture, with the primary aim being to avoid stiffening in a flexed position. Previous trials have analyzed the therapeutic effects of three modalities of physical therapy interventions: supervised individualized therapy, group therapy supervised by a physical therapist, and homework performed by the own patient. All these studies have demonstrated that physiotherapeutic exercise improves spinal mobility and also reduces functional impairment in these patients. Ramos et al reported that the best results in improving mobility and reducing functional impairment were obtained with group therapy, while individual therapy at home attained the worst results. This conclusion has been confirmed by the review of the Cochrane Musculoskeletal Group, which reported that supervised physiotherapeutic exercises in group have a greater short term beneficial effects than home exercise programs. Otherwise, the Cochrane Musculoskeletal Group has not found any randomized trial investigating different physical therapy protocols in patients with AS. Accordingly, we still do not know which particular treatment protocol to use when we meet a patient suffering from AS.

There are different exercises for AS. Conventional protocols of physical therapy consist of analytic flexibility exercises of the cervical, thoracic and lumbar spine, and analytic stretching of the erector spine muscles, hamstring muscles, and shoulder muscles. However, not all of these exercises are specific for patients with this disease. In France, Italy, and Spain there is a physical therapy method called “Global Posture Reeducation” (GPR), which was developed by Phillipe Souchard based on 20 yrs of clinical experience. This method deals with the existence of different muscle chains: the posterior static chain of the body (Fig. 1), the anterior diaphragmatic chain (Fig. 2), and the anterointernal chain of the pelvic (Fig. 3) and scapular girdles (Fig. 4). All these muscle chains are constituted by gravitational muscles (i.e., erector spine muscles, piriformis muscle, scalene muscles, suboccipital muscles, etc.) which work in synergist function depending on the muscle chain (e.g., the muscular function of the posterior static chain of the body is to permit the standing position against gravity). According to this method, the analytic stretching of any muscle could be expanded to include secondary adaptative or maladaptative changes in the rest of the muscle chain, so that the analytic stretching of any individual muscle would be inefficient, if not associated with a stretching of the whole muscle chain. Therefore, the GPR method employs specific strengthening and flexibility exercises in which these shortened muscle chains are stretched and strengthened.

From January to June of 2002, our research group performed a biomechanical analysis, based
on the principles of the GPR method, of the typical distortions frequently observed in AS patients: 

1. Protrusion of the jaw, caused by the shortening of both the anterior diaphragmatic chain (Fig. 2) and the posterior static chain of the body (Fig. 1).

2. Thoracic kyphosis, caused by the shortening of the anterior diaphragmatic chain (Fig. 2).

3. Loss of lumbar curve, provoked by the shortening of the posterior static chain of the body (Fig. 1).

4. Protraction and internal rotation of the scapular girdle, caused by the shortening of the anterointernal chain of the scapular girdle (Fig. 4).

5. Flexion and internal rotation of the pelvic girdle, provoked by the shortening of the anterointernal chain of the pelvic girdle (Fig. 3).

Based on that analysis, our research group developed an experimental intervention in which these muscle chains were strengthened according to the principles of the GPR method. There are many years of clinical experience in Europe and South America in the management of AS with the GPR method; however, we have not found any study in the peer-reviewed literature analyzing the effectiveness of this method in AS. The GPR method probably is based primarily upon expert opinions and clinical experiences in different countries. Therefore, the aim of this study was to assess, in a randomized controlled trial, changes on mobility and functional outcomes in AS patients who are treated with the GPR method. In addition, we expose the principles of this method in the management of patients suffering from AS. To our knowledge, this is the first paper analyzing different physical therapy protocols in patients with AS.

The purpose of this randomized controlled trial was to evaluate the impact of a 4-month comprehensive protocol of strengthening and flexibility exercises vs. conventional exercises for patients with AS on functional and mobility outcomes. We hypothesized that patients allocated to the comprehensive exercise protocol would demonstrate greater improvement in functional and mobility outcomes than those receiving conventional exercises.

**MATERIALS AND METHODS**

**Subjects**

Forty-five patients diagnosed with Ankylosing Spondylitis (AS) according to the modified criteria of New York and classified into four levels of functional affectation considered by the American College of Rheumatology participated in the present trial. Patients had been diagnosed and classified by their rheumatologist. In addition, functional classification for each patient was verified by the research group. All the included patients came from three hospitals in Madrid (Spain): Hospital Doce de Octubre, Hospital Severo Ochoa, and Fundación Hospital Alcorcón. Patients were excluded if they met criteria for functional class level IV, had a medical condition that impaired
function more than their AS, had osteoporosis, or had a history of fractures secondary to osteoporosis. The health situation of the patients was clinically stable, without current symptoms of any other concomitant chronic disease. All patients provided informed consent prior to beginning the trial. A demographic questionnaire was completed by all patients to assess clinical and demographic data. Patients were allocated to control or experimental groups using a random numbers table. The control group was treated with a conventional protocol of physical therapy, whereas the experimental group was treated with the protocol suggested by our research group. Patients were blinded to the physical therapy intervention.

The present clinical trial was supervised by the Research and Teach Unit of Physical Therapy, Occupational Therapy, Physical Medicine and Rehabilitation of the Universidad Rey Juan Carlos. It was approved by the Ethical Committee in Clinical Research of the University.

Physical Therapy Interventions

The study lasted 4 mos. During this period, patients received a weekly group session. Each experimental or control group comprised six to eight patients. An experienced physical therapist supervised each intervention. The characteristics and experience of the therapists will be exposed later. Each session lasted an hour, and there were 15 total sessions. During the study patients were emphasized not to modify their life habits. Ninety percent of the patients had been practicing swimming-pool exercises for 2 yrs. They were told not to receive any other physical therapy intervention during the study, which was an exclusion motive. This circumstance was reminded in all sessions. All the patients were taking different types of NSAID, and the rheumatologist assured that the pharmacological treatment was the same during the trial.

The control group received a conventional physical therapy intervention, which consisted of 20 exercises employed in previous studies (see Appendix 1). This protocol included analytic flexibility exercises of the cervical, thoracic, and lumbar spine. The intervention also included stretching of the erector spine muscle, hamstring muscles, and shoulder muscles. At the end of each session, chest expansion exercises and control abdominal and diaphragm breathing exercises were performed. This protocol was supervised by Physiotherapist 1 (PT1).

The experimental group received a different physical therapy intervention, according to the postural affectation of the AS. This intervention was based on the treatment of the shortened muscle chains, following the guideline described by the GPR method. This method employs specific strengthening and flexibility exercises in which these muscle chains are stretched and strengthened. Some of these exercises are: eccentric work of the erector spine muscles (Fig. 5), stretching of the posterior muscle chain in the pelvic region (Fig. 6), and different specific

FIGURE 4 Anterointernal chain of the scapular girdle. The anterior-internal chain of the scapular girdle is constituted by subscapular and mayor pectoralis muscles.
exercises aimed at stretching the shortened muscle chains (Figs. 7 and 8). This protocol was supervised by Physiotherapist 2 (PT2). Appendix 2 details the experimental intervention.

Each physical therapist had more than 2 yrs of experience in this type of treatment. Moreover, both therapists participated in developing the experimental protocol in these patients, and they had been working together for more than 4 yrs with patients suffering from AS. Therefore, the ability to motivate patients to adhere to each protocol was similar in both of them.

Outcome Measures

Each patient’s evolution was closely followed, with two checks made during the study: one check at the beginning (pretreatment) and a second check at the end of the 15 sessions (posttreatment). Checks during the trial were assessed by a third assessor who was blinded to the intervention group. This assessor was a physician with many years of experience with patients suffering from AS. He was trained during 1 month to assess correctly the different outcomes. Outcome measures were assessed without any previous warm-up and before any kind of exercise. Changes in activity, mobility and functional capacity were evaluated using the following previously validated indexes from the Bath group:

- BASMI\textsuperscript{17} (The Bath Ankylosing Spondylitis Metrology Index). The BASMI consists of five clinical measures used to assess the status of the axial skeleton: tragus to wall distance,\textsuperscript{18,19} the modified Schober test,\textsuperscript{20,21} cervical rotation,\textsuperscript{20,22} lumbar side flexion, and intermalleolar distance. A cervical goniometric device manufactured by Performance Attainment Associates (St. Paul, MN) was employed for active cervical rotation assessment. Both sides were measured and the mean of the obtained values was calculated.\textsuperscript{23} Psychometric properties of this index are good interobserver reliability on each clinical measure ($r = 0.99; P < 0.001$), and criterion validity (total metrology score) of $r = 0.992, P < 0.001$. Original authors of this index reported that these clinical measures could be analyzed independently. Jones et al subsequently established cut points on a 0–10 scale to assess a total score of the BASMI index.\textsuperscript{24} In the present study, clinical measures were assessed independently. Patients repeated all movements three times, and the mean of these values was employed in the analysis.

- BASFI\textsuperscript{25} (The Bath Ankylosing Spondylitis Functional Index). The BASFI consists of ten questions referring to the functional capacity of the patient with AS to perform the daily activities (see Appendix 3). All items are valued with a 10-cm horizontal visual analogue scale. The score on the BASFI is obtained from the sum of all values. Higher score of the BASFI reflects greater limitation. Psychometric properties of this index are good interobserver reliability ($r = 0.87; P < 0.001$) and good criterion validity.

- BASDAI\textsuperscript{26} (The Bath Ankylosing Spondylitis Disease Activity Index). The BASDAI consists of six questions related to five symptoms during the last week: fatigue, spinal and joint pain, tenderness, and morning stiffness (see Appendix 4). All items are valued with a 10-cm horizontal visual analogue scale. The score on the BASDAI is obtained from the sum of the values from the first five questions. Higher score of the BASDAI re-
flects greater disease activity. Psychometric properties of this index are good test-retest reliability ($r = 0.93; P < 0.001$), and face and content internal validity.

Readers might usefully explore psychometric properties of these indexes elsewhere. 27

**Statistical Analysis**

Data were introduced in the SPSS package, version 11.5. Mean and standard deviation of the values were calculated for each variable. The Kolmogorov Smirnov test showed a normal distribution of the quantitative outcomes ($P > 0.05$). The intragroup data within both groups were compared with the paired $t$ test. The intergroup data between both groups at the beginning of the trial were assessed with the unpaired $t$ test. The intergroup comparison between the improvement (prepost scores) in both groups was also achieved with the unpaired $t$ test. Statistical analysis was conducted at a 95% confidence level. A $P$ value less than 0.05 was considered as statistically significant.

**RESULTS**

Two patients suffering from osteoporosis were excluded. On the other hand, three patients did not finish the study for personal reasons. Therefore, the analyzed sample size was 40 patients. The initial sample size in the control group was 22 patients. However, two did not finish the trial, so that the final group sample comprised 20 patients. In the same way, the initial sample size in the experimental group was 21 patients, but one did not finish the study, which means that 20 patients were finally included.

At the beginning of the study, there were no statistically significant differences in any outcome of the BASMI index (tragus to wall distance $P = 0.6$; modified Schöber test $P = 0.1$; cervical rotation $P = 0.1$; lumbar side flexion $P = 0.2$ and intermalleolar distance $P = 0.1$), as well as in the BASFI ($P = 0.4$) and BASDAI ($P = 0.7$) indexes between both groups. Furthermore, there were no statistically significant differences on gender ($P = 0.6$), age ($P = 0.6$), length of the disease ($P = 0.2$), or ACR classification ($P = 0.1$) between both study groups. It could therefore be assumed that they were comparable in all respects at the start of the trial. Demographic data corresponding to each group are given in Table 1. Clinical measures and total scores of BASFI and BASDAI indexes are summarized in Table 2.

---

*FIGURE 6 Stretching exercise of the posterior muscle chain in the pelvic region.*

*FIGURE 7 Stretching exercise of the posterior static chain of the body wall seated.*
The intragroup comparison (paired t test analysis) showed an improvement in all the outcomes, mobility measures of the BASMI index, as well as in BASFI and BASDAI indexes within both groups. In the control group, the improvement in tragus to wall distance ($P = 0.009$) and in lumbar side flexion ($P = 0.02$) was statistically significant. Although the rest of the outcomes also improved, they did not reach a significant level ($P > 0.05$). In the experimental group, the improvement in all the clinical measures of the BASMI index ($P < 0.01$) and in the BASFI index ($P = 0.003$) was statistically significant. However, the improvement in the activity index (BASDAI) was not significant.

Table 2 summarizes the intragroup comparison between prepost scores within both groups.

The intergroup comparison (unpaired t test analysis) between the improvement (prepost scores) in both groups showed that the experimental group obtained a greater improvement than the control group in all the clinical measures of the BASMI index, except in tragus to wall distance, and in the BASFI index. Table 3 summarizes the intergroup comparison of prepost scores between both groups.

**DISCUSSION**

**Findings**

The main purpose of previous trials was to study the effectiveness of physical therapy in the management of AS. Nevertheless, the interventions were often poorly described and the exact content of the programs remains partly unclear. To our knowledge, this is the first paper analyzing different physical therapy protocols in patients with AS.

Our results showed that the improvement obtained with the experimental intervention was greater than the improvement obtained with a conventional physical therapy intervention. Both interventions get an improvement in all the outcome measures of the BASMI index, as well as in BASFI and BASDAI indexes. Only the improvement in tragus to wall distance and in lumbar side flexion was statistically significant in the control group. On the other hand, the improvement in all clinical measures of the BASMI index and in the BASFI index was statistically significant in the experimen-
The intergroup comparison between pre-post scores between both groups showed that the experimental group obtained a greater improvement than the control group in four clinical measures of the BASMI and in the BASFI index. Despite the apparent differences between groups at the beginning of the trial, the intergroup comparison (based on the unpaired *t* test) did not reveal any significant differences at the beginning of the trial. Therefore, it might be assumed that the experimental intervention was the responsible of the greater improvement obtained in the experimental group.

Based our results, we might assume that the treatment of the shortened muscle chains (Figs. 1–4), according to the GPR method, might be more beneficial than conventional interventions in patients with AS. We have to emphasize that the difference between these interventions is the integration of the affected muscles in different muscle chains shortened in these patients. The analytic stretching of any of these muscles could be expanded to include secondary adaptive or maladaptive changes in the rest of the muscle chain, so that the analytic stretching of any gravitational muscle would be inefficient. Further studies are required to elucidate the role of the muscle chains in AS patients.

Another purpose of the present trial was to compare the results of the experimental group with the results reported by previous papers. Although previous studies have analyzed the effects of the conventional intervention received by our control group, differences among patients, number of sessions, outcome measures, scores at the beginning of the trial, and other such variables make it difficult to make this comparison. A factor that influences the results is the group session. A group session implies positive reinforcement and an increase in the patients’ self-esteem when coming by themselves and finding support from other people with the same disease. There is some evidence suggesting that spinal mobility obtains a greater improvement with a group physical therapy program than with home exercises. Baumberger mentioned mutual encouragement, reciprocal motivation, and exchange of experience as some advantages of group physical therapy intervention. Therefore, future studies evaluating the therapeutically effects of different physical therapy protocols should be performed in group sessions.

### Outcome Measures

In the present study, changes in activity, mobility, and functional capacity were evaluated using the previously validated indexes from the Bath group: the BASMI, BASFI, and BASDAI indexes. The intragroup differences in functional and clinical outcomes following intervention by group in patients with ankylosing spondylitis are shown in Table 2.

#### Table 2: Intragroup differences in functional and clinical outcomes following intervention by group in patients with ankylosing spondylitis

<table>
<thead>
<tr>
<th>Outcome Measures</th>
<th>Control Group</th>
<th>Experimental Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tragus to wall</td>
<td>6.2 (5.2) cm</td>
<td>4.3 (4.1) cm</td>
</tr>
<tr>
<td>Schöber test</td>
<td>2.7 (1.3) cm</td>
<td>2.5 (1.3) cm</td>
</tr>
<tr>
<td>Lumbar side flexion</td>
<td>5.4 (2.6) cm</td>
<td>4.8 (2.3) cm</td>
</tr>
<tr>
<td>Intermalleolar distance</td>
<td>68.1 (18.2) cm</td>
<td>46.5 (21.2) cm</td>
</tr>
<tr>
<td>Basdai</td>
<td>47 (19)</td>
<td>28.3 (10)</td>
</tr>
</tbody>
</table>

Pretreatment, posttreatment values of each variable, values are expressed by mean (standard deviation); Pre-post treatment values of each variable, values are expressed by mean (standard deviation); P values come from dependent samples Student’s *t* test analysis. NS, no significance.
The clinical measures considered in the BASMI index reflect the axial status of patients with AS. This index had demonstrated sensitivity to change across the whole disease spectrum including patients with disease of long duration (see psychometric properties of BASMI index in Methods section). Although Jones et al. established cut points on a 0–10 scale to assess a total score of the BASMI index, authors of the present trial preferred to assess independently each clinical measure. Although Heikkila et al. suggested changes in the mobility (Schöber test and tragus to wall distance), and two measures (lumbar side flexion), two at the middle level (cervical rotation, and lateral flexion as the most sensitive mobility measures. Other measures that might be used, but at a lower level of sensitivity, are cervical rotation and intermalleolar distance. Finally, they did not suggest the Schöber test, thoracolumbar flexion, or tragus to wall distance.

The BASMI index has one measure in the top level (lumbar side flexion), two at the middle level (cervical rotation and intermalleolar distance), and two not suggested (Schöber test and tragus to wall distance). Therefore, conclusions from the study of Heikkila et al. suggest changes in the mobility measures of the BASMI index. The reason to assess the BASMI index is that it is the only validated index (criterion validity and interobserver reliability determined) in the peer-reviewed literature that assess the status of the axial skeleton in patients with AS. However, that situation does not refute the necessity of revision of these clinical measures.

The BASFI index showed to be sensitive to detect changes in patients with AS. Length of time in morning stiffness was included in this index, but this item is not used to calculate the total score (see psychometric properties of BASDAI index in Methods section). In the present study, changes in the BASDAI index after both interventions were not statistically significant. This situation could be caused by the small simple size. A greater number of patients might show statistically significant differences in prepost scores within both groups.

### Further Research

The scientific literature shows different papers which have analyzed the therapeutically effects at 8 mos, 9 mos, and 3 yrs after the intervention. Therefore, some questions remain to be answered. First, how long will the improvement maintain? The improvement would probably remain longer with the experimental protocol, since better results were achieved with this intervention. But, will the effect disappear at the same rate in both groups? Will the experimental group maintain a greater proportion of clinical improvement? That is why we propose to continue the study assessing patients after 6 mos and 2 yrs after the intervention.

### Limitations

The small sample size and the absence of a control group, without any physical therapy intervention, are some of the most significant limitations of this study. We have to consider that this trial has been comparative, taking an active control
group treated with a physical therapy intervention analyzed in previous papers, 9,11,16,28,29 Obviously, type 2 errors could have happened, so it is recommended to repeat the same procedure with a greater number of patients and with a control group without any therapeutically intervention

Another topic to discuss is the possible influence of some clinical features such as age, sex, work, length of the disease, and other circumstances. Any randomized controlled trial depends on the demographic and clinical characteristics of the sample, so it is difficult to obtain similar patients in different studies performed in different countries.

Finally, we did not consider the effects of the pharmacological treatment, which we know that helped patients during the study. All patients had been taking NSAID treatment for at least 10 yrs (controlled by their rheumatologist); however, during the trial all patients were asked not to modify their pharmacological treatment. On the other hand, something we positively thought could affect the results was the swimming-pool exercise. The whole sample size (90%) had already been doing it for several years and it was included in their habitual practice. Therefore, it was not modified because the exercises were the same in all patients, so that we considered that the obtained improvement would affect both groups in the same way.

CONCLUSION

The experimental group obtained a greater improvement than the control group in all clinical measures, except in tragus to wall distance, and in the functional index (BASFI). The experimental protocol developed by our research group, based on the GPR method and specific strengthening and flexibility exercises of the shortened muscle chains, offers promising results in the management of patients suffering from AS. Further trials are required.

ACKNOWLEDGMENTS

I would like to acknowledge to all members of the Universidad Rey Juan Carlos, to the Spanish Rheumatic Diseases League for the coordination of the Ankylosing Spondylitis Associations in Spain, and to each patient who took part in this trial. I would also like to thank Philippe E. Souchard (creator of the GPR method) and Ruben Fernández for their priceless support in every moment.

REFERENCES


35. Hidding CA, Van der Linden S, Gielen X: Continuation of group physical therapy is necessary in Ankylosing Spondylitis: Results of a randomized controlled trial. *Arthritis Care Res* 1994;7 (2):90–96

**APPENDIX 1: Conventional physical therapy intervention**

<table>
<thead>
<tr>
<th>Number</th>
<th>Position of the Patient</th>
<th>Exercise</th>
<th>Repetitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Seated</td>
<td>Cervical lateral flexion</td>
<td>2 series of 10 repetitions each one</td>
</tr>
<tr>
<td>2</td>
<td>Seated</td>
<td>Cervical rotation</td>
<td>2 series of 10 repetitions each one</td>
</tr>
<tr>
<td>3</td>
<td>Seated</td>
<td>Cervical flexo-extension</td>
<td>2 series of 10 repetitions each one</td>
</tr>
<tr>
<td>4</td>
<td>Seated</td>
<td>Thoracic rotation</td>
<td>2 series of 10 repetitions each one</td>
</tr>
<tr>
<td>5</td>
<td>Standing</td>
<td>Thoracic lateral-flexion</td>
<td>2 series of 10 repetitions each one</td>
</tr>
<tr>
<td>6</td>
<td>Standing</td>
<td>Thoracic flexo-extension</td>
<td>2 series of 10 repetitions each one</td>
</tr>
<tr>
<td>7</td>
<td>Standing</td>
<td>Thoracic muscles stretching</td>
<td>2 repetitions of 45 second each one</td>
</tr>
<tr>
<td>8</td>
<td>Standing</td>
<td>Hamstring muscles stretching</td>
<td>2 repetitions of 45 second each one</td>
</tr>
<tr>
<td>9</td>
<td>Standing</td>
<td>Gastrocnemius muscle stretching</td>
<td>2 repetitions of 45 second each one</td>
</tr>
<tr>
<td>10</td>
<td>Standing</td>
<td>Strengthening of quadriceps muscle</td>
<td>2 series of 8 repetitions each one</td>
</tr>
<tr>
<td>11</td>
<td>Kneeling</td>
<td>Psoas muscle stretching</td>
<td>2 repetitions of 45 second each one</td>
</tr>
<tr>
<td>12</td>
<td>Lying supine</td>
<td>Posterior pelvic girdle gliding</td>
<td>2 series of 8 repetitions each one</td>
</tr>
<tr>
<td>13</td>
<td>Lying supine</td>
<td>Active flexion of the upper cervical spine</td>
<td>2 series of 8 repetitions each one</td>
</tr>
<tr>
<td>14</td>
<td>Lying supine</td>
<td>Superior abdominal strengthening</td>
<td>2 series of 10 repetitions each one</td>
</tr>
<tr>
<td>15</td>
<td>Lying supine</td>
<td>Inferior abdominal strengthening</td>
<td>2 series of 10 repetitions each one</td>
</tr>
<tr>
<td>16</td>
<td>Lying supine</td>
<td>Lumbar spine rotation</td>
<td>2 series of 8 repetitions each one</td>
</tr>
<tr>
<td>17</td>
<td>Lying on the side</td>
<td>Coxofemoral abduction</td>
<td>2 series of 10 repetitions each one</td>
</tr>
<tr>
<td>18</td>
<td>Lying on the side</td>
<td>Shoulder abduction</td>
<td>2 series of 10 repetitions each one</td>
</tr>
<tr>
<td>19</td>
<td>Kneeling—hand position</td>
<td>Anteroposterior pelvic girdle gliding</td>
<td>2 series of 10 repetitions each one</td>
</tr>
<tr>
<td>20</td>
<td>Kneeling—hand position</td>
<td>Anteroposterior lumbar and thoracic gliding</td>
<td>2 series of 10 repetitions each one</td>
</tr>
</tbody>
</table>
APPENDIX 2: Experimental physical therapy intervention

In this appendix we will expose the scheme of the experimental intervention (divided into six phases) and the exercises to stretch the specific muscle chains. More details about these exercises may be found elsewhere.\(^{2,13}\) (*) Kleinrensink GJ, Stoeckart R, Mulder PG, et al: Upper limb tension tests as tools in the diagnosis of nerve and plexus lesions. Anatomical and biomechanical aspects. *Clin Biomech* 2000;15 (1):9–14.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Purpose of Each Phase</th>
<th>Exercises</th>
<th>Repetitions</th>
</tr>
</thead>
</table>
| 1. General warm-up | All exercises in this phase will be performed standing and/or walking | 1. Stretching exercise of the posterior muscle chain.  
2. Stretching exercise of the anterior muscle chain.  
3. Neural mobilization of the median nerve (*). | 2 series of 8 repetitions each one |
| 2. Specific warm-up | Exercises in this phase are focussed on improving the pelvic girdle mobility | 1. Anteroposterior pelvic girdle gliding.  
2. Extension-flexion motion of the lumbar spine (McKenzie method).  
4. Stretching exercise of the posterior muscle chain in the pelvic region (Fig. 6). | 2 series of 8 repetitions each one |
| 3. Dynamic axial exercise | Exercises in this phase will be performed lying supine and prone | 1. Prone exercises. Anterior pelvic girdle gliding.  
2. Anteroposterior pelvic girdle gliding in supine.  
3. Rotation stretching of the posterior muscle chain. | 2 series of 10 repetitions each one |
| 4. Static postural exercise | Exercises in this phase are focused on stretching and strengthening the shortened muscle chains | 1. Stretching exercise of the anterior muscle chain in supine (Fig. 8a).  
2. Stretching exercise of the posterior muscle chain seated (Fig. 8b).  
3. Stretching exercise of the posterior muscle chain seated on the wall (Fig. 7).  
4. Stretching exercise of the anterior muscle chain standing (Fig. 8c).  
5. Eccentric work of the erector spine muscles (Fig. 5). | All stretching postures have to be maintained during 3–4 mins each one |
| 5. Specific respiratory exercises | All respiratory exercises will be performed in a stretching posture during phase 4 | 1. Thoracic breathless.  
2. Expiratory breathless.  
3. Stretching of the anterointernal muscle chain of the scapular girdle. | 2 series of 10 repetitions each one |
| 6. Cool-down | This phase will consist on slightly neck and thoracic exercises. All exercises will be performed walking | 1. Cervical flexo-extension.  
2. Cervical lateral-flexion.  
4. Circular motion of the scapular girdle. | 1 series of 5 repetitions each one |
**APPENDIX 3: The Bath Ankylosing Spondylitis Functional Index: The BASFI index**

Please draw a mark on each line below to indicate your level of ability with each of the following activities during the last week.

- Putting on your socks or tights without help or aids (e.g., sock aid).
  - Easy ___ Impossible ___
- Bending forward from the waist to pick up a pen from the floor without an aid.
  - Easy ___ Impossible ___
- Reaching up to a high shelf without help or aids (e.g., helping hand).
  - Easy ___ Impossible ___
- Getting up out of an armless dining room chair without using your hands or any other help.
  - Easy ___ Impossible ___
- Getting up off the floor without help from lying in your back.
  - Easy ___ Impossible ___
- Standing unsupported for 10 mins without discomfort.
  - Easy ___ Impossible ___
- Climbing 12–15 steps without using a handrail or walking. One foot on each step.
  - Easy ___ Impossible ___
- Looking over your shoulder without turning your body.
  - Easy ___ Impossible ___
- Doing physically demanding activities (e.g., physiotherapy exercises, gardening, or sports).
  - Easy ___ Impossible ___
- Doing a full day of activities, whether at home or at work.
  - Easy ___ Impossible ___

**APPENDIX 4: The Bath Ankylosing Spondylitis Disease Activity Index: The BASDAI index**

Please place a mark on each line below to indicate your answer to each question, relating to the past week.

- How would you describe the overall level of fatigue/tiredness you have experienced?
  - ___ None ___ Very severe
- How would you describe the overall level of AS neck, back, or hip pain you have had?
  - ___ None ___ Very severe
- How would you describe the overall level of pain/swelling you have had in joints other than the neck, back, or hips?
  - ___ None ___ Very severe
- How would you describe the overall level of discomfort you have had from any areas tender to touch or pressure?
  - ___ None ___ Very severe
- How would you describe the overall level of morning stiffness you have had from the time you wake up?
  - ___ None ___ Very severe
- How long does your morning stiffness last from the time you wake up?
  - 0 hrs ___ 1/2 ___ 1 hr ___ 1 1/2 ___ 2 or more hrs

Item 6 is not employed to assess the total score of the BASDAI index.
Authors:
David Green, MD, PhD
Susan Sullivan, RN
Janet Simpson, RN
Robert C. Soltysik, MS
Paul R. Yarnold, PhD

Affiliations:
From the Departments of Medicine and Physical Medicine and Rehabilitation, Feinberg School of Medicine, Northwestern University, the Rehabilitation Institute of Chicago, Chicago, Illinois.

Correspondence:
All correspondence and requests for reprints should be addressed to David Green, MD, PhD, 676 North St. Clair Street, Suite 850, Chicago, IL 60611. Supported, in part, by grant H133G990046 from the National Institute on Disability and Rehabilitation Research.

0894-9115/05/0-0420/0
American Journal of Physical Medicine & Rehabilitation
Copyright © 2005 by Lippincott Williams & Wilkins
DOI: 10.1097/01.phm.0000163714.73660.70

Evolving Risk for Thromboembolism in Spinal Cord Injury (SPIRATE Study)

ABSTRACT


Design: A comprehensive review of the charts of all patients admitted between late 1999 and early 2003 for rehabilitation after spinal cord injury. Only records, including evidence for objective testing for VTE (ultrasound, venography, lung scanning) were included, and patients having inferior vena cava filter placement or previous VTE were excluded. Analysis variables included type and location of spinal cord injury, American Spinal Injury Association classification, concomitant injuries, surgical procedures, complications, preexisting illnesses, and use of antithrombotic prophylaxis. Using univariate optimal discriminant analysis, data from the current group of patients were compared with a previous study of 243 subjects examined between 1992 and 1995.

Results: The current sample consisted of 76 persons with acute spinal cord injury, of whom six had VTE (7.9%). As compared with the frequency of VTE in the previous patient sample (21%), this represented a significant decrease ($P < 0.01$). The major differences between the current and previous patient samples were a decrease in the use of unfractionated heparin (15.8% vs. 56.8%, $P < 0.0001$) and an increase in the use of low molecular weight heparin (81.6% vs. 59.7%, $P < 0.0001$).

Conclusions: VTE has been a common and occasionally lethal complication in persons with spinal cord injury. The recent switch from unfractionated heparin to low molecular weight heparin for the prevention of VTE has coincided with a decrease in the frequency of this complication in patients with spinal cord injury.

Key Words: Thromboembolism, Spinal Cord Injury, Risk Factors, Complications
We previously reported an analysis of risk factors for thromboembolism in persons with acute spinal cord injury who were inpatients on the spinal cord injury (SCI) unit at the Rehabilitation Institute from September 1992 through December 1995. We found that the frequency of venous thromboembolism (VTE) was 21% and that older patients with concomitant cancer were at high risk, as were women between the ages of 36 and 58 yrs and men with flaccid paralysis. The patients in this earlier study generally received either unfractionated heparin or low molecular weight heparin (LMWH) for prophylaxis against thrombosis.

However, since 1995, there has been increasing use of LMWHs, both in the acute phase and in the rehabilitation phase of SCI. The recently published Consensus Conference of the American College of Chest Physicians recommends prophylaxis with LMWH in acute SCI, begun once primary hemostasis is evident and continuing on into the rehabilitation phase. LMWH began to supplant unfractionated heparin in 1999 at our institution. Therefore, we examined patients admitted to our unit between 1999 and 2003 to determine whether VTE had become less frequent than recorded during the earlier period. We noted the numbers of patients that had received unfractionated heparin or LMWH and collected data on a variety of possible risk factors for thrombosis.

METHODS

Subjects and Data Collection

The study was reviewed and approved by the Office for the Protection of Research Subjects of Northwestern University. The records of all admissions to the SCI unit of the Rehabilitation Institute during the period from November 1999 through February 2003 were retrieved. Patients having VTE during their acute hospital stay, or placement of a vena cava filter, were excluded. The remaining charts were reviewed for evidence that an objective assessment for deep vein thrombosis or pulmonary embolism had been performed at some point during the patient’s hospitalization. Most patients had a venous ultrasound examination performed during their rehabilitation stay, usually to determine whether they had developed asymptomatic deep vein thrombosis during their acute care hospitalization or, less commonly, to investigate symptoms such as leg pain or swelling. A positive study reported the presence of a noncompressible venous segment or visualized a thrombus in either a proximal or distal deep vein.

All charts of patients having objective assessments for VTE were then abstracted. For each subject, we recorded demographic information, the location and completeness of the SCI, the American Spinal Injury Association classification, and a description of concomitant injuries. In addition, hospital length of stay, concomitant disorders, complications, and all medical and surgical treatments were noted. In particular, the choice of thromboprophylaxis was recorded. At the time of admission, orders were written for either unfractionated heparin or LMWH, which was continued for the entire inpatient rehabilitation stay. Exceptions to this practice were for patients already receiving warfarin (six patients) or those in whom anticoagulants were contraindicated because of recent intracranial, gastrointestinal, or genitourinary bleeding, who were fitted with compression boots (four patients).

Statistical Analysis

Samples were compared on individual attributes using univariate optimal discriminant analysis. The use of LMWH vs. unfractionated heparin within the current sample was assessed using a z test for correlated proportions.

RESULTS

There were 329 admissions during the 40-mo period; 272 were traumatic SCI. Many of these patients had vena cava filter placement or VTE during their acute hospital stay, or lacked evidence of objective testing for VTE, and therefore were not included. The final study cohort consisted of 76 subjects. Table 1 shows a comparison of the characteristics of this current group with the patients reported previously from our institution (previous group). It can be seen that the demographics of the current group are similar to the previous group, but a larger percentage had cervical injuries and spinal fusion. Most importantly, in the current group of patients, the use of LMWH greatly exceeded that of unfractionated heparin (81.6% vs. 15.8%, P < 0.0001).

Only 6 of 76 patients (7.9%) in the current group had VTE (five deep vein thromboses, one pulmonary embolism), compared with 51 of the 243 patients (21%) in the previous group (P < 0.01). In the current group, only the presence of rib fractures showed statistically marginal association with VTE (P = 0.07). Other characteristics, such as age, sex, spinal surgery, completeness of motor paralysis, and flaccid paralysis, did not demonstrate significant associations with VTE.

DISCUSSION

The most important finding to emerge from our study of current as compared with previous patients with spinal cord injury is a marked decline in the frequency of VTE. This was not because the current group of patients had less severe spinal
cord injuries, fewer cervical spine fractures, differential percentage of patients classified as American Spinal Injury Association classification A, or number of patients requiring spinal fusion; these were comparable between samples. Also, other risk factors for VTE, such as age, rib and long bone fractures, and concomitant disorders,4,7 were often coexistent. Rather, the most striking difference between the two samples was the replacement of unfractionated heparin in the current group with LMWH.

As compared with unfractionated heparin, LMWH is better absorbed from subcutaneous sites, has a longer half-life, and demonstrates less binding to cells and plasma proteins, making more free drug available for anticoagulation.8 Thus, from a consideration of the pharmacologic properties of the two drugs, it is biologically plausible that LMWH would provide improved thromboprotection as compared with unfractionated heparin. This supposition is confirmed by clinical trials, which demonstrate that patients assigned to LMWH have fewer thromboembolic events with no increase in bleeding.2,4,7,9

Our study has some important limitations. Because fewer patients meeting the inclusion criteria of this study were admitted to our spinal cord unit during the more recent period, the sample size was smaller, fewer thromboses were observed, and the assessment of possible risk factors was more uncertain. This may explain why we were unable to confirm the associations noted in our earlier study. Furthermore, an observational study such as this cannot prove a causal relationship between the decline in the frequency of VTE and the increasing use of LMWH, although the evidence cited above supports this hypothesis. In summary, it seems that the switch from unfractionated heparin to LMWH is now paying dividends in terms of fewer thromboembolic complications for patients with spinal cord injuries.

**ACKNOWLEDGMENTS**

We thank David Chen, MD, and Diane Hartwig for assistance in patient location.

**REFERENCES**


<table>
<thead>
<tr>
<th>TABLE 1 Characteristics of previous group and current group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristic</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Age in yrs, mean (±SD)</td>
</tr>
<tr>
<td>Length of stay in days, mean (±SD)</td>
</tr>
<tr>
<td>Female sex, %</td>
</tr>
<tr>
<td>Spinal fusion, %</td>
</tr>
<tr>
<td>Cervical injuries, %</td>
</tr>
<tr>
<td>ASIA-A, %</td>
</tr>
<tr>
<td>Rib fracture, %</td>
</tr>
<tr>
<td>Long bone fracture, %</td>
</tr>
<tr>
<td>Cancer, %</td>
</tr>
<tr>
<td>Unfractionated heparin, %</td>
</tr>
<tr>
<td>Low molecular weight heparin, %</td>
</tr>
</tbody>
</table>

ASIA-A, American Spinal Injury Association classification A.

*Sum is >100% because some patients were switched from unfractionated heparin to low molecular weight heparin when this became available.
Efficacy of Computer-Aided Dosing of Warfarin Among Patients in a Rehabilitation Hospital

ABSTRACT


Objective: To determine whether computer-aided dosing of warfarin is superior to physician dosing to maintain a patient in a rehabilitation hospital within a target international normalized ratio goal.

Design: Randomized, double-blinded, clinical trial in an inpatient rehabilitation hospital. A total of 30 consecutive patients admitted receiving warfarin were randomized to either clinician dosing or computer-aided warfarin dosing for the duration of their hospitalization. The main outcome measures included the percentage of days in a therapeutic anticoagulation range and the number of blood draws. Exclusion criteria included short length of stay (n = 110, 39%) and a physician declared international normalized ratio target range of <2.0 (n = 67, 23%). A total of 73 patients were excluded because of heme-positive stools at admission, recent gastrointestinal bleed, early discharge or consent refusal. Dawn AC software was used to determine warfarin dosage and frequency of blood draws to maintain a target international normalized ratio of 2.0–3.0 for the computer-dosed group (n = 14). Several physicians recommended warfarin dosages for the second group (n = 16). Two were dropped from the computer model secondary to lost data files for these two patients.

Results: Computer-aided dosing of warfarin resulted in 61.7% of days within the therapeutic range (international normalized ratio, 2–3), whereas clinician dosing resulted in only 44.1%. There were no significant differences in the number of blood draws or demographic variables between the two groups.

Conclusion: Computers were significantly better at maintaining patients within a therapeutic international normalized ratio range than physicians. There were no significant differences in the number of recommended blood draws.

Key Words: Anticoagulants, Warfarin, Computer-Assisted Drug Therapy, Deep Vein Thrombosis, Rehabilitation
Nonambulatory hospitalized patients continue to be at great risk for developing deep vein thrombosis (DVT) during their hospitalization. As the risk of DVT is increased among those who have been immobilized, have paralysis, have had a recent surgery, who have had a cardiovascular accident or a myocardial infarction, or who are >40 yrs of age, many patients admitted to a rehabilitation hospital have an increased risk for DVT or pulmonary embolism. For those with a documented DVT or pulmonary embolism, anticoagulation is usually recommended, with the most accepted international normalized ratio (INR) range of 2–3.1 The long-term anticoagulation drug of choice continues to be warfarin. Current accepted guidelines for anticoagulation therapy suggest an INR range of 2.0–3.0 for most patients with at least one risk factor for DVT.2,3

Despite clear guidelines, physicians continue to underdose and even defer anticoagulation treatment in patients who warrant anticoagulation.4,5 Physicians continue to be concerned about warfarin-induced complications such as gastrointestinal bleeds.5 Some have found that, despite close monitoring, hospitalized patients are outside of a therapeutic INR range for >60% of the time.5 Given these findings, researchers have attempted to explore more effective means to ensure the accurate prophylaxis of patients at risk.

A recent study examining the use of computer-aided dosing for hospitalized acute care patients found computer programs to be significantly better than physicians at the initiation of warfarin treatment; specifically, patients dosed by computer programs reached a therapeutic INR in a significantly shorter period of time than patients dosed by physicians.7 Another trial found that computer dosing was superior to manual dosing of outpatients who had recently undergone total hip arthroplasty.8 Studies involving ambulatory patients have found computer-aided dosing to be superior at maintaining a therapeutic INR in the outpatient population.9–11 Patients of resident physicians who used computerized support to aid in dosing of warfarin reached therapeutic levels more quickly12 and more consistently than patients whose physicians did not use such models.13 Additionally, computerized dosing has been shown to decrease the burden on outpatient anticoagulation clinics14 and on medical staff time.15 One study found that computers were most effective at maintaining outpatients within the higher INR ranges;16 another study has shown that in various settings, physicians tend to generally underdose with warfarin.6 To the investigators’ knowledge, there have been no previous studies that address the utility of computer-aided dosing for both the introduction and maintenance of warfarin within a therapeutic INR range over the course of a rehabilitation hospitalization.

The goal of the current study was to determine whether computer-aided dosing would be significantly better than physician dosing at maintaining hospitalized rehabilitation patients within a therapeutic INR. The investigators further attempted to determine whether the use of a computer-aided system could reduce the number of blood draws needed to monitor a patient’s anticoagulation blood values.

**METHODS**

Approval for this study was obtained from the institutional review board. Informed consent was obtained from all participants in accordance with the institutional review board. A total of 280 consecutively admitted patients who were prescribed warfarin for anticoagulation were identified in a 1-yr period. All patients were admitted to a free-standing, 288-bed, academic rehabilitation center. Exclusion criteria included a length of stay of <2 wks (39%, n = 110), a physician-requested INR target range of <2.0 (23%, n = 67), and postadmission complications (heme-positive stools, gastrointestinal bleed, early discharge; 25%, n = 71) that required an early termination of warfarin dosing. A total of 30 eligible patients agreed to participate. Subjects were randomized using a random-number table into one of two groups: group P (physician dosing) or group C (computer dosing). Patients placed in group C (n = 14) were cared for by physicians who received instructions for warfarin dosing and for timing and frequency of blood draws from a computer-generated program, Dawn AC (4S Information Systems, London, UK). The Dawn AC program was chosen because it had been used with success in previous trials.7 Patients in group P (n = 16) were dosed by physicians who selected the timing of blood draws and dosing of warfarin based on their clinical judgment and without the aid of computer-generated instruction. The goal of both groups was to maintain patients within a target INR of 2.0–3.0.

Data for group C were collected by trained study physicians. Data points were verified by trained researchers to ensure the accuracy of the data points.

**STATISTICAL ANALYSIS**

Linear interpolation of missing INR scores (days when there was no blood drawn) was used. This method has been previously shown to be a reasonably accurate means of bridging data points (INR values) between draws.17 A Splus program was written to impute missing INR scores by linearly interpolating between the previous and next existing INR scores. Missing INR scores that were
on the boundaries of a patient’s data (the first or last day INR values) were excluded because there was not enough information to impute. A two-sample t test of proportions with continuity correction was done to determine whether computer-dosed patients fared better than physician-dosed patients. The Splus code is available on request.

Each day, for each patient, was recorded as a success or failure depending on whether the patient’s INR score was within the target range. A two-sample test of proportions with continuity correction was used for patients in groups P and C. To determine whether the number of ordered blood draws differed between the two groups, a two-sample t test was performed because the normality assumption for the t test was borderline and the sample was rather small (n = 30).

Logistic regression was used to test if the other measured variables (age, sex, and primary diagnosis) were confounding the relationship between the study group and the proportion of days in the optimal INR range. A type I error rate of 0.05 was used for all tests.

RESULTS

The total number of data points (INR values) was 1014, excluding 36 days during which the INR score could not be imputed. During the data review process, two patients from the computer-dosed groups were identified as having missing data sheets. As these data points could not be retrieved, these patients were dropped from the study. The original group of 32 was therefore trimmed to 30 through this loss of data sheets. A total of 1014 data points for 30 patients were analyzed.

Baseline patient characteristics can be found in Table 1. Age and sex were similar in groups C and P. The mean length of stay was slightly longer in group C than in group P (38.7 vs. 31.7 days, respectively). The medical indication for the initiation of anticoagulation is represented in Table 2. A larger percentage of patients in group C had a history of atrial fibrillation (28% vs. 6%), and group P had a larger proportion of patients with a history of DVT (50% vs. 7%). Baseline patient characteristics (age, sex, primary diagnosis) did not significantly affect the relationship between the dosing and the proportion of days in the optimal INR range.

A significant difference was identified when analyzing the proportion of days in the optimal INR range between the two groups. The proportion of time that patients were in therapeutic the INR range was significantly longer in group C than in group P (61.7% vs. 44.1%, respectively, P < 0.05, 95% confidence interval, 0.113–0.238) (Fig. 1).

There was no significant difference in the number of blood draws between the two groups (P = 0.170) (Fig. 2). There was no incident DVT or pulmonary embolism recorded for any of our subjects during the course of the study.

DISCUSSION

This study compares the INR values of two methods for dosing of warfarin. One reflects the standard of care, that of physicians who routinely review the results of the most recent laboratory values and decide, based on experience and the knowledge of warfarin use, the next appropriate warfarin dose and dosing interval. Using a prospective, randomized method, this study demonstrates that the computerized dosing method may provide superior results than can be obtained by experienced clinicians. As there are several studies that demonstrate that inpatients are often underdosed, there is a very real concern that the patients often do not receive the intended prophylaxis. Previous studies seem to suggest that there is a tendency for physicians to underdose, thus increasing the risk of incomplete anticoagulation. This study demonstrates that patients are more likely to be within

**TABLE 1** Patients’ characteristics at baseline (n = 30)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group C (Computer Dosed)</th>
<th>Group P (Physician Dosed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total, % (n)</td>
<td>47 (14)</td>
<td>53 (16)</td>
</tr>
<tr>
<td>Sex, % (n)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>43 (6)</td>
<td>50 (8)</td>
</tr>
<tr>
<td>Female</td>
<td>57 (8)</td>
<td>50 (8)</td>
</tr>
<tr>
<td>Mean age, yrs (SD)</td>
<td>66.0 (15.5)</td>
<td>72.9 (10.8)</td>
</tr>
<tr>
<td>Mean length of stay, days (SD)</td>
<td>38.7 (15.6)</td>
<td>31.7 (16.5)</td>
</tr>
</tbody>
</table>

**TABLE 2** Primary anticoagulation indication (n = 30)

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Group C (Computer Dosed)% (n)</th>
<th>Group P (Physician Dosed)% (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep vein thrombosis</td>
<td>7 (1)</td>
<td>50 (8)</td>
</tr>
<tr>
<td>History of valvular pathology</td>
<td>14 (2)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>28 (4)</td>
<td>6 (1)</td>
</tr>
<tr>
<td>Stroke</td>
<td>36 (5)</td>
<td>38 (6)</td>
</tr>
<tr>
<td>Other</td>
<td>14 (2)</td>
<td>6 (1)</td>
</tr>
</tbody>
</table>
the therapeutic range if dosing is assisted by computer. We speculate that previously published studies that have shown physicians to undertreat their patients do so due to fear that patients will experience complications associated with overdosing, for example, gastrointestinal bleeds. This tendency to underdose might be easily solved through an increase in the average warfarin dose in such a way as to correct the deficiency without overcompensation. Significantly, our data do not demonstrate that any of the patients were beyond an INR of 4, and patients were noted to be in the 3–4 range only 18% of the time. There were no adverse events noted for any patient. This study therefore suggests that improved efficacy in therapeutic warfarin dosing due to computer-assisted technology does not come at the expense of overdosing. When frequency of blood draws were compared between groups, no significant difference was found. Computer-assisted warfarin dosing did not require more frequent blood draws, an important factor that affects both cost control and patient comfort.

A number of tools have been used to improve the ability to maintain warfarin-induced anticoagulation at appropriate therapeutic values. These have included computers, nomograms, and flexible protocols. In general, it seems that these tools have allowed for the reduction of the number of measured INR values that are above the therapeutic range, shorten the time to achieve a steady state, or reduce the number of dose adjustments that are necessary. In a multicenter, randomized study of computerized anticoagulant dosing, Poller et al. reviewed 285 patients randomized to a computer-generated dosing group or a traditional dosing group. In this study, the computer-generated dosing was better for achieving a target INR, with a mean time within INR better than the traditional dosing group. Stabilization, however, was not significantly different between the two groups. In another study by Wilson and James, computer-assisted management of warfarin treatment in an outpatient clinic seemed to be more efficient and resulted in no difference between groups in achieving appropriate anticoagulation. In another outpatient study, 688 patients were reviewed with a computerized dosing program. An improvement of 38% was achieved among these patients in realiz-

---

FIGURE 1 Percentage of time within therapeutic international normalized ratio (INR) range.

FIGURE 2 Number of blood draws in both the physician-dosed group and the computer-dosed group. There was no significant difference between the number of blood draws for each group (P = 0.170).
ing INR within the recommended therapeutic range.\textsuperscript{11} For comparison, using three computerized systems for anticoagulation monitoring, Poller et al.\textsuperscript{16} randomized patients to one of the computerized groups or to medical staff treatment. The computerized models seemed to be better at assigned dosing when the target INR was 3–4.5. Notably, there was a reduction in the number of times that the patients were undertreated. Comparisons between nomograms and physician-adjusted warfarin therefore have suggested possible advances, but they have not led to any generally agreed upon software or protocol for the initiation and maintenance of anticoagulation with warfarin.

The findings of our study have important implications for lowering both morbidity and mortality among patients receiving anticoagulation. With better dosing of warfarin, uncomfortable and potentially fatal events such as DVT or pulmonary embolism may be significantly reduced.

As this study was limited by its power, future studies are indicated that have a larger sample size. There was likely a great deal of variability between the dosing habits of different physicians, a factor that was not controlled for in this study. Also, there were more patients with DVTs in group P, and group C had more patients with nonrheumatic atrial fibrillation. These factors were adjusted for by logistic regression.

This study demonstrates that the utilization of a computer to assist with warfarin dosing is superior to physician dosing for the maintenance of patients within a therapeutic INR. As thrombembolism continues to remain a significant cause of morbidity and mortality in hospitalized patients, these data may provide beneficial information for institutions that wish to optimize the treatment of patients requiring anticoagulation. In this study, we review the effect on patients who were taking Coumadin for full anticoagulation. There is no reason to believe that these result might not also apply to different INR target ranges.

REFERENCES
Training of Somatosensory Discrimination After Stroke
Facilitation of Stimulus Generalization

ABSTRACT


Objective: Task-specific learning typifies perceptual training but limits rehabilitation of sensory deficit after stroke. We therefore investigated spontaneous and procedurally facilitated transfer of training effects within the somatosensory domain after stroke.

Design: Ten single-case, multiple-baseline experiments were conducted with stroke participants who had impaired discrimination of touch or limb-position sense. Each experiment comprised three phases: baseline, stimulus-specific training of the primary discrimination stimulus, and either stimulus-specific training of the transfer stimulus or stimulus-generalization training. Both the trained and transfer stimuli were monitored throughout using quantitative, norm-referenced measures. Data were analyzed using individual time-series analysis and meta-analysis of intervention effects across case experiments.

Results: Stimulus-specific training was successful for trained texture and proprioceptive discriminations, but it failed to show spontaneous transfer to related untrained stimuli in the same modality in seven of eight experiments in which this was possible. In contrast, intramodality transfer was obtained with stimulus-generalization training in four of five experiments that investigated stimulus-generalization training of texture discrimination. Findings were confirmed by meta-analysis.

Conclusions: Our findings demonstrate generalization of training within a somatosensory modality poststroke, provided that a program designed to enhance transfer is used. This has implications for the design of efficient rehabilitation programs.

Key Words: Sensation, Rehabilitation, Cerebrovascular Accident, Transfer of Training
Generalization of intervention effects to untrained tasks is central to sensorimotor neurorehabilitation. Task-specific training, although effective, has the potential to be very costly, and novel tasks will present continuing problems. Discovery of effective training methods also able to achieve transfer of gains to novel tasks is thus essential. Furthermore, investigation of the boundaries of transfer is increasingly regarded as a key ingredient to a proper understanding of learning effects in the sensorimotor system. In the sensory domain, investigation of stimulus-generalization effects could help clarify what is encoded about the stimuli during the training experience.

"Generalization of training" or "transfer of learning" to an unpracticed task has been investigated with healthy subjects in various domains, including perception. Empirical investigations in the perceptual learning literature suggest that transfer is possible in some instances. However, discrimination training is often highly specific to the task. A general finding is that similarity between the trained and transfer tasks is an important determinant of positive transfer.

The dimensions of similarity that seem to matter include whether the novel transfer task is within the same perceptual dimension as the trained task, the similarity of the required response, and the similarity in method of processing the information. Body location, density of receptors, and organization of the somatosensory cortex may also affect transfer of learning in somatosensory discrimination. The training conditions under which transfer might be expected also need to be considered. Principles that may enhance transfer of training have been identified in the motor-learning literature. However, exposure and feedback may be adequate for perceptual transfer, at least for unimpaired subjects.

Generalization of training within the same perceptual dimension or sensory modality has received only limited investigation in stroke patients. Spontaneous improvement in related visual functions was found in a controlled study that trained a single visual function of patients with cerebral blindness. In contrast, we found stimulus-specific training effects across tactile and proprioceptive discrimination tasks poststroke. Studies that have trained multiple somatosensory modalities simultaneously, used objects with multisensory demands, or employed sensorimotor tasks have reported improvement in untrained sensory tasks, suggesting the potential for generalized somatosensory effects. These results are consistent with findings of spontaneous transfer across multidimensional tasks in healthy subjects. Although spontaneous transfer may be achieved by unimpaired subjects when tasks and processing activities are highly similar, it is unknown whether these conditions are adequate for individuals in whom the perceptual system is damaged. Furthermore, the potential for transfer of training under different training conditions has not been systematically investigated within the somatosensory domain after stroke.

We have focused our research on rehabilitation of somatosensory deficits, particularly touch and proprioceptive discrimination. These deficits are characteristic of the loss experienced and are reported in approximately 50% of stroke patients. Such deficits pose substantial difficulties in reception of sensory information and exploration of the environment and have detrimental effects on spontaneous use of hands, object manipulation, and precision grip. Moreover, sensory loss has a negative effect on personal safety, functional outcome, and quality of life, and it influences length of stay in the hospital (see Carey for review), highlighting the functional importance of training these capacities.

This investigation of generalization effects followed our initial findings that stroke-induced impairments of texture discrimination and limb-position senses are trainable. Improvements were specific to the stimuli trained and confirmed the expected independence of tactile and proprioceptive performance. Although lack of transfer across tactile and proprioceptive discrimination was expected, those data do not address transfer across stimuli of the same type, such as textures with different distinctive features of roughness. Efficient retraining of somatosensory discrimination abilities requires an understanding of the similarity required to achieve successful transfer.

We therefore conducted a series of ten controlled, multiple-baseline, single-case experiments to investigate transfer across stimuli within the same sensory-perceptual dimension using either stimulus-specific training (SST) or stimulus-generalization training (SGT). The design permitted training the primary discrimination stimuli while simultaneously monitoring the influence of this treatment on the transfer stimuli, and the design delayed introduction of generalization-enhanced training to after this control phase. Study 1 investigated spontaneous transfer of training effects within tactile and proprioceptive domains after SST. Study 2 investigated transfer to novel stimuli within the tactile dimension using a program modified to facilitate SGT.

Transfer tasks were selected to require the same sensory-perceptual dimension and body location as the trained task, as these features of similarity seem to be important in perceptual learning. Tactile stimuli were two types of textured
surfaces: plastic grids and fabrics. These stimuli require processing within the same tactile domain and are both explored with the fingertip but have different surface characteristics. The proprioception tasks involved discrimination of imposed wrist positions in the flexion-extension and ulnar-radial deviation planes of movement. They required the same type of limb-position discrimination and the same body location (i.e., the wrist), but they used positions within different planes of movement. Norm-referenced, quantitative, and reliable measures of these tactile and proprioceptive discriminations were available as outcome measures.12,14,15

MATERIALS AND METHODS

Study 1: Spontaneous Transfer of Training Effects with SST

Subjects

Five stroke patients with tactile or proprioceptive discrimination impairment, as tested by the measures described below, were studied. They were medically stable, had adequate comprehension of instructions for assessment, had no peripheral neuropathy or previous central nervous system dysfunction, and were free of unilateral spatial neglect based on clinical observation and standard neuropsychological tests. Patients meeting these criteria were selected sequentially, as they presented, and gave voluntary informed consent. The project was approved by the human ethics committees of LaTrobe University and participating hospitals and conformed to the Helsinki Declaration.

Materials

Measures of touch discrimination were the Tactile Discrimination Test (TDT)15 and the Fabric Matching Test (FMT).12 The TDT employed finely graded plastic surfaces marked by ridges at set spatial intervals. Texture grids were presented in sets of three, with two surfaces identical and one different. The FMT comprised two identical sets of ten cotton-based fabrics, ranked from smooth to rough by unimpaired subjects. These measures have high retest reliability and good discriminative validity and normative standards.14,16

For the proprioception domain, the Wrist Position Sense Test (WPST) quantified the subject’s ability to indicate wrist position after an imposed movement in the flexion-extension14 or ulnar-radial deviation12 plane of movement. A box-like apparatus included two protractor scales to indicate target and response positions and splints for forearm and hand. The hand splint was attached to a lever allowing freedom of movement at the wrist. The box occluded vision of subject’s wrist position and of the examiner’s lever manipulations. Each test comprised 20 predetermined wrist positions.

High retest reliability and good discriminative validity and normative standards have also been established for these tests.14,16

Procedure

Five single-case experiments were conducted to investigate spontaneous transfer of training to related stimuli in study 1. Two experiments investigated transfer within the tactile-discrimination dimension and three within proprioception. In the tactile experiments, texture grids were employed as the primary stimuli to be trained and the fabrics as the transfer stimuli. Imposed wrist positions within the flexion-extension plane were the primary trained proprioception stimuli and positions within the ulnar-radial deviation plane the transfer stimuli.

The typical case experiment had three phases, each phase comprising ten sessions in which both the trained and transfer tasks were assessed. Assessment sessions were scheduled 48–72 hrs apart. In the first phase, both tasks were monitored only. In the second phase, the texture grid or flexion-extension stimuli were trained over ten treatment sessions, interspersed between assessment sessions, while baseline monitoring continued on the transfer response. In the third phase, treatment was introduced for the transfer stimuli over ten sessions. This permitted investigation of the subject’s potential for a SST effect on the transfer task. Follow-up was conducted after an interval equivalent to the combined time of baseline and intervention phases (i.e., 12–14 wks after the end of training).

Test Procedure and Scoring

The TDT was administered after standard instructions were given.15 Subjects tactually explored each set of comparison surfaces with their preferred finger and indicated the odd texture in a three-alternative, forced-choice design. Five different triplets, spanning Weber ratios of 0.033–1.0, were each presented ten times in random order to obtain the discrimination limen. The limen was derived from fitting a cumulative normal distribution to the probabilities of correct responses. In the FMT, subjects attempted to tactually match each of the ten test surfaces with identical comparison surfaces.12 Fabrics were presented behind a curtain, using a predetermined random order and a standard set of instructions. The preferred finger used in the TDT was also used for the fabric test. Response and target rank orders were correlated with Spearman’s rho to quantify fabric discrimination ability. Rho values were transformed with Fisher’s z, and the scale was inverted by subtracting scores from the allocated maximum value of
3.5 z score to make graphical presentation of therapeutic change consistent in direction with TDT scores.

In the WPST for flexion-extension and ulnar-radial deviation, the examiner moved the subject’s hand, via a lever, to 20 different predetermined wrist positions. Average absolute error between actual and response positions was then calculated as the index of limb position sense for each test.

**SST**

The SST program was designed to maximize improvement of the specific sensory discriminations trained. Principles of training included: repeated presentation of targeted discrimination tasks (as learning is reported to be maximal for the specific task trained); progression from easy to more difficult discriminations; attentive exploration of stimuli with vision occluded; use of anticipation trials; feedback on salient sensory features of the stimuli, accuracy of judgments, and method of exploration; comparison of the sensation with the other hand; use of vision to facilitate intermodal calibration of sensory information; summary feedback; and intensive training. These principles were applied to training each stimulus, as described below.

In the SST program, training tasks were the same as the assessment tasks, but typically used a restricted range and number of stimuli. Five sets of grids, with differences in surfaces ranging from 3.3% spatial increase to 100% spatial increase, were used for training texture grids, as previously described. Subjects explored each set of comparison grid surfaces with their preferred finger and indicated the odd texture. Training of fabric discrimination used the fabrics of the FMT. Subjects were required to match comparison surfaces to a restricted range of target surfaces and, in some trials, to identify which fabrics they had touched. Training employing a subset of fabrics with four levels of difference, from large to small differences, was provided, as above. When correct discriminations occurred in at least three of four consecutive occasions, the next-finer stimulus difference was introduced. Summary feedback on the individual’s accuracy of judgments and method of exploration was also provided at the end of each training session. Training was intensive, with sessions lasting 40–60 mins, three times a week.

**Data Analysis**

First, each single-case experiment was analyzed separately for stimulus-specific and generalized training effects. Second, these results were combined in a meta-analysis to obtain an overall conclusion.

Time-series data were analyzed using visual (graphical) and statistical analyses. For graphical analyses, standard single-case charts were evaluated visually by the authors and a panel of three trained, independent analysts who were naïve to the purpose of the study. The analysts’ type I error (false alarms) and type II error (missed treatment effect) rates were calibrated using computer-generated charts in which null or non-null effects were presented unidentified and in random order. Analysts were required to make separate judgments of systematic change between phases according to defined criteria.

Statistical interrupted time-series analyses (ITSA) were also conducted for each single-case experiment. The initial step comprised model iden-
tification, in which three curvilinear models of learning likely to adequately describe the intervention effect were compared for goodness of fit using the curvilinear regression routines of SPSS. The models $Y = a + b \ln X$, $Y = aX^n$ and $Y = ae^{bX}$ were compared, consistent with earlier investigations.\(^8\) Residuals from the models were examined by autocorrelation and partial autocorrelation methods to detect remaining serial dependence in the data. The best-fitting model with residuals free of positive serial dependence was selected to describe the data and perform ITSA. After investigation and deletion of outliers, ITSA were tested for trend and level effects between phases to evaluate intervention effects, as previously described.\(^8\) A systematic difference between phases was judged to be present if either a statistically significant trend or level effect was present. Determination of a spontaneous transfer effect involved (a) identification of an intervention effect on the primary trained response together with (b) identification of a simultaneous, systematic change in a therapeutic direction in the transfer response.

Finally, meta-analyses were conducted to complement individual subject ITSA and assist a generalized conclusion across the case experiments. The type I error probabilities of each ITSA were converted to $z$ scores and then averaged to obtain a pooled estimate.\(^17\) Separate meta-analyses examined stimulus-specific intervention effects, spontaneous generalization effects, and the effect of the intervention designed to facilitate generalized improvement in texture discrimination.

**Study 2: Transfer of Training Across Textured Surfaces Using a Program Designed to Facilitate Generalization to Novel Stimuli**

**Subjects**

Five further stroke patients, who met selection criteria previously described and had tactile discrimination impairment, were investigated. Patients meeting these criteria were selected sequentially, as they presented, and gave voluntary informed consent.

**Materials**

The FMT was again employed, as previously described. The Grid Matching Test (GMT)\(^12\) tested texture-grid matching ability. The test used the same finely graded plastic grids as for the TDT but employed a stimulus-matching procedure, as in the FMT. Sixteen surfaces, with spatial intervals ranging from 1500 to 3000 $\mu$m, were placed on test and comparison texture wheels and presented for tactual exploration through openings in the test apparatus. Reliability, discriminative validity, and normative data have been obtained for this test.\(^18\)

**Procedure**

This study comprised five multiple-baseline, single-case experiments. Performance on the GMT and FMT was monitored throughout the time series. The first phase comprised baseline only. In the second phase, SST (described in study 1) was introduced for texture grids while the transfer response was monitored. In the third phase, SGT was introduced while monitoring of the untrained transfer response continued. Follow-up was conducted 12–14 wks after the end of training.

**Test Procedure and Scoring**

The stimulus-matching procedure was used for both tests. For the GMT, the subject was required to tactually match each of eight test surfaces (selected from the set of 16) against the identical set, with vision occluded. For the FMT, subjects were required to match each of the ten test surfaces with the ten comparison surfaces. Testing took approximately 10–15 mins for each test. Stimulus-matching ability was quantified by correlating the response and test values with Pearson’s $r$ (GMT) and Spearman's rho (FMT). Correlation coefficients were then normalized with Fisher's $z$ transform. A higher score reflects better matching.

**SGT**

The SGT program was designed to facilitate transfer of training effects to untrained, novel stimuli. To achieve this, the additional principles of variation in stimulus and practice conditions, intermittent feedback, and tuition of training principles\(^1,2,19\) were included, as these have been associated with enhanced transfer and retention. A variety of training surfaces (e.g., paper, glass, leather, and rubber, but not fabrics) were employed in the SGT program. The surfaces comprised a range of distinctive features of roughness, including contour, surface pattern, and grit. Each different type of surface was graded from smooth to rough across five stimuli and included small to large differences. First, large differences were introduced across a subset of surface types, followed by medium and fine differences. Thus, progressive grading of difficulty was across stimuli and within stimuli, and subjects had the opportunity of making similar discriminations across novel surface types. Feedback on salient sensory features, accuracy, and exploration method (as for SST) was given intermittently rather than at every trial, and subjects were encouraged to check the accuracy of their own performance. Feedback was also given on the transfer task of identifying new distinctive features of roughness in novel stimuli. Specific tuition on principles underlying training, such as use
of anticipation trials and feedback, and how these apply across tasks that the client may encounter in other environments were also included. Repeated presentation of stimuli, attentive exploration with vision occluded, anticipation trials, summary feedback, and intensive training were also incorporated, as for SST.

Data Analysis

Case charts were analyzed visually and statistically for trend and level effects. Individual time series were graphed for visual analysis, and statistical analysis was conducted, after model identification, as previously described.

RESULTS

Study 1

Background data of subjects is detailed in Table 1. All subjects showed marked impairment on the TDT (i.e., unable to consistently discriminate the largest texture difference of 100% spatial increase [criterion of normality, <37.7% spatial increase\textsuperscript{15}]). Impaired performance was also evident across multiple modalities, using the following tests and normative standards: wrist proprioception using the WPST for flexion-extension (criterion of normality, 9.5 degrees\textsuperscript{14,16}); pressure discrimination using Semmes-Weinstein monofilaments (normal threshold = 0.02–0.13 filament\textsuperscript{20}); hot/cold discrimination using the Roylan hot and cold discrimination kit (A629-1) (age-matched, healthy controls detect at least 9 out of 10 correct\textsuperscript{16}).

Table 1: Background information of subjects in study 1

<table>
<thead>
<tr>
<th>Subject</th>
<th>S01</th>
<th>S02</th>
<th>S03</th>
<th>S04</th>
<th>S05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, yrs</td>
<td>44</td>
<td>60</td>
<td>59</td>
<td>51</td>
<td>50</td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>Male</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>Hand dominance\textsuperscript{a}</td>
<td>Left (consistent)</td>
<td>Right (consistent)</td>
<td>Left</td>
<td>Right (consistent)</td>
<td>Left</td>
</tr>
<tr>
<td>Affected side</td>
<td>Left</td>
<td>Right</td>
<td>Left</td>
<td>Right</td>
<td>Right</td>
</tr>
<tr>
<td>Time since stroke onset</td>
<td>10.5 wks</td>
<td>6.5 wks</td>
<td>5.5 wks</td>
<td>13.5 wks</td>
<td>8 wks</td>
</tr>
<tr>
<td>Site of lesion (based on CT scan)</td>
<td>Right frontoparietal infarct</td>
<td>Right temporal lobe, corona radiata, and lentiform nucleus infarct</td>
<td>Left frontoparietal and corona radiata infarct</td>
<td>Left lentiform-internal capsule hemorrhage</td>
<td>Right basal ganglia, internal capsule hemorrhage</td>
</tr>
<tr>
<td>Motor function (affected UL)</td>
<td>Limited voluntary movement</td>
<td>Isolated finger control</td>
<td>No voluntary movement of wrist and fingers</td>
<td>No functional movement</td>
<td>No functional movement</td>
</tr>
<tr>
<td>Sensory function (affected UL)</td>
<td>TDT = 100 PSI; WPST = 23.9 degrees; Pressure = 6.65; H/C = 5/10</td>
<td>TDT = 100 PSI; WPST = 11.3 degrees; Pressure = 4.31; H/C = 7/10</td>
<td>TDT = 100 PSI; WPST = 32.4 degrees; Pressure = 4.31; H/C = 6/10</td>
<td>Mildly reduced attention processes; slow information processing</td>
<td>NAD; alert; concentration, memory, and learning intact</td>
</tr>
<tr>
<td>Cognitive function</td>
<td>Slight decrease in complex planning skills, otherwise NAD</td>
<td>Impaired language: integrity of problem solving and memory</td>
<td>NAD; attention, memory, and learning intact</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CT scan, computed tomographic scan; UL, upper limb; TDT, Tactile Discrimination Test\textsuperscript{15}; PSI, percentage spatial increase; WPST, Wrist Position Sense Test in flexion-extension plane\textsuperscript{15}; Pressure, log\textsuperscript{12} force of Semmes-Weinstein monofilament (in milligrams) detected at preferred finger\textsuperscript{17}; H/C, hot/cold discrimination using Roylan hot and cold discrimination kit (A629-1), number of correct judgments out of ten stimuli presented to distal pad of preferred finger are reported; N/A, not available; NAD, no abnormality detected.

\textsuperscript{a}Hand dominance determined from Annett\textsuperscript{21} questionnaire of hand dominance.

Model Identification

The logarithmic model \( Y = a + b \ln X \) accounted for most variance in the intervention data for eight of the ten time series, without high positive autocorrelation in the residuals of any series. Statistically significant negative autocorrelation (lag 2) was identified only in two baseline phases. The analysis of residuals thus confirmed selection of a logarithmic model for ITSA.

Intervention Effects in the Trained Response

Intervention effects were clearly apparent in three of the five first-trained primary responses after SST (Fig. 1, S01 to S03). These improvements were detected above any changes between the first and second phase of baseline of the second discrimination response. Improvements were marked and immediate, with changes in performance noted after the first training session. Performance scores achieved by the end of training were within the normal range\textsuperscript{15} and similar to those of the other hand in all three cases. The panel of judges unanimously detected a systematic change (\( P \sim 0.00 \)) in all three series and did not identify a significant
parallel change in the transfer response, suggesting a specific training effect. Statistical analysis confirmed both trend and level effects in the three series (Table 2). Improvements achieved by the end of the training were well maintained. In the remaining two series, no intervention effect could be claimed (Fig. 1, S04 and S05). Both series showed an improvement during baseline, limiting the scope for training effects.

Spontaneous Transfer of Training

Spontaneous transfer of training to the second response was not evident in two of the three cases in which an intervention effect was present in the primary trained response (Fig. 1, S01 and S03). Lack of a concomitant and systematic change in the transfer responses was confirmed by the panel of judges. Although a statistically significant change in trend was obtained for subject 3 (Table 3, S03) the change was not in the therapeutic direction. Subject 2 demonstrated variable performance on the transfer task (FMT) during sessions 1–10 of baseline and a reduced variability and possible improvement during sessions 11–20 of baseline. It was difficult to distinguish whether these observations represented continuation of the pattern established in the first baseline phase or a change in the second baseline phase; therefore, judgment was reserved. The panel of visual analysts failed to indicate a systematic change. However, a statistically

FIGURE 1  Case charts of spontaneous transfer effects within tactile and proprioceptive domains. In phase 1 (sessions 1–10), performance was monitored under baseline conditions for both trained and transfer stimuli. In phase 2 (sessions 11–20), stimulus-specific training was introduced to the primary trained response (texture grids or flexion-extension wrist positions). Performance of the transfer response (fabrics or ulnar-radial wrist positions) was simultaneously monitored under extended baseline conditions to permit observation of spontaneous transfer effects. In phase 3 (sessions 21–30), stimulus-specific training was introduced to the transfer response and continued for the primary trained response but at a lower intensity. Texture grid limen (PSI), score from the Tactile Discrimination Test in units of percentage spatial increase.15 Fabric matching z-score, score from Fabric Matching Test in Fisher's z units.12,16 Fisher's z scale was inverted so that graphical representation of the Fabric score was consistent with the Tactile Discrimination Test score (i.e., higher values represent poorer performance). Flex-ext position error (deg), flexion-extension wrist position sense average error, in degrees, based on the Wrist Position Sense Test.14 Uln-rad position error (deg), ulnar-radial deviation wrist position sense average error, in degrees, based on the Ulnar-Radial Wrist Position Sense Test.12,16 S01–S05, subjects 1–5.

significant change in level, from 2.6 to 2.1, on the fabric-matching score was observed (Table 3). The identified change was in a therapeutic direction but of small magnitude and did not seem to follow the characteristic logarithmic function. In the remaining two cases (subjects 4 and 5), an intervention effect in the first-trained primary response was not identified; therefore, generalization of a training effect was not testable.

### Effect of SST on the Transfer Stimuli

Although the fabric and ulnar-radial position stimuli did not demonstrate evidence of spontaneous generalization effects, they did respond to SST

### TABLE 2 Stimulus-specific training effects: Summary of trend and level effects from interrupted time series analyses

<table>
<thead>
<tr>
<th>Subject</th>
<th>Time Series</th>
<th>Trend Effect</th>
<th>Level Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$t(df)$</td>
<td>$P$, 1-Tail</td>
</tr>
<tr>
<td>Study 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S01</td>
<td>Texture grids</td>
<td>$t(26) = 5.67$</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td></td>
<td>Fabrics</td>
<td>$t(15) = 3.10$</td>
<td>$&lt;0.004$</td>
</tr>
<tr>
<td>S02</td>
<td>Texture grids</td>
<td>$t(25) = 8.83$</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td></td>
<td>Fabrics</td>
<td>$t(16) = 2.67$</td>
<td>$&lt;0.009$</td>
</tr>
<tr>
<td>S03</td>
<td>Flex-ext</td>
<td>$t(26) = 2.51$</td>
<td>$&lt;0.01$</td>
</tr>
<tr>
<td>S04</td>
<td>Flex-ext</td>
<td>$t(26) = 3.22$</td>
<td>$&lt;0.002$</td>
</tr>
<tr>
<td>S05</td>
<td>Flex-ext</td>
<td>$t(26) = 4.06$</td>
<td>$&lt;0.0002$</td>
</tr>
<tr>
<td></td>
<td>Ulnar-radial</td>
<td>$t(16) = 1.52$</td>
<td>$&gt;0.05$</td>
</tr>
<tr>
<td>Study 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S06</td>
<td>Texture grids</td>
<td>$t(16) = 4.57$</td>
<td>$&lt;0.0002$</td>
</tr>
<tr>
<td>S07</td>
<td>Texture grids</td>
<td>$t(16) = 2.92$</td>
<td>$&lt;0.005$</td>
</tr>
<tr>
<td>S08</td>
<td>Texture grids</td>
<td>$t(16) = 0.09$</td>
<td>$&gt;0.05$</td>
</tr>
<tr>
<td>S09</td>
<td>Texture grids</td>
<td>$t(21) = 3.34$</td>
<td>$&lt;0.002$</td>
</tr>
<tr>
<td>S10</td>
<td>Texture grids</td>
<td>$t(18) = 3.84$</td>
<td>$&lt;0.0006$</td>
</tr>
</tbody>
</table>

*Flex-ext, wrist position stimuli in the flexion-extension plane of movement; ulnar-radial, wrist position stimuli in the ulnar-radial deviation plane of movement.*

*Bold $P$ values indicate a significant effect in a therapeutic direction; italic $P$ values indicate a significant difference between phases, but an intervention effect was not claimed due to pre-existing improvement in baseline. Comparisons are between baseline (second phase of baseline for transfer stimuli) and stimulus-specific training phases.*

### TABLE 3 Spontaneous transfer effect to the transfer discrimination response

<table>
<thead>
<tr>
<th>Subject</th>
<th>Time Series</th>
<th>Trend Effect</th>
<th>Level Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$t(df)$</td>
<td>$P$, 1-Tail</td>
</tr>
<tr>
<td>Study 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tactile discrimination</td>
<td>Fabrics</td>
<td>$t(16) = 0.60$</td>
<td>$&gt;0.05$</td>
</tr>
<tr>
<td>S02</td>
<td>Fabrics</td>
<td>$t(16) = 0.38$</td>
<td>$&gt;0.05$</td>
</tr>
<tr>
<td>Proprionception discrimination</td>
<td>Ulnar-radial</td>
<td>$t(16) = 2.29$</td>
<td>$&lt;0.02$</td>
</tr>
<tr>
<td>S04</td>
<td>Ulnar-radial</td>
<td>$t(16) = 1.95$</td>
<td>$&lt;0.04$</td>
</tr>
<tr>
<td>S05</td>
<td>Ulnar-radial</td>
<td>$t(16) = 0.26$</td>
<td>$&gt;0.05$</td>
</tr>
<tr>
<td>Study 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tactile discrimination</td>
<td>Fabrics</td>
<td>$t(16) = 1.32$</td>
<td>$&gt;0.05$</td>
</tr>
<tr>
<td>S07</td>
<td>Fabrics</td>
<td>$t(16) = 1.41$</td>
<td>$&gt;0.05$</td>
</tr>
<tr>
<td>S08</td>
<td>Fabrics</td>
<td>$t(16) = 1.20$</td>
<td>$&gt;0.05$</td>
</tr>
<tr>
<td>S09</td>
<td>Fabrics</td>
<td>$t(21) = 2.32$</td>
<td>$&lt;0.02$</td>
</tr>
<tr>
<td>S10</td>
<td>Fabrics</td>
<td>$t(18) = 0.37$</td>
<td>$&gt;0.05$</td>
</tr>
</tbody>
</table>

*Ulnar-radial, wrist position stimuli in the ulnar-radial deviation plane of movement.*

*Bold $P$ values indicate a significant effect in a therapeutic direction; italic $P$ values indicate a significant difference between phases but not in a therapeutic direction. Comparisons are between the first and second phase of baseline for the transfer stimulus.*
in all cases (Fig. 1). Improvements were clearly defined and marked relative to the second phase of baseline. All subjects performed within the normal range on the FMT and WPST for ulnar-radial deviation by the end of training. In the ulnar-radial series, the lowest error scores were achieved. Independent visual analysis \((P \sim 0.00)\) and statistically significant trend and level effects (Table 2) confirmed the presence of intervention effects relative to the control baseline. Improvements achieved by the end of training were well maintained at follow-up in all cases.

**Study 2**

Background details of subjects are presented in Table 4. Subjects showed severe deficit on the GMT (−0.12 to 0.29 \(z\) score compared with 0.62 \(z\) score, the criterion of normality) and moderate to severe impairment on the FMT (criterion of normality, >1.61 \(z\) score). Most patients had the maximum impairment score on the TDT. Performance on the WPST varied.

**Model Identification**

The logarithmic model previously identified in several of our studies and in study 1 again successfully accounted for baseline and intervention data, leaving no statistically significant or suspiciously high positive autocorrelation in the residuals of any time series.

**Effects of SST**

The case charts (Fig. 2) suggested a systematic change between baseline and SST phases for the GMT beyond any parallel change in the transfer responses of each of the subjects. Improvements were marked, often achieving scores within the normal performance range and comparable with the other hand. These effects of SST were confirmed for each subject by the ITSAs (Table 2, S06 to S10).

**Spontaneous Transfer to Untrained Fabric Stimuli After SST**

Spontaneous transfer to the untrained FMT response was not evident in the second baseline phase when SST of texture grids was introduced (Fig. 2). This was clear for subjects 6 – 8. Subjects 9 and 10 showed more variability. Lack of a transfer effect was confirmed by the results of the ITSA in all but one case (Table 3). A significant trend effect was identified in subject 9. However, the change did not follow the characteristic pattern achieved by training, and performance was variable; thus, a spontaneous transfer effect was not claimed.

**Transfer to Novel Textured Stimuli After SGT**

Discrimination of fabric surfaces, which had not been trained in any of the three phases, im-

<table>
<thead>
<tr>
<th>TABLE 4 Background information of subjects in study 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
</tr>
<tr>
<td>Age, yrs</td>
</tr>
<tr>
<td>Sex</td>
</tr>
<tr>
<td>Hand dominance(^a)</td>
</tr>
<tr>
<td>Affected side</td>
</tr>
<tr>
<td>Time since stroke onset</td>
</tr>
<tr>
<td>Site of lesion (CT scan)</td>
</tr>
<tr>
<td>Motor function (affected UL)</td>
</tr>
<tr>
<td>Sensory function (affected UL)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Cognitive function</td>
</tr>
</tbody>
</table>

\(^a\)Hand dominance determined from Annett\(^{21}\) questionnaire of hand dominance.
proved on introduction of the SGT program in four of the five experiments (subjects 6–9), as shown in Figure 2. Improvements were often marked, achieving scores within the normal performance range. They followed a pattern similar to the SST effects, although the magnitude of early change

FIGURE 2 Case charts of stimulus-generalization training effects in the tactile domain. In phase 1 (typically sessions 1–10), performance was monitored under baseline conditions for both trained and transfer stimuli. In phase 2 (typically sessions 11–20), stimulus-specific training (SST) was introduced to the texture grids and spontaneous transfer effects were monitored in the transfer-fabric stimulus. In phase 3 (typically sessions 21–30), stimulus-generalization training (SGT) was introduced, and monitoring continued for the fabric response to permit observation of transfer of training effects. Grid-matching and fabric-matching test scores are in Fisher's z units. In contrast to Figure 1, higher values represent better performance. Case chart S06:trained/untrained grids provides an example of the case charts in which the grid-matching score was separately calculated for grids that were trained and those that were not trained. Stimulus-specific training effects are evident in the trained grid score, and spontaneous transfer of training effects are evident in the untrained grid score. S06–S10, subjects 6–10.
was often smaller. Subject 10 also showed an improvement trend in phase 3; however, this was not significantly different from the preexisting trend evident in phase 2 (Fig. 2, Table 5). Statistical analyses are summarized in Table 5.

**Spontaneous Transfer from Trained to Untrained Grids After SST**

To investigate whether transfer might occur across stimuli with the same texture characteristic, we trained for only half of the grid surfaces, while monitoring performance on all (Fig. 2, S06: *trained/untrained grids*). Spontaneous transfer from trained texture grids to untrained texture grids did occur with the SST program in four of the five case experiments (Table 6). Subject 10 was the only subject who did not show either spontaneous transfer to stimuli with the same distinctive feature (texture grids) or facilitate transfer to novel stimuli (fabrics).

**Meta-analyses**

A meta-analysis investigating the overall effect of SST with the plastic texture grids was performed on the seven subjects providing appropriate data (subjects 1, 2, and 6-10). This meta-analysis very clearly confirmed the positive effect of training on the discrimination of texture grids according to both level (Z = 7.61, P < 0.001) and trend (Z = 8.46, P < 0.001) effects. Discriminations within the same difficulty range, but using plastic grids with spatial intervals that were not used during training trials, were available from five experiments (subjects 6-10). This meta-analysis also showed significantly improved performance in both the level (Z = 3.43, P < 0.001) and trend (Z = 4.61, P < 0.001) effects. These strong positive results in discriminations of trained and untrained texture grids contrast markedly with the results obtained for spontaneous transfer–related changes in the discrimination of fabric textures after SST of plastic grids. Meta-analysis of these time series failed to show a statistically significant effect in either level (Z = -0.206, P > 0.41) or trend changes (Z = 0.126, P > 0.55), despite the increase in sensitivity afforded by the pooling of seven sets of experimental results. A meta-analysis that directly contrasted the effect of specific grid training on discrimination of plastic grids vs. discrimination of fabrics showed a clearly superior effect in the discrimination of the plastic grids according to both level (Z = 5.24, P < 0.001) and trend (Z = 6.07, P < 0.001) results. Importantly, when the SGT program was introduced in the second group of five experiments, both level (Z = 4.10, P < 0.001) and trend (Z = 5.57, P < 0.001) effects responded with significant changes in the untrained transfer stimuli.

**DISCUSSION**

Several conclusions seem reasonable. First, rapid, task-specific training effects can be obtained with the SST program. Second, improvements obtained in a particular discrimination task do not generalize spontaneously to related stimuli within the same sensory-perceptual dimension in stroke patients after SST. For example, training on grid surfaces did not generalize to fabrics. However, spontaneous transfer of training does seem to occur across stimuli that share the same distinctive feature (e.g., trained and untrained grids) after SST. Third, performance on related untrained stimuli within the same sensory-perceptual dimension can be improved to a clinically significant degree by using a program deliberately oriented to enhance generalization (the SGT program). Transfer of training from practice with paper, glass, and rubber to discrimination of fabric surfaces illustrates this point.

**Conditions of Training and Influence on Transfer**

Successful transfer across stimuli seems to have been influenced by the training principles employed. The SST program employed principles designed to maximize learning to discriminate a particular stimulus type. Positive training and transfer effects within the specific stimulus type trained were achieved. However, spontaneous

<table>
<thead>
<tr>
<th>Subject</th>
<th>Time Series</th>
<th>Trend Effect</th>
<th>Level Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>t(df)</td>
<td>P, 1-Tail</td>
</tr>
<tr>
<td>S06</td>
<td>Fabrics</td>
<td>t(16) = 5.37</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>S07</td>
<td>Fabrics</td>
<td>t(16) = 3.41</td>
<td>&lt;0.002</td>
</tr>
<tr>
<td>S08</td>
<td>Fabrics</td>
<td>t(16) = 4.13</td>
<td>&lt;0.0004</td>
</tr>
<tr>
<td>S09</td>
<td>Fabrics</td>
<td>t(21) = 1.77</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>S10</td>
<td>Fabrics</td>
<td>t(23) = 1.13</td>
<td>&gt;0.05</td>
</tr>
</tbody>
</table>

Bold P values indicate a significant effect in a therapeutic direction. Comparisons are between the second phase of baseline and the stimulus-generalization training phase for the transfer stimulus.
TABLE 6 Stimulus-specific training effect on trained grids and spontaneous transfer to untrained texture grids

<table>
<thead>
<tr>
<th>Subject</th>
<th>Texture Grids</th>
<th>Trend Effect</th>
<th>Level Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>t(df)</td>
<td>P, 1-Tail</td>
</tr>
<tr>
<td>S06</td>
<td>Trained</td>
<td>t(16) = 5.49</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>Untrained</td>
<td>t(16) = 3.02</td>
<td>&lt;0.004</td>
</tr>
<tr>
<td>S07</td>
<td>Trained</td>
<td>t(16) = 1.49</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td></td>
<td>Untrained</td>
<td>t(16) = 0.62</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>S08</td>
<td>Trained</td>
<td>t(16) = 0.67</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td></td>
<td>Untrained</td>
<td>t(16) = 0.41</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>S09</td>
<td>Trained</td>
<td>t(21) = 4.77</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>Untrained</td>
<td>t(21) = 5.04</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>S10</td>
<td>Trained</td>
<td>t(18) = 5.30</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>Untrained</td>
<td>t(18) = 0.40</td>
<td>&gt;0.05</td>
</tr>
</tbody>
</table>

Bold P values indicate a significant effect in a therapeutic direction. Comparisons are between baseline and stimulus-specific training phases for the trained grids and between the first and second phase of baseline for the untrained grids.

Transfer to related tasks within the same sensory-perceptual dimension did not occur. To facilitate generalized improvements, a program that added variation in training stimuli, intermittent feedback, and tuition of training principles was required. Repeated exposure under conditions of attentive exploration with vision occluded was usually insufficient for therapeutic change, as indicated by stable baseline performance in 75% of the baseline time series investigated. Thus, conditions of training seem important to outcome, with implications for the design of efficient rehabilitation programs. The principles of training derived from studies of learning with normal subjects seem applicable to stroke patients, confirming the value of learning theory for models of neurologic rehabilitation.

Specificity of Learning and Information Processing Within the Sensory System

Perceptual learning studies with unimpaired subjects suggest that discrimination training is often highly specific to the task and receptor location and to the method of processing information. We also found highly specific training effects in tasks employing the same sensory dimension (tactile or proprioceptive) and body location (fingertip or wrist) when SST was undertaken. Spontaneous transfer was, however, found across stimuli of the same type that employed the same receptor location and method of processing the information (e.g., from trained to untrained grids and across wrist positions within the same plane of movement [positions trained were different to those assessed]). These findings suggest spontaneous transfer of training to novel stimuli of the same type and of comparable difficulty when SST is employed.

The high degree of specificity observed is also consistent with evidence of highly specific deficits resulting from cerebral damage. Moreover, studies of behaviorally induced neural plasticity suggest that changes in cortical maps are specific to the trained task and specific body location in monkeys and humans. Organization of the somatosensory system is highly specific, as evidenced by body location and modality-specific columnar organization and submodality-specific neurons in the primary somatosensory cortex and neuro-axis. These observations indicate a high degree of specificity of information processing within modalities and body locations. However, the sensory system is also characterized by convergence of somatosensory information and presence of distributed sensory networks. Thus, distinctive sensory features of a multidimensional texture might be integrated in a unified perception at a level of convergence in the system. This level of processing would be consistent with the learning and transfer across textures with multiple features of roughness after SGT.

What Is Learned

A common view supported by behavioral and neurophysiologic evidence proposes that distinctive features of difference are learned and form the basis of transfer of training. In the TDT and GMT, the distinctive feature is likely to be the spatial interval of the grids, whereas the physical characteristics of fabrics are likely to be based on nonspatial neural coding mechanisms. This inference is supported by observation of stimulus-specific transfer within, but not across, these stimuli. The distinctive feature of the limb-position task is presumed to be the change in location of the limb or relative wrist angle. Our findings suggest that the distinctive feature also needs to be defined in relation to the
defined plane of movement or to receptors involved in the discrimination. This conclusion is consistent with observations indicating selective activation of muscle receptors, joint receptors, and skin areas in association with specific positions. Thus, it seems necessary that the distinctive features of stimulus difference are highly similar and specific to the receptor population activated to obtain spontaneous generalization.

Using the SGT program, transfer was obtained across tactile stimuli with potentially different distinctive features (i.e., from training on rubber, glass, leather, and sandpaper to transfer on fabrics). Training of multidimensional stimuli with a wide range of distinctive features in the SGT program is likely to increase the probability of exposure to distinctive features relevant to the transfer stimulus, with consequent transfer of learning. Investigations with unimpaired subjects have suggested that transfer is facilitated when stimuli are more complex and potentially share some distinctive features. Physical characteristics of difference trained in our study included surface type, surface contour, grit, friction/abrasion, stimulus patterns with different spatial frequencies, surface irregularities, and various combinations of these. This set of features seems to represent a broader dimension of roughness and covers aspects of roughness that are both spatial and nonspatial. Thus, success of the SGT may be influenced by the fact that the varied stimuli used included nonspatial discriminations that may be more directly related to the transfer stimulus. Further, it may be the dimension of roughness, rather than a specific unitary feature of it, that is learned and transferred in this training. Other studies obtaining generalized training effects in stroke patients have included nonspatial discriminations. This conclusion is consistent with observations indicating selective activation of muscle receptors, joint receptors, and skin areas in association with specific positions. Thus, it seems necessary that the distinctive features of stimulus difference are highly similar and specific to the receptor population activated to obtain spontaneous generalization.

Study Limitations

Investigation of training effects was limited to ten subjects (20 time series). Although this is a small number for group-based studies, each intensive single case is a controlled experiment involving within-subject control and replication of training effects. This approach is well suited to the heterogeneous nature of stroke and allows for variable findings across individuals that might be obscured with group analysis. In addition, systematic replication of training effects across ten subjects with different background characteristics and across tactile and proprioceptive modalities provides generalizability of findings not usually afforded to single case studies. Finally, meta-analysis of individual experiments improves power, permits an overall conclusion, and provides a quantitative evaluation of the generality of treatment effects. Thus, our findings provide a strong foundation for larger controlled studies.

The sequence of baseline, SST, and SGT was constant, and therefore, it may be argued that the exposure and SST phases led to an overlearning effect with an impact on transfer and effectiveness of the SGT program. However, this explanation is unlikely given that observed SST effects worked...
quickly, that even prolonged exposure (e.g., 20 sessions) did not effect changes, and that changes observed for SGT were of similar rapidity. These observations support the conclusion that improvements in the SGT phase were due to a generalization training effect, rather than due to additional dosage of stimulus exposure, or SST.

**Implications for Therapeutic Interventions and Further Investigations**

Clinically significant improvements were obtained with training, providing justification for therapeutic intervention. Practice or exposure alone was not usually sufficient to achieve the changes characteristic of perceptual learning. The findings also suggest that the nature of training is crucial to outcome. SST may be important if there is a need to train specific work or daily living tasks. However, to facilitate more generalized improvements and minimize reliance on resource-intensive, task-specific training, a program that also includes variation within and across related training stimuli, intermittent feedback, and tuition of training principles is recommended. In the majority of cases, ten training sessions seem to be sufficient to achieve clinically significant improvement. However, the goal of healthcare providers is not only to provide an effective training program but also an efficient training program. The highly specific nature of training suggests that clinicians need to accurately identify the specific sensory dimension that requires training and use graded stimuli that cover the range of distinctive features for the perceptual dimension targeted. Programs cannot rely on spontaneous transfer; they need to systematically train for transfer. This issue of generalization of training effects has importance in relation to rehabilitation of a range of abilities in stroke patients and to rehabilitation in general.

The present studies provide some guidance on both the nature of transfer expected and on the conditions of training required to achieve transfer. Further systematic investigation of these variables is indicated. For example, transfer across macrospatial and microspatial features of surfaces will help to determine if the boundary of spontaneous transfer is at the level of a particular spatial pattern or broader spatial features of roughness. Transfer across dimensions such as smoothness–roughness, hardness–softness, and compressional elasticity within multidimensional texture stimuli is also indicated, particularly after SGT. Investigation of transfer across different joints of the upper limb for proprioception or different fingers for texture is also necessary. Identification of sites of cerebral reorganization underlying the recovery of somatosensory function after stroke may provide new insights into cerebral networks involved in perceptual learning and recovery. Systematic investigation of the relative contribution of training principles in the treatment package would also help to identify the critically important elements associated with successful learning and transfer. Finally, the ability to modify sensory abilities experimentally opens up an experimental paradigm for future investigations of the relationship between sensation and other abilities, such as pinch grip, and the effect of improving sensation on motor function.

**ACKNOWLEDGMENTS**

We thank the participants for their involvement in the study and the therapists and management at Royal Talbot Rehabilitation Centre, Cedar Court Healthsouth Rehabilitation Hospital, and Hampton Rehabilitation Hospital, Victoria, Australia, for their support.

**REFERENCES**

After Stroke: Training Effects and Quantitative Measurement [PhD thesis]. Melbourne, LaTrobe University, 1993


20. Directions for Using Semmes-Weinstein Monofilaments. San Jose, North Coast Medical, 1987


Diagnostic Precision of Ultrasonography in Patients with Carpal Tunnel Syndrome

ABSTRACT


Objective: To evaluate the diagnostic value of ultrasonography in patients with electrophysiologically confirmed carpal tunnel syndrome.

Design: A prospective ultrasonographic study of 35 wrists with electrophysiologically confirmed carpal tunnel syndrome and of 40 normal wrists. Receiver-operating-characteristics curves for the ultrasonographic measurements of median nerve were plotted to identify the most optimal cutoff values.

Results: The ultrasonographic measurements of median nerves were found to be increased significantly in patients with carpal tunnel syndrome when compared with controls, particularly in terms of cross-sectional area ($P < 0.001$) and the bowing of the flexor retinaculum ($P < 0.01$) but not in the flattening ratio ($P > 0.05$). According to receiver-operating-characteristics curve results, the most optimal cutoff value for the cross-sectional area of the median nerve was obtained at the level of middle carpal tunnel, which was 9.3 mm$^2$, with a sensitivity of 80% and specificity of 77.5%. The optimal cutoff value for the bowing of the flexor retinaculum was 3.7 mm, with a sensitivity of 71.4% and specificity of 55%. No optimum cutoff value could be identified from the receiver-operating-characteristics curves for the flattening ratio of median nerve.

Conclusion: Ultrasonographic examination of the median nerve seems to be a promising method in the diagnosis of carpal tunnel syndrome, evaluating the morphologic changes of the median nerve in patients with clinical signs and symptoms. Further studies with wider series are needed to confirm our preliminary results.

Key Words: Carpal Tunnel Syndrome, Ultrasonography, Median Nerve, Electrodiagnosis
Carpal tunnel syndrome (CTS) is the most common entrapment neuropathy, caused by the compression of the median nerve in the carpal tunnel beneath the flexor retinaculum. It is considerably more frequent in women (3:1 to about 10:1), and 50% of the cases are bilateral. The pathophysiology and pathogenesis are the subject of continued and extensive investigation; however, it has been suggested that a nonspecific tenosynovitis is the most common cause of CTS, but this remains unproven. The diagnosis of CTS is clinical and based on the proper interpretation of symptoms and signs, with confirmation by electrodiagnostic testing. The nerve conduction study is the most definite diagnostic test for CTS, being positive in 91–98% of patients with clinically diagnosed CTS; however, there are some reports that this technique itself is limited in its accuracy, particularly due to the lack of standardized diagnostic criteria for CTS.

In 1993, the Quality Standards Subcommittee of the American Academy of Neurology stated that the benefits of diagnostic imaging techniques such as magnetic resonance imaging or ultrasonography had yet to be fully established. So far, several studies with magnetic resonance imaging and ultrasonography in the diagnosis of CTS have been performed, the majority of which were published in radiology journals. However, most of those studies were reviewed to be not well designed, such that the criteria for the diagnosis of CTS were inconclusive or that the study design was improper.

The aim of this study was to evaluate the diagnostic value of ultrasonography in patients with electrophysiologically confirmed CTS.

**MATERIALS AND METHODS**

The study was carried out between December 2002 and June 2003 with full ethics committee approval. The patients in the electrophysiologically diagnosed CTS group were randomly selected among the patients referred to our electrophysiology laboratory with clinical symptoms and signs of idiopathic CTS. Of those, patients with electrophysiologic CTS were asked to participate, and only the patients accepted to participate were included in the study. Because the patients of the study were all women, the control group was also derived from the healthy female volunteers, without any complaint of CTS and with normal neurologic examination, employed in the hospital. Thus, we studied 36 wrists with electrophysiologically diagnosed CTS of 26 patients and 40 non–dominant-side, electrophysiologically normal wrists of 40 controls. All subjects gave their informed consent. A complete neurologic examination was performed in all patients and controls included in the study. Sensory abnormalities in the area of median nerve distribution, atrophy in the thenar muscles, or motor weakness were evaluated in the physical examination of CTS.

We considered electrophysiologic evaluation as the gold standard for this study. Patients with predisposing factors such as diabetes mellitus, thyroid disease, rheumatoid arthritis, myxedema, amyloidosis, acromegaly, and wrist fractures were not included in the study because of the possible variations in the involvement pattern of the median nerve due to underlying disease.

Electrophysiologic evaluations of all subjects were conducted by the same investigator with a Medtronic Keypoint 4C electrophysiologic measurement system. The following tests were performed for the diagnosis of CTS: (1) median sensory nerve conduction velocities of digits 1, 2, and 3 and of palm-to-wrist segments; and (2) median nerve distal motor latency. Patients with one or more abnormal tests defined above were diagnosed as having CTS. The tests for CTS included mixed nerve conduction study of the median nerve across the wrist–elbow segment in the symptomatic arm. Motor and sensory conductions of the ulnar nerve were also analyzed at the suspected limb in all patients to exclude a probable polyneuropathy associated with CTS. Reference values for the electrodagnostic tests obtained from previous studies of a reference population in our electrophysiology laboratory were used in the diagnosis of CTS. The values varying by >2 SD from the mean normal values were considered to be abnormal. The skin surface temperature was maintained at >31°C during the conduction studies.

All ultrasonographic examinations of carpal tunnel in patients and controls were performed with a high-resolution, real-time ultrasonograph (Logiq 9, General Electric, Milwaukee, WI) with a M12L Matrix Array, 5–13 MHz probe. A single examiner blinded to the electrophysiologic data, and without querying each participant regarding clinical status, performed the imaging in the same week of the electrophysiologic study. To ensure unbiased examination, the wrists of all participants were evaluated bilaterally in a resting neutral position with the palm up. However, nondominant limbs of 40 controls (40 wrists) and the affected sides (36 wrists) of 26 patients with CTS were taken into account for the study. Examination of each wrist took approximately 6 mins.

Quantitative ultrasonographic measurements of the median nerve, consisting of measurement of cross-sectional area (proximal nerve swelling), bowing of the flexor retinaculum, and flattening of the nerve within the carpal tunnel, described by Buchberger et al as the diagnostic triad of
CTS, were performed and compared between patients and controls. The full course of the median nerve in the carpal tunnel was assessed in both the transverse and sagittal planes. Axial images of the nerve were obtained at three levels, proximally (at radioulnar articulation), in the middle (at the level of pisiform), and distally (at the level of hook of hamate), in the carpal tunnel. The echogenic rim surrounding the nerve was included during the sonographic measurements. Anteroposterior and transverse cross-sectional diameters (R₁ and R₂) at each level were measured. The cross-sectional area was calculated indirectly in square millimeters, assuming an elliptical shape (area = \( \pi(R₁ \times R₂)/4 \)) (Fig. 1, left and right). The flattening of the nerve was determined as the ratio of the nerve’s major to minor axis at three levels. Palmar bowing of the flexor retinaculum, which is the displacement (measured in millimeters) of the retinaculum, was measured as the distance from the palmar apex of retinaculum to a straight line drawn between the tubercle of trapezium and hook of hamate bone.²

For statistical calculations, the statistical package program SPSS version 11.5 (SPSS, Chicago, IL) was used. Arithmetic mean values and standard deviations of the data were determined and a t test to compare the groups was carried out. Receiver-operating-characteristics (ROC) curves plotting the sensitivity vs. 1-specificity were used for optimal possible cutoff values of the measured ultrasonographic variables with the highest sensitivity and specificity. The sensitivity and specificity of ultrasonographic screening measures were determined in relation to the result of electrophysiologic studies.

RESULTS

One of the patients with right-sided CTS was excluded from the study after the ultrasonographic examination, which revealed anatomic variation of the bifid median nerve. Consequently, a total of 35 wrists with electrophysiologically diagnosed CTS of 25 patients and 40 non–dominant-side, electrophysiologically normal wrists of 40 controls were included in the calculations. The mean ages of the patients and controls were 45.16 ± 10.4 and 41.5 ± 6.7 yrs, respectively. The mean duration of the complaints was 3.7 ± 1.6 yrs. Paresthesias, as the most common complaint, was present in all the patients and was associated with pain in 12 of the wrists (34.3%). None of the patients had forearm pain. Thenar sensation deficit was found only in four hands (11.4%) of three patients. Atrophy of the thenar eminence or thenar muscle weakness was not detected in any patient. CTS was diagnosed bilaterally in 10 of 25 patients (40.0%). The dominant side only was affected in 11 patients (44.0%). There was no patient with absent compound muscle or nerve action potentials during the nerve conduction studies. Median nerve conduction test results of the patients with CTS and controls are presented in Table 1.

Ultrasonographic measurements of the median nerve in patients with CTS and controls were compared. The mean cross-sectional areas of the median nerve at three levels were found to be significantly increased in patients with CTS (\( P < 0.001 \)). Similarly, the mean bowing of flexor retinaculum was significantly greater in patients than in controls (\( P < 0.01 \)). There were no significant differences between patients and controls in terms of the mean flattening ratios of the median nerve at three levels (\( P > 0.05 \)). Values of ultrasonographic measurements of the median nerve in patients with CTS and controls are given in Table 2.

The sensitivity and specificity of the ultrasonographic quantitative measurements of the median nerve, formed for the cross-sectional area, the bow-
The ROC curves of the cross-sectional area of the median nerve at the proximal, middle, and distal levels represented good tests, with areas under the curve of 0.796 (P < 0.001), 0.833 (P < 0.001), and 0.806 (P < 0.001), respectively, with the middle level being the most significant (Fig. 2). The ROC curve of the bowing of the flexor retinaculum represented a fair test with an area of 0.680 (P < 0.01) (Fig. 3). No optimum cutoff values could be identified from the ROC curves of the flattening ratio of the median nerve at three levels. The ROC curves of the flattening ratio at the proximal, middle, and distal levels represented worthless tests, with areas under the curve of 0.480 (P = 0.770), 0.402 (P = 0.146), and 0.560 (P = 0.370), respectively (Fig. 4). The values of optimal cutoffs, sensitivity, and specificity of each criterion are given in Figures 2–4.

DISCUSSION

There are many conditions associated with CTS, but in some patients, an underlying disease process cannot be identified, and the description of “idiopathic CTS” is used. Some of the studies in
patients without an attributable cause for CTS show that the anatomy of the hand, wrist, and carpal tunnel may play a role for the development of CTS, encouraging the possible explanation of high prevalence of bilateral CTS. However, it was also reported that carpal canal size was not found to be a risk factor for CTS, implying other anatomic risk factors and work-related factors to be possibly responsible for the development of CTS. As a matter of fact, any condition that exerts pressure on the median nerve at the wrist can possibly cause this syndrome. Therefore, imaging techniques like ultrasonography may be very important to detect or exclude those possible pathologic changes. In this study, the morphologic changes in median nerve, rather than the carpal tunnel size and shape, were basically investigated by ultrasonography during the course of CTS.

Buchberger et al. demonstrated the enlargement and deformation of the median nerve with ultrasonography for the first time and described the characteristic features that could be found in patients with CTS. Afterward, a number of studies concerning ultrasonography in the diagnosis for CTS were carried out. However, there were some methodologic shortcomings in the designs of those studies with ultrasonography in the diagnosis of CTS, including improper control groups, inappropriate methodology, or undefined criteria for the electrodiagnosis of CTS, although electrodiagnostic tests were taken as the gold standard in many of these studies.

The most predictive feature of ultrasonography in the diagnosis of CTS reported previously was the increase in the cross-sectional area of the median nerve. The level of pisiform bone (the middle part of the carpal tunnel), which was regarded as the level of maximum swelling, was used to measure the cross-sectional area of the median nerve in most of those studies. We measured the cross-sectional area of the median nerve at three levels, proximal, middle, and distal part, of the carpal tunnel, all of which were significantly increased in patients with CTS, a finding consistent with previous studies. However, according to ROC curve results, a cross-sectional area of the median nerve at the middle of the carpal tunnel with a optimum cutoff value of 9.3 mm² yielded the highest sensitivity and specificity: 80% and 77.5%, respectively. The cutoff values for the cross-sectional area at the proximal and distal part of the carpal tunnel were 8.5 and 9.5 mm², respectively, with higher sensitivity and specificity.

Buchberger et al. reported a critical value of 10.7 mm² for the cross-sectional area of the median nerve at the pisiform bone. Sarria et al. and Nakamichi and Tachibana also measured the cross-sectional area of the median nerve at three levels under the flexor retinaculum, and they also observed a significant increase in the size of the

![Graphs show receiver-operating-characteristic (ROC) curves for median nerve cross-sectional area at three levels.](image1)

![Graph shows receiver-operating-characteristic (ROC) curve for bowing of the retinaculum.](image2)

FIGURE 2

**FIGURE 2** Graphs show receiver-operating-characteristic (ROC) curves for median nerve cross-sectional area at three levels.

**FIGURE 3** Graph shows receiver-operating-characteristic (ROC) curve for bowing of the retinaculum.
median nerve at those levels under the retinaculum. However, contrary to our study and previous studies, they found the level of distal carpal tunnel to be more sensitive than the middle or proximal levels regarding the cross-sectional area of the median nerve. Nakamichi and Tachibana\textsuperscript{13} reported a cutoff value of 13 mm\textsuperscript{2}, with a sensitivity of 57\% and specificity of 97\%, at the distal level, whereas, Sarria et al.\textsuperscript{2} found the best cutoff value as 11 mm\textsuperscript{2}, with a sensitivity of 75\% and specificity of 57.1\%, at the distal level. Wong et al.\textsuperscript{25} also measured the cross-sectional area of the median nerve at three levels of the carpal tunnel, and, using a ROC curve, a cutoff value of \(>9.8\) cm\textsuperscript{2} at the tunnel inlet was mentioned to provide a diagnostic sensitivity of 89\% and specificity of 83\%. In an other study, Duncan et al.\textsuperscript{12} found the cross-sectional area of \(>9\) mm\textsuperscript{2} to be highly predictive of CTS using two different methods. They obtained a sensitivity of 82.4\% and specificity of 97.1\% with a direct method and a sensitivity of 76.5\% and specificity of 88.2\% with an indirect method. Swen et al.\textsuperscript{11} suggested 10 mm\textsuperscript{2} as the optimal cutoff value for the median nerve cross-sectional area at the entrance of the carpal tunnel, with a sensitivity of 70\% and specificity of 63\%. We had regular results at three levels in terms of the cross-sectional area of the median nerve using our diagnostic criteria, with higher sensitivity and adequate specificity. Among the previously mentioned diagnostic criteria of CTS, the cross-sectional area of the median nerve at three levels, particularly at the middle level, was found to be the most useful criterion. Our cutoff values, sensitivity, and specificity levels were similar to those reported by Duncan et al.\textsuperscript{12} and Wong et al.\textsuperscript{25}

In some of the previous studies, cross-sectional area of the median nerve was measured at a single location.\textsuperscript{11,12} In our study, we demonstrated the enlargement of the median nerve at three levels similar to the results of Wong et al.,\textsuperscript{25} Nakamichi and Tachibana,\textsuperscript{24} and Sarria et al.\textsuperscript{2} These findings suggest a diffuse enlargement of the median nerve beneath the flexor retinaculum; however, the nerve may occasionally be within normal limits when measured at a single location, resulting in the examiner underestimating or overlooking the enlargement.

Bowing or displacement of the flexor retinaculum is another predictive feature of ultrasonography in the diagnosis of CTS. It was also mentioned as a constant finding in CTS patients in previously reported studies.\textsuperscript{2,15,17} Buchberger et al.\textsuperscript{15} reported a critical value of 3.7 mm, whereas Sarria et al.\textsuperscript{2} found a value of 2.5 mm for bowing of the flexor retinaculum. In our study, optimum cutoff value for the displacement of flexor retinaculum was 3.7 mm, with a moderate sensitivity (71.4\%) but weak specificity (55\%), as mentioned in the study of Buchberger et al.\textsuperscript{15}

Information regarding the flattening ratio of the median nerve in CTS seems to be controversial. Buchberger et al.\textsuperscript{15,17} reported the flattening ratio of the median nerve to increase in patients with CTS. However, contrary to this report, Duncan et al.\textsuperscript{12} found the flattening ratio to be highly variable and remarked it to be poorly predictive of CTS. In this study, we also did not find a significant difference in patients and controls in terms of flattening ratio measured at three levels, as Sarria et al.\textsuperscript{2} This is possibly related to the swelling of the median nerve throughout both transverse and axial axes.

In this study, we used the nerve conduction tests as the diagnostic gold standard for CTS and compared the ultrasonographic quantitative measurements of the median nerve between patients and healthy controls. Our data showed that ultrasonographic measurements of the median nerve in patients with CTS were significantly different compared with normal controls, particularly in terms of cross-sectional area of the median nerve and bowing of the flexor retinaculum but in the flattening ratio. These findings proved the presence of statistically significant morphologic changes, con-

\textbf{FIGURE 4} Graphs show receiver-operating-characteristic (ROC) curves for flattening ratio of median nerve at three levels.
firmed the possible associated anatomic changes in the median nerve of the patients with CTS. In various studies, different ultrasonographic criteria for the diagnosis of CTS have been presented. However, the lack of consensus on the optimal criteria for CTS may be attributable to the differences in equipment, patient populations, and sonographic techniques. In our study, the sensitivity and specificity of the measured quantitative variables with ultrasonography indicated the cross-sectional area of the median nerve to be the most appropriate criterion in the diagnosis of CTS.

CTS is diagnosed on the basis of the patient’s history and clinical examination, and electrodiagnostic and imaging studies may be useful in confirming the diagnostic impression. In the electrodiagnostic studies, the nerve conduction tests reflect the status of nerve fibers. However, in the diseased nerve, if there are nerve fibers unaffected by disease or injury, the test results may seem normal. Focally compressed nerves may reveal decreased conduction velocity through the area of compression, enabling localization of the disease. A normal conduction velocity, however, does not exclude the presence of compression. A major limitation of the electrodiagnostic testing by means of nerve conduction studies is the lack of standardized diagnostic criteria for CTS. In consequence, it was reported that the diagnosis was missed in 16–34% of the patients with clinically defined CTS.

Recent advances in the sophistication of ultrasonography systems and high-frequency transducers provide wide near-field view, allowing visualization of internal nerve anatomy and disease in the surrounding structures that affect the nerve, such as edema, inflammation, and tumors. In our study, we also determined anatomic variation of the median nerve with ultrasonography in one of the patients with CTS.

Compared with electrophysiologic study, ultrasonography is a noninvasive, pain-free, rapid, and well-tolerated imaging technique, requiring shorter examination time with reduced cost. It may provide more objective data.

In conclusion, ultrasonographic evaluation of the median nerve and carpal tunnel may be a useful method in the diagnosis of CTS. The promising preliminary results of this study suggest a potential for ultrasonography in the evaluation of CTS, requiring confirmation by further studies with wider series before generalizing these findings to wider populations.

REFERENCES


18. Seror P: Carpal tunnel sonography by the rheumatologist versus nerve conduction study by the neurologist. J Rheumatol 2001;28:2560–1


Association of Hemoglobin Levels, Acute Hemoglobin Decrease, Age, and Co-Morbidities with Rehabilitation Outcomes After Total Knee Replacement

ABSTRACT


Objective: A study was undertaken to assess the association of preoperative and postoperative hemoglobin levels with rehabilitation outcomes, age, and selected co-morbidities.

Design: Charts of 49 patients admitted to rehabilitation after total knee arthroplasty due to degenerative joint disease were reviewed. Outcome measures included rehabilitation admission and discharge motor FIM™ scores, motor FIM gain, and rehabilitation length of stay.

Results: Patients with higher preoperative hemoglobin levels had higher rehabilitation admission motor FIM scores ($r = 0.38, P < 0.01$) and lower motor FIM gains ($r = -0.45, P < 0.001$). Patients who had higher hemoglobin levels at rehabilitation admission had higher admission motor FIM scores and shorter length of stay. Patients with diabetes had lower preoperative hemoglobin levels. Patients with hypertension had longer length of stay. Older patients had lower admission and discharge motor FIM scores and longer length of stay.

Conclusions: Patients admitted to rehabilitation after total knee replacement have the potential to improve motor function, regardless of their preoperative and rehabilitation admission hemoglobin levels and the decrease in hemoglobin levels. However, those admitted to rehabilitation with lower hemoglobin levels, those with lower admission motor FIM scores, those who are older, and those who have hypertension may expect longer hospital stays to reach their functional goals.

Key Words: Total Knee Replacement, Hemoglobin, Functional Outcomes, Age, Co-morbidities
Anemia is a very common problem in postoperative arthroplasty patients, but postoperative anemia management is very controversial. Some surgeons prefer to use blood transfusions to keep hemoglobin levels at >10.0 g/dl, whereas others prefer a cutoff of 8.0 g/dl. However, blood transfusion has risks such as reaction, infection, increased duration of hospitalization, and fluid overload. Berman et al. reported the average transfusion was 2.6 units per arthroplasty. The major reason for a decrease in hemoglobin after total joint arthroplasty is blood loss during surgery. Yochelson et al. reported that 98% of patients after total joint arthroplasty were anemic. Andrews et al. found that hemoglobin fell 1.3 g/dl in the first postoperative week for persons with preoperative hemoglobin levels at or above 12.0 g/dl who did not receive preoperative ferrous sulfate and recommended preoperative oral iron supplement to reduce postoperative decrease in hemoglobin. Keating et al. reported the average decrease in hemoglobin was 3.85 g/dl in a group of patients undergoing unilateral total knee replacement and 5.42 g/dl in a group of patients undergoing bilateral total knee replacements. Preoperative hemoglobin was a strong predictor of transfusion risk in patients undergoing unilateral and bilateral total knee replacements. Lofthouse et al. recommended preoperative administration of epoetin alfa to reduce requirements in elderly patients having primary total arthroplasty.

Anemia can cause significant medical and functional problems. In one case study, Diamond found that severe postoperative anemia was associated with markedly reduced tolerance for therapies and prolonged length of stay (LOS). A study by Gruson et al. showed hospital LOS and mortality rates at 6 and 12 mos were significantly higher for hip fracture patients who were anemic at admission. However, Carson et al. studied 8787 consecutive hip fracture patients and found that preoperative transfusion in patients with hemoglobin levels of 8.0 g/dl or higher did not seem to influence the risk of 40- or 90-day mortality in an elderly population; 90.5% patients with hemoglobin concentration of <8.0 g/dl received blood transfusion in his sample. Kim et al. reviewed 332 patients with total hip replacement and reported that no correlation was found between preoperative hemoglobin level, level of hemoglobin at discharge from the hospital, or decrease in hemoglobin level from presurgery to discharge and number of days to discharge from the hospital.

Forrest et al. found the only factor that correlated with LOS on a surgical unit was age. The factors that correlated with the need for inpatient rehabilitation were age and diabetes. Patients with American Society of Anesthesiologists scores of 3 or 4 were most likely to require admission to a rehabilitation unit. However, Forrest et al. did not study the patient’s functional recovery and LOS in rehabilitation.

There is a paucity of literature that describes the effect of acute anemia and hemoglobin decrease from presurgery to rehabilitation admission on functional recovery among patients after total knee arthroplasty. There are no standardized guidelines for managing anemia in the rehabilitation setting. Therefore, a retrospective study was undertaken to characterize changes in hemoglobin levels from presurgery to postsurgery and assess the effect of these changes on rehabilitation outcomes.

- **Hypothesis 1**: Hemoglobin levels preoperatively and at admission from rehabilitation and the difference between these two (i.e., the decrease in hemoglobin) will be related to functional levels at admission and discharge from rehabilitation and to the amount of functional gain achieved during rehabilitation.
- **Hypothesis 2**: Low hemoglobin levels preoperatively and at rehabilitation admission and large decreases in hemoglobin will increase rehabilitation LOS.
- **Hypothesis 3**: Older patients will have lower functional levels, smaller functional gains, and longer LOS compared with younger patients.
• Hypothesis 4: Persons with selected co-morbidities (i.e., heart disease, diabetes, and hypertension) will have lower hemoglobin levels, be less functional, make smaller functional gains, and have longer LOS than persons without those co-morbidities.

METHOD
Procedure
The Institutional Review Board of Baylor College of Medicine and the Veterans Affairs Medical Center in Houston approved the protocol before collection and analysis of data. The study was conducted at the Veterans Affairs Medical Center in Houston. All data were derived retrospectively from the computerized medical records.

Sample
From 1999 through 2002, 49 patients were admitted from the orthopedic ward for inpatient rehabilitation after total knee arthroplasty due to degenerative joint disease surgery. Patients were excluded if they had a psychosis, peripheral vascular disease, rheumatoid arthritis, stroke, or postoperative surgical complications. The mean age of the sample was 64.4 yrs (SD = 11.1, median = 63, range = 47–85 yrs), and the mean LOS in rehabilitation was 10.2 days (SD = 5.4, median = 9, range = 3–28 days). Other characteristics of the sample are displayed in Table 1.

Measures
Demographic Data and Rehabilitation LOS
Age, race/ethnicity, sex, and LOS were extracted from the medical record.

| TABLE 1 Demographic characteristics of the subjects (n = 49) |
|---------------------------------|----------------|--------|
| Ethnicity                      | Number | Percentage |
| White                          | 38     | 78     |
| Black                          | 8      | 16     |
| Hispanic                      | 2      | 4      |
| Other                          | 1      | 2      |
| Sex                            |        |        |
| Male                           | 47     | 96     |
| Female                         | 2      | 4      |
| Heart disease                  |        |        |
| Yes                            | 9      | 18     |
| No                             | 40     | 82     |
| Diabetes                       |        |        |
| Yes                            | 20     | 41     |
| No                             | 29     | 59     |
| Hypertension                   |        |        |
| Yes                            | 30     | 61     |
| No                             | 19     | 39     |

Hemoglobin
Hemoglobin levels were measured presurgery and at admission to rehabilitation. The presurgery hemoglobin values were collected from the preoperative laboratory results. Rehabilitation admission hemoglobin values were obtained from the first blood draw after patients were admitted to rehabilitation. A decrease in hemoglobin was calculated as the difference between presurgery hemoglobin and rehabilitation admission hemoglobin levels.

Functional Independence
The motor subscale of the FIM™ instrument was administered at admission to and discharge from rehabilitation. Motor FIM gain was calculated as the difference between rehabilitation admission motor FIM and discharge motor FIM.

Co-morbidities
Selected co-morbidities were obtained from the history taken before surgery and included heart disease, diabetes, and hypertension. Heart disease was defined as a history of any of the following medical problems: myocardial infarction, coronary artery disease, angina, coronary artery bypass surgery, coronary angioplasty, valvular heart disease, and arrhythmia.

Statistical Analysis
Descriptive analyses were performed for each study variable, including mean values, standard deviations, median values, and ranges for continuous variables (i.e., age, hemoglobin levels and change, FIM scores and change, and LOS) and number and percentage for categorical variables (ethnicity, sex, and presence or absence of each co-morbidity).

Because many of the continuous variables were not normally distributed, nonparametric analyses were performed to assess relationships among variables. Spearman’s rho correlations were computed between measures of hemoglobin, motor FIM, rehabilitation LOS, and age. Mann-Whitney U rank tests were performed to assess the relation of the hemoglobin, FIM, and LOS measures to the three selected co-morbidities. A P value of <0.05 was considered to be statistically significant.

To determine if there was a specific level of hemoglobin below which LOS began to increase, rehabilitation hemoglobin was divided into seven groups (7.6–7.9, 8–8.0, 9–9.9, 10–10.9, 11–11.9, 12–12.9, and 13–13.4 g/dl). The mean LOS for each of these seven groups was calculated and graphed. The resulting graph was examined for a potential cut point below which LOS began to increase.
RESULTS

Hemoglobin

The mean presurgery hemoglobin level was 13.8 mg/dl (SD = 1.4, median = 13.8, range = 9.0–16.5 mg/dl). The mean hemoglobin level at admission to rehabilitation was 10.1 g/dl (SD = 1.4, median = 9.8, range = 7.6–13.4 g/dl). Presurgery hemoglobin level was significantly higher than rehabilitation hemoglobin (P < 0.05). The mean decrease in hemoglobin was 3.7 g/dl (SD = 1.5, median = 3.9, range = 0.3–6.3 g/dl), a 26.8% decrease.

Functional Independence

The mean rehabilitation admission motor FIM score was 53.6 (SD = 9.7, median = 53, range = 32–73), and the mean rehabilitation discharge motor FIM score was 76.2 (SD = 6.6, median = 77, range = 50–84). The motor FIM gain was 22.6 (SD = 9.2, median = 23, range = 3–38), a 42.2% improvement.

Co-morbidities

The presence or absence of a history of heart disease, diabetes, and hypertension are displayed in Table 1. Hypertension was most prevalent, followed by diabetes, and heart disease.

Relation of Measures of Hemoglobin, Functional Independence, Rehabilitation LOS, and Age

As displayed in Table 2, patients whose presurgery hemoglobin levels were higher tended to have higher rehabilitation admission motor FIM scores (r = 0.38, P < 0.01) and lower motor FIM gain (r = −0.45, P < 0.001). However, presurgery hemoglobin levels were unrelated to discharge motor FIM scores and LOS. Patients who had higher hemoglobin levels at rehabilitation admission tended to have higher admission motor FIM scores and shorter LOS. Rehabilitation hemoglobin was not related to discharge motor FIM scores or motor FIM gain. The amount of the decrease in hemoglobin levels was unrelated to (a) admission motor FIM score, (b) discharge motor FIM score, and (c) motor FIM gain. Older patients had lower admission and discharge motor FIM scores and longer LOS in the rehabilitation unit.

Displayed in Figure 1 are the results of categorizing rehabilitation hemoglobin and calculating the mean LOS for each category. The relation of hemoglobin level and LOS is roughly linear (Spearman’s rho = 32, P < 0.2), with LOS being higher for lower hemoglobin levels. There is no particular cut point at which LOS begins to increase abruptly. However, it should be noted that the number of patients in the more extreme groups on either end is very small.

Relation of Co-morbidities and Hemoglobin Levels, Functional Independence, and Rehabilitation LOS

Co-morbidities that were assessed included heart disease, diabetes, and hypertension (Table 3). Heart disease was unrelated to hemoglobin levels,
motor FIM scores, and LOS. Hypertension was related only to LOS, with persons with hypertension having longer LOS than those without hypertension. Diabetes was related only to preoperative hemoglobin levels, with those with diabetes having lower hemoglobin.

DISCUSSION

Total knee arthroplasty is a very common surgery. Although knee arthroplasty is generally considered to be a reliable and safe procedure, it is associated with a risk of mortality. After knee arthroplasty, anemia is a very common complication. Anemia can cause clinical symptoms such as shortness of breath, dizziness, blurred vision, hypotension, tachycardia, and headache. Our study found mean hemoglobin levels to be 13.8 g/dl pre-surgery and 10.1 g/dl at admission to rehabilitation, with an average hemoglobin decrease of 3.7 g/dl (26.1%). A 26.1% hemoglobin decrease is a significant physiological change, especially for elderly persons, and the median age for our sample was 63 yrs. The average decrease of 3.7 g/dl in our study is greater than the 1.3 g/dl reported by Andrews et al.4 but similar to the average decrease of 3.85 g/dl reported by Keating et al.5 for patients undergoing unilateral total knee replacement. However, in contrast to the 98% who were anemic at admission to rehabilitation reported by Yochelson et al., only one person (2%) had a rehabilitation admission hemoglobin of <8.0 g/dl in our study. In our sample, 74% had one (33%), two (35%), or all three (6%) of the three selected co-morbidities—hypertension, diabetes, and heart disease. However, despite the decrease in hemoglobin, all of the patients in our sample made some functional gain (median = 23 points) as measured by the change in motor FIM score between admission to and discharge from rehabilitation.

It was hypothesized (hypothesis 1) that hemoglobin levels preoperatively and at admission to rehabilitation and the difference between these two (i.e., the decrease in hemoglobin) would negatively affect functional levels at admission to and discharge from rehabilitation and the amount of functional gain achieved during rehabilitation. Hypothesis 1 was partially supported. Preoperative hemoglobin and rehabilitation hemoglobin were related to admission motor FIM scores (persons with higher hemoglobin had better admission motor FIM scores), but neither hemoglobin level was related to discharge motor FIM scores. Preoperative hemoglobin was significantly related to FIM gain, but in the opposite of the expected direction, most likely due to ceiling effects on FIM gain. Persons with lower preoperative hemoglobin levels made larger functional gains than persons with higher preoperative hemoglobin. Contrary to our hypothesis, hemoglobin decrease was unrelated to

<table>
<thead>
<tr>
<th>TABLE 3 Association of measures of hemoglobin, functional independence, and rehabilitation length of stay (LOS) with presence or absence of heart disease, hypertension, and diabetes*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preop Hemoglobin, Rank (Mean)</td>
</tr>
<tr>
<td>--------------------------</td>
</tr>
<tr>
<td>Heart disease</td>
</tr>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>Hypertension</td>
</tr>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>Diabetes</td>
</tr>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
</tr>
</tbody>
</table>

*Mann-Whitney U rank test.

bp < 0.05
admission and discharge motor FIM scores and FIM gain.

The hypothesis (hypothesis 2) that preoperative and rehabilitation admission hemoglobin levels and the decrease in hemoglobin would be related to LOS was partially supported. Persons with higher rehabilitation admission hemoglobin had shorter rehabilitation LOS. There was no relation between preoperative hemoglobin level or hemoglobin decrease and LOS. Kim et al.\textsuperscript{10} reported that neither preoperative nor discharge hemoglobin levels were related to LOS.

The hypothesis (hypothesis 3) that age would be related to functional levels at admission to and discharge from rehabilitation, functional gains, and LOS was partially supported. Older patients had lower admission and discharge FIM scores and longer LOS; however, age was unrelated to FIM gain. Forrest et al.\textsuperscript{11,12} found that age correlated with LOS on a surgical unit and age was also related to the need for rehabilitation, but they did not study the patients’ functional recovery and LOS in rehabilitation.

The hypothesis (hypothesis 4) that persons with heart disease, diabetes, or hypertension would have lower presurgery and rehabilitation admission hemoglobin levels, greater hemoglobin drop, lower rehabilitation admission and discharge motor FIM scores, smaller FIM gains, and longer LOS was partially supported. Persons with diabetes had lower admission and discharge FIM scores and LOS was partially supported. Older patients had lower admission and discharge FIM scores and longer LOS; however, age was unrelated to FIM gain. Forrest et al.\textsuperscript{11,12} found that age correlated with LOS on a surgical unit and age was also related to the need for rehabilitation, but they did not study the patients’ functional recovery and LOS in rehabilitation.

There are some limitations to these findings. First, the sample size is small, which reduces statistical power to detect associations between variables. In our sample, the lowest presurgery hemoglobin level was 9.0 g/dl and the lowest rehabilitation hemoglobin level was 7.6 g/dl, and none of the patients had clinical symptoms of anemia before surgery or at admission to rehabilitation. This limits the generalizability of our findings to persons who do not have symptomatic anemia. When rehabilitation hemoglobin levels were categorized into groups, the sample size of each group was quiet small. A larger sample size would be needed before a definitive statement could be made about a cutoff score for rehabilitation hemoglobin below which LOS begins to increase.

**CONCLUSION**

The results indicate that patients admitted to rehabilitation after total knee replacement have the potential to improve motor function, regardless of their preoperative and rehabilitation admission hemoglobin levels and the decrease in hemoglobin levels. However, those admitted to rehabilitation with lower hemoglobin levels, those with lower admission motor FIM scores, those who are older, and those who have hypertension may expect longer hospital stays to reach their functional goals. These findings suggest that providing treatment of low hemoglobin before or after surgery may be helpful in improving function and reducing LOS. This could, in turn, reduce hospital costs. However, the optimal cutoff hemoglobin indicating the need for treatment is not established, and the type of treatment (e.g., providing ferrous sulfate before surgery or blood transfusion after surgery) remains controversial.

**REFERENCES**

12. Forrest GP, Roque JM, Davoudi ST: Decreasing length of stay after total joint arthroplasty: Effect on referrals to rehabilitation units. \textit{Arch Phys Med Rehabil} 1999; 80:192–4
14. \textit{Guide for the Uniform Data Set for Medical Rehabilitation (Adult FIM), Version 4.0.} Buffalo, NY, Uniform Data System for Rehabilitation, University of Buffalo Foundation Activities, 1994
Self-Assessment Exam Questions

CME Article Number 4: X. Wang, et al.

1. According to this article, which group of patients stayed longer in the rehabilitation unit?
   A. Patients with lower rehabilitation admission hemoglobin.
   B. Patients with lower preoperative hemoglobin.
   C. Patients with a greater hemoglobin drop.
   D. Patients with diabetes.

2. In this study, older patients significantly differed from younger patients in which one of the following ways?
   A. Older patients had a greater hemoglobin drop from presurgery to admission to rehabilitation.
   B. Older patients had lower FIM scores during rehabilitation.
   C. Older patients had lower preoperative hemoglobin levels.
   D. Older patients had lower rehabilitation hemoglobin levels.

3. Patients with a greater decline in hemoglobin from pre-surgery to rehabilitation admission also had which one of the following?
   A. Lower FIM scores on admission to the rehabilitation unit.
   B. Lower FIM scores on discharge from the rehabilitation unit.
   C. Longer hospital stays.
   D. FIM gains that were not significantly different from the gains made by persons with less decline in hemoglobin.

4. What was the average hemoglobin drop after knee replacement surgery in this study?
   A. 3.7 g/dl.
   B. 1.1 g/dl.
   C. 3.85 g/dl.
   D. 2.0 g/dl.

5. From this study, it can be concluded that:
   A. Patients with co-morbidities should not be admitted for rehabilitation because they do not make meaningful progress.
   B. The transfusion cutoff point is hemoglobin level 7.6 g/dl.
   C. After knee replacement, patients should get blood transfusions to keep admission rehabilitation hemoglobin levels above 10.00 g/dl.
   D. Older patients are likely to have longer lengths of stay in the rehabilitation unit.
The answers to any essay questions must be typed or computer printed on a separate piece of paper and attached to this page.

After finishing this exam:

1. Check your answers with the correct answers on page 472.
2. Complete the CME Evaluation and Certification on the following page and mail to Bradley R. Johns, Managing Editor, CME Dept.-AAP, American Journal of Physical Medicine & Rehabilitation, 7240 Fishback Hill Lane, Indianapolis, IN 46278.
3. This educational activity must be completed and postmarked by December 31, 2006. AAP Members may complete and submit this CME Answering Sheet and the following CME Evaluation and Certification page online through the members-only section of the AAP web page at www.physiatry.org.

Answering Sheet

AMERICAN JOURNAL OF PHYSICAL MEDICINE & REHABILITATION
Vol. 84, No. 6 • June 2005

CME Article Number 4: X. Wang, et al.
Circle the appropriate answers.

1. A  B  C  D
2. A  B  C  D
3. A  B  C  D
4. A  B  C  D
5. A  B  C  D
<table>
<thead>
<tr>
<th>Question</th>
<th>Poor</th>
<th>Satisfactory</th>
<th>Outstanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was the article consistent with the stated objectives?</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Did reading this article prepare you to achieve its stated objectives?</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Is reading this article likely to enhance your professional effectiveness?</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Was the article format conducive to learning?</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Suggestions for future topics:

I, ________________________________, certify that I have met the criteria for CME credit by studying the designated materials, by responding to the self-assessment questions, by reviewing those parts of the article dealing with any question(s) answered incorrectly, and by referring to the supplemental materials listed in the references.

This educational activity is designated for 1½ category 1 CME credits.

Indicate total credits claimed: ___________ (maximum of 1½ hours)

Signature of Participant __________________________________________ Date ________________

Are you a member of the AAP?  ☐ Yes ☐ No

Do you have an individual subscription to the Journal?  ☐ Yes ☐ No

If you are not an AAP member or a Journal subscriber, have you enclosed payment of $15 with your exam?  ☐ Yes ☐ No

Payment Options:  ☐ Check (Payable to AAP) ☐ VISA ☐ Mastercard

Card # ____________________________ Exp. date ____________ Amount $ ___________

Signature _____________________________________________________________

Please provide your name as it should appear on your certificate of credit and also provide your mailing address.

Name ________________________________________________________________ Position ____________________

Institution ____________________________________________________________

Street Address _________________________________________________________

City, State, Zip _______________________________________________________

Phone ___________________________ Fax ________________________ Email _______________________

Office Use ____________________________
Assembling a Toolkit to Measure Geriatric Rehabilitation Outcomes

ABSTRACT

Objective: To gather and assemble relevant patient-based outcome measures with emphasis placed on the older adults' level of functioning and activity performance.

Design: The study was conducted in two phases. First, a set of relevant measurement instruments was identified, and their was value analyzed according to general characteristics and metrologic criteria. Second, this “toolkit” was pretested on 22 older adults with respect to the burden of assessment and the quality of the data.

Results: The toolkit includes eight measurement instruments related to mobility, basic activities of daily living, independent living, leisure, physical functioning, psychologic functioning, social functioning, and caregiver status. Participants' acceptance of the toolkit was high, with all subjects completing the toolkit in two sessions (30–90 mins each). The leisure participation and satisfaction measure was the most difficult to complete. Distributional properties were adequate to ascertain variability between subjects, except for a ceiling effect found for the social functioning measure.

Conclusion: Measurement tools that are used in combination are needed to optimize the applicability and utility of outcome results. The toolkit has the potential to become a valuable method for researchers and clinicians reporting geriatric rehabilitation outcomes.

Key Words: Aged, Outcome Assessment (Health Care), Rehabilitation, Patient Satisfaction, Geriatric Assessment
Rehabilitation focuses on the treatment of disability and the restoration of an individual’s capability to live life to its full potential. Recent literature supports the application of the principles of rehabilitation to the geriatric population and, in the last decade, programs addressing the needs of this population have grown both in number and importance, especially for the frail older population.

The multidimensional approach of rehabilitation is particularly relevant for addressing interactions between disease, body structure, body functions, and the many factors that may influence the activities of older adults.

Measuring and reporting treatment outcomes are becoming critically important in light of rapidly changing health care systems. Rehabilitation outcome research is expected to evaluate treatment efficacy and efficiency, and support clinical reasoning. Many research questions have yet to be studied. Examples include determining the optimal level of treatment intensity, and understanding the role of environmental adaptation in sustaining functional gains. Data on the effectiveness of geriatric rehabilitation are scarce, and the need for increased research activities in this area has recently been emphasized.

A major limitation is the lack of shared outcome measures. In the geriatric rehabilitation literature, a variety of disparate tools were found to be in use, a limitation identified as fragmentation of instruments across settings. This fragmentation also occurs across programs. Several well-validated tools are used, such as the Berg Balance Scale. This is not, however, the case for all measurement instruments. Equivocal outcome results are likely to occur with the use of unpublished (homemade) measures or measures that lack specificity for older adults, or the ability to detect small but meaningful changes in status over time. In May 2003, the first Canadian Consensus Workshop on Geriatric Rehabilitation was organized with the goal to develop a research agenda in the field. The need to both establish a consensus on shared outcome measures and to develop benchmarks for multisite studies was emphasized as high priorities for the coming years.

The objective of the study described here was to assemble relevant measurement instruments to assess the effectiveness of hospital-based geriatric rehabilitation and permit comparable outcome data to be shared across sites and across programs.

**CONCEPTUAL FRAMEWORK**

To measure geriatric rehabilitation outcomes, it is important to first clarify what is meant by “outcomes.” Outcomes may be considered from a variety of perspectives: at the individual level or at the program level, according to diagnosis types, or in relation to in- or outpatient settings. Outcomes may also be tailored according to the needs of those requiring and using the data. In a previous study, we developed a conceptual framework of key assessment areas for the evaluation of rehabilitation outcomes in older adults. This study was based on an older consumer approach because of the need to support patient-based outcome measures. Accordingly, emphasis was placed on the older adults’ level of functioning and activity performance. This approach excluded other variables, such as comorbidities, which would be important to healthcare professionals, or length of stay, which would be important to hospital managers. The development of the conceptual framework involved four stages, including a review of the literature on potential outcome variables semistructured interviews with older adult informants to record their thoughts about important rehabilitation outcomes, merging of data and identification of relevant outcome areas, and validation procedure using a panel of interdisciplinary experts.

The conceptual framework (Fig. 1) is composed of four primary outcome domains related to important activities for community-living older adults: mobility activities, basic activities of daily living, activities of independent living, and leisure activities. Moreover, to fully understand the level of performance in these activity domains, the framework allows for four brief evaluations of underlying functioning components, including physical functioning, psychologic functioning, social functioning, as well as factors related to the caregiver status and use of available resources. Definitions for these eight outcome domains appear in Table 1. Although the area of health status measures, such as quality of life, has begun to receive considerable attention, this domain area was not addressed separately in this conceptual framework. Activities and functioning components are, however, predominant ingredients of the older adult’s quality of life. They thus provide an operational definition to the broader concept of quality of life. The domains put...
forward in this framework resemble in terms of concepts and terminology those of existing and highly valued models in rehabilitation. Existing rehabilitation models may be appropriate for in-depth analysis of the disablement experience across the lifespan. However, they do not respond to the specific need to pinpoint outcome domains of assessment following geriatric rehabilitation. Accordingly, in the conceptual framework used for this study, attention is given to life issues that typically affect the older person, including leisure activities, social functioning, and caregiver status.

It is argued here that geriatric rehabilitation outcomes should measure not only individuals’ performance in functioning domains, such as physical functioning, but should also permit an examination of the extent to which individuals participate in valued life activities. A geriatric rehabilitation-specific toolkit incorporating each of these important assessment domains is needed.

### DESIGN

This study involved two phases of nonexperimental, developmental research activities. First, a set of relevant measurement instruments was identified. Second, the toolkit was pretested in the field to examine its feasibility with respect to the burden of assessment for the older adults as well as the quality of the data.

#### 2.1 Phase 1: Selecting a Set of Pertinent Measurement Instruments

Measurement instruments include questionnaires, tests, and indexes. These categories differ according to the approach used for gathering data, which can be self-reported, performance-based, or based on the evaluator’s judgment. Measurement instruments using any of the three categories of data gathering were admissible for inclusion in this study. To assemble the literature, the following procedure was implemented for each domain of geri-

---

**TABLE 1 Conceptual framework for the assessment of geriatric rehabilitation outcomes**

<table>
<thead>
<tr>
<th>Outcome domain</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Activity domains</strong></td>
<td></td>
</tr>
<tr>
<td>Mobility activities</td>
<td>Refer to the activities involved in moving around, inside the home and outside. Mobility includes walking or propelling a wheelchair, transferring to and from various positions (seated, standing, lying) and surfaces.</td>
</tr>
<tr>
<td>Basic activities of daily living</td>
<td>Refer to self-care activities involving bodily functions (basic needs). They include personal hygiene, dressing, eating and taking one’s medication.</td>
</tr>
<tr>
<td>Activities of independent living</td>
<td>Refer to routine activities, performed inside and outside the house, that are required to function in and maintain a home environment. They include managing personal affairs (mail, finances, using the telephone), household chores (housecleaning, cooking, laundry), community errands (shopping, appointments), and using modes of transportation.</td>
</tr>
<tr>
<td>Leisure activities</td>
<td>Refer to pleasurable activities that a person engages in to spend his or her free time. The participation in and satisfaction with leisure activities are both considered.</td>
</tr>
<tr>
<td><strong>Functioning domains</strong></td>
<td></td>
</tr>
<tr>
<td>Physical functioning</td>
<td>Refers to the physical abilities required to perform the activity domains previously defined (mobility activities, basic activities of daily living, activities of independent living, leisure activities). It includes dexterity, balance, comfort (viewed as absence of pain), and endurance</td>
</tr>
<tr>
<td>Psychological functioning</td>
<td>Refers to the emotional and cognitive components underlying the older adults’ role performance and self-image. It includes psychological well-being, attention, memory, ability to plan/organize, and ability to adapt to situations.</td>
</tr>
<tr>
<td>Social functioning</td>
<td>Refers to the interpersonal relationships that the older adult maintains with people around him or her. It includes being in contact with people through some social network (not living in isolation), being able to communicate (verbally or by writing), and having access to information.</td>
</tr>
<tr>
<td>Caregiver status and use of available resources</td>
<td>Refers to the well-being of the person who assists the older adult at home, and to the kind of resources that are being used. It includes the physical and psychological aspects involved in providing ongoing assistance to someone, and the use and availability of home services</td>
</tr>
</tbody>
</table>
atrie rehabilitation outcome. First, a comprehensive list of relevant measurement instruments was assembled from scale compendiums and review articles. Measurement instruments that were not described as targeting older adults or had not been used with this population were excluded. Second, descriptive and psychometric data related to relevant measurement instruments were extracted for in-depth review and synthesis. When the information was incomplete, additional data were sought by i) consulting the reference list of the source compendium or review article for publications concerning the instrument of interest, ii) scanning articles acquired from this list for additional references, and iii) conducting computerized searches in Medline, CINHAL, Current Contents, and PsycINFO (1967–2002) databases, entering the full name of each scale and its abbreviation as key words.

To assess the value of existing measurement instruments per domain, the data were organized according to six evaluation criteria including relevance, pragmatic aspects of administration, reliability, validity, responsiveness to change, and language.

Relevance. The first evaluation criterion was the match between the outcome domain definition and the concept the instrument was intended to measure. Priority was given to generic tools that address all facets included in the domain definition and that had been developed for older adults. Tools with a scope of content that is too narrow were not retained. For instance, depression questionnaires are limited in their capacity to address the positive aspect of psychologic functioning, and were not retained.

Pragmatic aspects of administration. The second evaluation criterion referred to technical aspects that would facilitate the administration of the measurement instrument on two grounds: firstly, in the context of a home visit, and secondly, in light of the application of a combination of tools. Seen from this perspective, the measurement instrument needed to be easy to use, take as little time as possible, and require little portable material, if any.

Reliability. Reliability reflects the amount of error inherent in a measurement. The third evaluation criterion was thus the quality of the reliability evidence for the particular instrument with respect to its proposed use with older adults. Based on classic test theory, the most common types of reliability include test-retest stability, intra- and interobserver reproducibility, and internal consistency. We considered both the magnitude of the estimates and the appropriateness of the types of reliability examined.

Validity. Validity addressed the degree of confidence one can place on the inferences drawn from the measurement instrument scores. As with reliability, this fourth evaluation criterion dealt with the quality of the evidence showing that it is pertinent and appropriate to use the particular tool for the assessment of older adults discharged from rehabilitation programs.

Responsiveness to change. This fifth evaluation criterion was defined as the ability of an instrument to measure a meaningful or clinically important change in a clinical state. Data showing that the measurement instrument can demonstrate changes occurring in community-dwelling older adults were examined. Noteworthy is the array of methods and indices that have been suggested for this purpose in the literature.

Language. The study was conducted in the province of Quebec, Canada. This country is characterized by two official languages, English and French. The last evaluation criterion was the availability of the measurement instrument in both languages. This requirement was somewhat relaxed for timed tests with no predetermined written instructions. The majority of instruments of interest were originally published in English. Accordingly, the quality of the French translation was assessed with respect to the robustness of its methodology, based on published guidelines.

The research team then held a 1-day seminar to identify the most relevant outcome measurement instruments. Three researchers (BS, JD, CW) who were not involved in the literature review independently rated each instrument on its ability to satisfy the previous requirements. For each evaluation criteria, a rating of 0, 0.5, or 1 was attributed, for none, partial, or complete fulfillment based on published evidence, and the total score was computed for each instrument (possible range 0–6). The ratings were then compared and discrepancies were discussed among the team of researchers, who also integrated the point of view of those involved in the literature review (RN, LD) and that of a rehabilitation practitioner (CA). The measurement instruments were then ordered from the most to the least performing for each domain of geriatric rehabilitation outcome. The best instruments were then selected, taking into account the complementary nature of the instruments with respect to the domain addressed.

2.2 Phase 2: Pilot Testing of Feasibility

This phase of the study examined the extent to which it was feasible to generate quality data without imposing too much of a burden on the respondents. Specifically, the question addressed was 2-fold. First, it was important to know the extent to which the respondents would accept completing individual measurement instruments and also the
toolkit as a whole. Second, it was important to ascertain that the data obtained would be useful with respect to its distributional properties. Sufficient variability in scores is important for the toolkit to be usable in outcome studies. On the other hand, floor and ceiling effect and high percentage of missing values would discourage the application of the tools.

Pretest Study Sample

The study included 22 community-living older adults recruited from the day center of a general and specialized geriatric care hospital: the Institut universitaire de gériatrie de Montréal, located in the Province of Quebec, Canada. The day center provides therapy, rehabilitation, prevention, and health promotion activities to enable seniors to continue living at home. It accommodates 25 users daily, and a total of 106 persons were registered during the study period (May–August 2003). Clients were eligible for inclusion to the study if they were known to the staff to be cognitively unimpaired and well motivated, and 60 yrs of age and over. Clients were excluded if they were unable to communicate either in French or in English. A caregiver was defined as a person close to the older adult, voluntarily providing at least 2 hrs of support per week. This definition excludes professional or paid assistance. If such a person could be identified, he or she was also invited to participate to the study. A total of 25 subjects were asked to participate; two refused, and one was unable to take part for medical reasons. The project was approved by the Research and Ethics Committees of the participating hospital. Table 2 presents subject characteristics.

Procedure

Interviews were conducted in a quiet setting at the day center facility. The participants were told the purpose of the study was to examine the acceptability and relevance of tools related to their day-to-day activities and functioning. The measurement instruments were administered on two sessions. The order of presentation was determined so as to alternate between those requiring physical and cognitive performance. Also, instruments that could induce emotional response by touching on psychologic and social issues were retained for the end. Thus, a climate of trust and comfort had more time to develop between the participant and the interviewer. Self-report questionnaires were enlarged in 14-point print and administered with the assistance of the interviewer when needed. The interviewer (IP) was a project staff member with training in research (MSc) and occupational therapy (ongoing). She developed the protocol manual under the supervision of the established researchers, and conducted practice sessions with two older adults prior to commencing the study.

TABLE 2 Characteristics of participants

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age (range)</td>
<td>79.6 (60–92)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>13 (59%)</td>
</tr>
<tr>
<td>Male</td>
<td>9 (41%)</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
</tr>
<tr>
<td>Widowed</td>
<td>11 (50%)</td>
</tr>
<tr>
<td>Married or cohabiting</td>
<td>10 (45%)</td>
</tr>
<tr>
<td>Never married</td>
<td>1 (5%)</td>
</tr>
<tr>
<td>Educational level</td>
<td></td>
</tr>
<tr>
<td>5–8 yrs</td>
<td>4 (18%)</td>
</tr>
<tr>
<td>9–13 yrs</td>
<td>13 (59%)</td>
</tr>
<tr>
<td>14–17 yrs</td>
<td>4 (18%)</td>
</tr>
<tr>
<td>18+ yrs</td>
<td>1 (5%)</td>
</tr>
<tr>
<td>Mobility device</td>
<td></td>
</tr>
<tr>
<td>Walker</td>
<td>12 (54%)</td>
</tr>
<tr>
<td>Cane</td>
<td>8 (36%)</td>
</tr>
<tr>
<td>Wheelchair</td>
<td>2 (9%)</td>
</tr>
<tr>
<td>Scooter</td>
<td>1 (5%)</td>
</tr>
<tr>
<td>None</td>
<td>7 (32%)</td>
</tr>
<tr>
<td>Economic status</td>
<td></td>
</tr>
<tr>
<td>More than sufficient income</td>
<td>6 (27%)</td>
</tr>
<tr>
<td>Sufficient income</td>
<td>15 (68%)</td>
</tr>
<tr>
<td>Low income</td>
<td>1 (5%)</td>
</tr>
<tr>
<td>Identified caregiver</td>
<td>18 (81%)</td>
</tr>
</tbody>
</table>

Descriptive statistics were used to describe the results.

RESULTS

3.1 Phase 1: Selection of a Set of Pertinent Measurement Instruments

For each measurement instrument, the results include a brief overview of the measurement instrument with respect to i) category; ii) number of items and of domains, and rating system; and iii) the mean score across the six evaluation criteria and the three raters (possible range 0–6). Tables 3–10 list the measurement instruments for each domain of assessment, from the most to the least performing.

For some measurement instruments, the items represented a broader conceptual range, compared with this study’s outcome domains. The majority of the items on the Functional Status Index,9 for example, relate to home chores, social/role activities, and hand activities (see Table 5). This index also includes items related to gross mobility and personal care. It was nevertheless associated with the activities of independent living domain because most of the content adequately covers this concept. On the other hand, some measurement instruments focus on a specific aspect of an outcome domain. Pain questionnaires, for example, were also included in the review because pain can be a major component of physical functioning in an older person.
3.2 Components of the Toolkit

This section presents the selected tools with respect to their relevance, pragmatic aspects of administration, reliability, validity, responsiveness to change, and language.

### TABLE 3 Reviewed tools for mobility activities

<table>
<thead>
<tr>
<th>Name and Source</th>
<th>Category</th>
<th>Items/Rating System</th>
<th>Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional Autonomy Measurement System (SMAF)-Mobility (Hébert et al., 1993)</td>
<td>Index</td>
<td>6 items</td>
<td>5.8</td>
</tr>
<tr>
<td>Time Up and Go (Podsalio and Richardson, 1991)</td>
<td>Test</td>
<td>1-item</td>
<td>5.7</td>
</tr>
<tr>
<td>Functional Independence Measure—Mobility and Locomotion (Granger and Hamilton, 1987)</td>
<td>Index</td>
<td>3 and 2 items</td>
<td>5.3</td>
</tr>
<tr>
<td>Gait Speed (Imms and Edhlm, 1981)</td>
<td>Test</td>
<td>1-item</td>
<td>5.2</td>
</tr>
<tr>
<td>Modified Time Up and Go (Mercer et al., 2002)</td>
<td>Test</td>
<td>1-item</td>
<td>4.7</td>
</tr>
<tr>
<td>Patient Evaluation Conference System—Physical Mobility (Harvey and Jellinek, 1981)</td>
<td>Index</td>
<td>7 items</td>
<td>3.7</td>
</tr>
<tr>
<td>Functional Status Rating System—Mobility (Forer, 1981)</td>
<td>Index</td>
<td>5 items</td>
<td>3.7</td>
</tr>
</tbody>
</table>

Note: A complete reference list of all reviewed tools can be found on the American Journal of Physical Medicine & Rehabilitation Website (http://www.amjphysmedrehab.com).

### TABLE 4 Reviewed tools for basic activities of daily living

<table>
<thead>
<tr>
<th>Name and Source</th>
<th>Category</th>
<th>Items/Rating System</th>
<th>Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional Autonomy Measurement System (SMAF)-ADL (Hébert et al., 1993)</td>
<td>Index</td>
<td>7 items</td>
<td>5.7</td>
</tr>
<tr>
<td>Barthel Index (Mahoney &amp; Barthe, 1965)</td>
<td>Index</td>
<td>10 items</td>
<td>5.7</td>
</tr>
<tr>
<td>Physical Self-Maintenance Assessment (Lawton &amp; Brody, 1969)</td>
<td>Index</td>
<td>6 items</td>
<td>4.3</td>
</tr>
<tr>
<td>Health Assessment Questionnaire (Fries, 1980)</td>
<td>Questionnaire</td>
<td>8 items (+ 6 questions about assistance and pain)</td>
<td>3.7</td>
</tr>
<tr>
<td>Rapid Disability Rating Scale-2 (Linn and Linn, 1982)</td>
<td>Index</td>
<td>18 items; 3 domains</td>
<td>3.5</td>
</tr>
<tr>
<td>LIFE-H (Fougéryrollas and Noreau, 1998)</td>
<td>Questionnaire</td>
<td>70 items; 13 domains</td>
<td>3.5</td>
</tr>
<tr>
<td>Index of Independence in ADL (Katz and Akpom, 1976)</td>
<td>Index</td>
<td>6 items</td>
<td>3.5</td>
</tr>
<tr>
<td>Old American Resources and Services Multidimensional Functional Assessment Questionnaire-ADL (Fillenbaum, 1988)</td>
<td>Questionnaire</td>
<td>15 items</td>
<td>3.3</td>
</tr>
</tbody>
</table>

ADL, Activities of Daily Living; a complete reference list of all reviewed tools can be found on the American Journal of Physical Medicine & Rehabilitation Website (http://www.amjphysmedrehab.com).

### Système de Mesure de l’Autonomie Fonctionnelle (SMAF; Functional Autonomy Measurement System)

The SMAF<sup>10</sup> is a 35-item scale based on the World Health Organization classification of disable-
ment (ICIDH-1) (World Health Organization, 1980). The SMAF measures functional ability in six areas: activities of daily living (ADL; seven items), mobility (six items), communication (three items), mental functions (five items), instrumental activities of daily living (IADL; eight items), and social functioning (six items; added in 2003). The communication and mental functions subscales were not retained for the study because other tools were selected in relation to those areas. The disability for each item is scored on a 5-point scale: independent (0), difficulty (0.5), needs supervision (1), needs help (2), and dependent (3). A higher score indicates a higher level of dependence. The SMAF must be administered by a health professional, who scores the subject after obtaining information either by questioning the subject and proxies, or by observing or directly testing the subject. Test-retest reliability coefficients (intraclass correlation coefficient and confidence interval) of the SMAF subscales are high: ADL, 0.96 (0.92–0.97); mobility, 0.91 (0.83–0.95); IADL, 0.95 (0.91–0.97); and social functioning, 0.96 (0.93–0.98).

### Table 5: Reviewed tools for activities of independent living

<table>
<thead>
<tr>
<th>Name and Source</th>
<th>Category</th>
<th>Items/Rating System</th>
<th>Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional Autonomy Measurement System (SMAF)-IADL (Hébert et al., 1993)</td>
<td>Index</td>
<td>8 items, 5-point rating scale; Subscale score</td>
<td>5.0</td>
</tr>
<tr>
<td>Instrumental Activities of Daily Living (Lawton and Brody, 1969)</td>
<td>Index</td>
<td>7 items, Guttman scaling; Overall score</td>
<td>4.0</td>
</tr>
<tr>
<td>Functional Activities Questionnaire (Pfeffer, 1984)</td>
<td>Questionnaire</td>
<td>10 items, 6-point rating scale; Overall score</td>
<td>3.0</td>
</tr>
<tr>
<td>Functional Status Index (Jette, 1980)</td>
<td>Questionnaire</td>
<td>18 items; 5 domains, 3 rating scales (5-, 4-, and 4-point); Subscale score</td>
<td>2.5</td>
</tr>
</tbody>
</table>

IADL, Instrumental Activities of Daily Living; a complete reference list of all reviewed tools can be found on the American Journal of Physical Medicine & Rehabilitation Website (http://www.amjphysmedrehab.com).

### Table 6: Reviewed tools for leisure activities

<table>
<thead>
<tr>
<th>Name and Source</th>
<th>Category</th>
<th>Items/Rating System</th>
<th>Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profil individuel en loisir- Satisfaction (Individual Leisure Profile-Satisfaction) (Carbonneau et al., 1994)</td>
<td>Questionnaire</td>
<td>24 items; 3 domains, 4-point rating scale; Subscale scores</td>
<td>4.0</td>
</tr>
<tr>
<td>Measuring Leisure Satisfaction- Satisfaction (Beard and Ragheb, 1980)</td>
<td>Questionnaire</td>
<td>24 items; 6 domains, 5-point rating scale; Subscale score</td>
<td>3.3</td>
</tr>
<tr>
<td>Mesure d’intérêt au loisir (Measuring Interest for Leisure) (Ouellette, 1991)</td>
<td>Questionnaire</td>
<td>50 items; 8 domains, 5-point rating scale; Subscale and overall scores</td>
<td>2.7</td>
</tr>
</tbody>
</table>

Note: A complete reference list of all reviewed tools can be found on the American Journal of Physical Medicine & Rehabilitation Website (http://www.amjphysmedrehab.com).

### Profil Individuel en Loisir-Satisfaction (ILP; Individual Leisure Profile-Satisfaction)

The satisfaction sections A, B, and C of the Individual Profile in Leisure comprise 25 items assessing the client’s use of free time (A), satisfaction with needs and expectations (B), and satisfaction with leisure activities (C). An ordinal scale ranging from 0 to 3 points is used for each item, and a high score out of 72 indicates a high degree of satisfaction with leisure. Internal consistency is excellent ($\alpha = 0.90$ (A), 0.92 (B), 0.89 (C)). Norms were collected for adults aged 50 and over ($n = 940$). The original French version was translated to English with a back translation method.

### Timed Up and Go (TUG)

This test measures the time taken, in seconds, to stand up from a regular arm chair, walk a 3-meter distance at a comfortable pace, turn around, return to the chair and sit down again. It requires that the person wears regular footwear and customary walking aid without assistance. It is widely used in geriatric rehabilitation. Its con-
### TABLE 7 Reviewed tools for physical functioning

<table>
<thead>
<tr>
<th>Name and Source</th>
<th>Category</th>
<th>Items/Rating System</th>
<th>Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box &amp; Block Test (Mathiowetz et al., 1985)</td>
<td>Test</td>
<td>2 items</td>
<td>5.0</td>
</tr>
<tr>
<td>Physical Performance Test (PPT) (Reuben and Siu, 1990)</td>
<td>Test</td>
<td>7 or 9 items</td>
<td>5.0</td>
</tr>
<tr>
<td>Timed Chair Rise (Rossiter-Fornoff et al., 1995)</td>
<td>Test</td>
<td>1 item</td>
<td>4.7</td>
</tr>
<tr>
<td>2-Minute Walk Test, 6-minute Walk Test (Butland et al., 1982)</td>
<td>Test</td>
<td>1 item</td>
<td>3.8</td>
</tr>
<tr>
<td>Short Physical Performance Battery for Lower Extremity Function (Gulranik et al., 1995)</td>
<td>Test</td>
<td>3 items</td>
<td>3.3</td>
</tr>
<tr>
<td>Pain rating scales</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Numeric Pain Rating Scale</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual Analogue Scale (unknown author)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>101-point Numerical Rating Scale (unknown author)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-point Box Scale (Downie et al., 1989)</td>
<td>Questionnaire</td>
<td>1 item/score</td>
<td>2.7</td>
</tr>
<tr>
<td>4 and 5-point Verbal Rating Scale (Tursky, 1976)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical Outcomes Study Pain Measures (Sherbourne, 1992)</td>
<td>Questionnaire</td>
<td>12 items</td>
<td>3.3</td>
</tr>
</tbody>
</table>

Note: A complete reference list of all reviewed tools can be found on the American Journal of Physical Medicine & Rehabilitation Website (http://www.amjphysmedrehab.com).

### TABLE 8 Reviewed tools for psychological functioning

<table>
<thead>
<tr>
<th>Name and Source</th>
<th>Category</th>
<th>Items/Rating System</th>
<th>Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Well-Being Schedule (Dupuy, 1978)</td>
<td>Questionnaire</td>
<td>18 items; 6 domains</td>
<td>4.8</td>
</tr>
<tr>
<td>3MS (Teng and Chui, 1987)</td>
<td>Test</td>
<td>15 items; 8 domains</td>
<td>4.3</td>
</tr>
<tr>
<td>Mini-mental State Evaluation (Folstein et al., 1975)</td>
<td>Test</td>
<td>Total score</td>
<td>4.2</td>
</tr>
<tr>
<td>Philadelphia Geriatric Morale Scale (Lawton, 1972)</td>
<td>Questionnaire</td>
<td>15 items; 3 domains</td>
<td>3.7</td>
</tr>
<tr>
<td>Life Satisfaction Index—revised (Neugarden et al., 1963)</td>
<td>Questionnaire</td>
<td>20 items; 5 domains</td>
<td>3.3</td>
</tr>
<tr>
<td>Memorial University of Newfoundland Scale of Happiness (MUNSH) (Kuzma and Stones, 1980)</td>
<td>Questionnaire</td>
<td>24 items; 2 domains</td>
<td>3.0</td>
</tr>
<tr>
<td>Short Portable Mental Status Questionnaire (SPMSQ) (Pfeiffer, 1975)</td>
<td>Test</td>
<td>10 items</td>
<td>3.0</td>
</tr>
<tr>
<td>Rand Mental Health Inventory (Rand Corporation and Ware, 1979)</td>
<td>Questionnaire</td>
<td>28 items; 5 domains</td>
<td>2.5</td>
</tr>
<tr>
<td>Évaluation rapide des fonctions cognitives (Rapid Assessment of Cognitive Function) (Billon, 1994)</td>
<td>Test</td>
<td>12 items; Overall score</td>
<td>2.5</td>
</tr>
<tr>
<td>Short Test of Mental Status (Kokmen et al., 1987)</td>
<td>Test</td>
<td>8 items; Maximum of 3 to 8 points per item; Overall score</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Note: A complete reference list of all reviewed tools can be found on the American Journal of Physical Medicine & Rehabilitation Website (http://www.amjphysmedrehab.com).
vergent validity has been confirmed with the Berg Balance Scale ($r = -0.72$), gait speed ($r = -0.50$ to $-0.88$), and Barthel Index ($r = -0.51$). The TUG is sensitive (87%) and specific (87%) to identify elderly prone to falls and to detect differences in performance in Parkinson disease ($t = 2.4, P = 0.035$). Norms have been published for community-dwelling elderly. Reliability coefficients are very good with cognitively unimpaired elderly individuals: interrater (ICC = 0.92–0.99; Spearman = 0.93) and test-retest (ICC = 0.89–0.99). However, lower reliability coefficients (ICC = 0.56) have been reported with cognitively impaired elderly.

**Box and Blocks Test (BBT)**

This test measures gross manual dexterity. It consists of moving 2.5-cm wooden blocks from one side of a box to the other, over a 15.2-cm partition. The score is given by the total number of blocks transferred in 1 minute. Test-retest reliability esti-

### TABLE 9 Reviewed tools for social functioning

<table>
<thead>
<tr>
<th>Name and Source</th>
<th>Category</th>
<th>Items/Rating System</th>
<th>Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMAF-Social (Pinsonneault et al., 2002)</td>
<td>Index</td>
<td>6 items; 5-point rating scale; Overall subscale score</td>
<td>4.5</td>
</tr>
<tr>
<td>Short and long versions of the Social Support Questionnaire (Sarson et al., 1983)</td>
<td>Interview-based questionnaire</td>
<td>6 items (27 items); 2 parts Listing of people + 6-point rating scale; Overall scores</td>
<td>4.3 (3.5)</td>
</tr>
<tr>
<td>Short version of the Medical Outcomes Study–Social Support Survey (Sherbourne and Stewart, 1991)</td>
<td>Interview-based questionnaire</td>
<td>6 items; 5-point rating scale Overall score</td>
<td>3.3</td>
</tr>
<tr>
<td>Social Provision Scale (Russell and Cutrona, 1984)</td>
<td>Questionnaire</td>
<td>24 items; 6 domains 4-point rating scale; Subscale and overall scores</td>
<td>3.2</td>
</tr>
<tr>
<td>Older American Resources Study—Social (Duke University, 1975)</td>
<td>Interview-based questionnaire</td>
<td>9 items; Various categorical and ordinal scales (2 to 11 levels); Overall score</td>
<td>3.0</td>
</tr>
<tr>
<td>RAND Social Health Battery (Donald and Ware, 1982)</td>
<td>Questionnaire</td>
<td>11 items; 2 domains 3 or 6-point rating scale; Subscale and overall scores</td>
<td>2.7</td>
</tr>
</tbody>
</table>

Note: A complete reference list of all reviewed tools can be found on the American Journal of Physical Medicine & Rehabilitation Website (http://www.amjphysmedrehab.com).

### TABLE 10 Reviewed tools for assistance and resources

<table>
<thead>
<tr>
<th>Name and Source</th>
<th>Category</th>
<th>Items/Rating System</th>
<th>Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Montgomery Borgatta Caregiver Burden Scale (Montgomery et al., 1985)</td>
<td>Self-report Questionnaire</td>
<td>14 items; 3 domains 5-point rating scale; Subscale scores</td>
<td>4</td>
</tr>
<tr>
<td>Outil de dépistage des aidants familiaux (Screening tool for family caregiver) (Guberman et al., 2001)</td>
<td>Interview-based questionnaire</td>
<td>12 items 4-point rating scale; Overall score</td>
<td>3.7</td>
</tr>
<tr>
<td>Burden Interview (Zarit et al., 1985)</td>
<td>Interview-based questionnaire</td>
<td>22 items 5-point rating scale; Overall score</td>
<td>3.3</td>
</tr>
<tr>
<td>Caregiver Burden Inventory (Novack and Guest, 1989)</td>
<td>Questionnaire</td>
<td>24 items; 5 domains 5-point rating scale; Subscales scores</td>
<td>2.8</td>
</tr>
<tr>
<td>Caregiver Strain Index (Robinson, 1983)</td>
<td>Questionnaire</td>
<td>13 items; 4 domains Dichotomic scoring; Overall score</td>
<td>2.7</td>
</tr>
</tbody>
</table>

Note: A complete reference list of all reviewed tools can be found on the American Journal of Physical Medicine & Rehabilitation Website (http://www.amjphysmedrehab.com).
mates are excellent with elderly adults with and without functional limitations (ICC = 0.89–0.97). Convergent validity with a comprehensive test on upper extremity performance, the TEMPA, has been reported. It has been used to monitor motor clinical changes with various diagnostic groups such as post-stroke, multiple sclerosis, and postsplinting status. Norms are available for an elderly population.

**Numeric Pain Rating Scale (NPRS)**

This scale consists of a single-item self-report scale of pain intensity. The subject circles a subjective numerical value qualifying his level of pain during the last 24 hrs (0 = least pain experienced; 10 = worst pain experienced). Test-retest reliability estimates have been reported with ICC values ranging from 0.67 to 0.96. Convergent validity was supported by correlating the instrument with the Visual Analogue Scale r = 0.79–0.95.

**General Well-Being Scale (GWBS)**

This self-report questionnaire contains 18 questions related to positive and negative aspects of well-being. The first 14 questions use an ordinal scale (0–5) and the last four questions use a paired-comparison technique with opposite descriptors at each end of an interval scale (0–10). The total score ranges from 0 to 110 points, which can be interpreted as distress for lower scores and as positive well-being for higher scores. A reliability study with adults aged 50–75 indicated good results for internal consistency (Cronbach’s α = 0.90–0.92) and test-retest stability (ICC = 0.82). The same study supported construct validity by identifying a one-factor structure explaining 50.2% of the variance. Convergent validity was supported by correlations ranging from 0.47 to 0.82 with depression and anxiety scales. One study addressed sensitivity to change after inpatient psychiatric treatment.

**The Modified Mini-Mental Scale (3MS)**

Teng & Chui proposed this modified, slightly longer, and more detailed version of the Mini Mental State Examination (2MS). Scores on the 2MS can be extracted for comparability purposes. It is used as a screening tool to assess the cognitive status of elderly individuals with respect to orientation, memory, attention, simple language, and construction. The 3MS comprises 15 items, an expanded score of 100 points, and detailed standardized testing and scoring procedures. With a cutoff score set at 77/78, sensitivity and specificity to detect dementia are estimated as 87–88% and 89–90%, respectively. Excellent test-retest and interrater reliability coefficients were found, ranging from 0.91 to 0.95. Norms with age- and education-specific reference values are available.

**Montgomery Borgatta Caregiver Burden Scale**

The short version of this self-report questionnaire measures three domains: i) objective burden (six items), ii) subjective demand burden (four items), and iii) subjective stress burden (four items). A five-point scale (1–5) is used to score each item and to compute subscale and total scores. The internal consistency of the three scales is respectable to very good, with Cronbach’s α ranging from 0.68 to 0.90. The construct validity was further confirmed by factorial analysis whereby the three-factor structure was found to explain 70% of the variance. Convergent validity was studied using the original 22-item version, using correlations and multiple regression analysis to show the association between the caregiver burden and variables such as caring tasks, personal characteristics, behaviors, and degree of assistance. The validation of a French version of the instrument is ongoing.

### 3.3 Pilot Testing of Feasibility

Participant acceptance of the toolkit was high, with all subjects who agreed to participate completing both evaluation sessions. Each session lasted 30–90 mins. A majority of subjects (n = 18, 82%) identified a caregiver. From the 18 caregivers identified by the participants, 13 (72%) completed and returned the questionnaire. Most instruments were well received by the participants. However, some sections or specific items within individual tools were found to be more difficult to complete. The Individual Leisure Profile-Satisfaction measures three domains: i) objective burden (six items), ii) subjective demand burden (four items), and iii) subjective stress burden (four items). The NPRS was repeated in the two sessions to verify that no major change occurred between sessions (physical level). Again, the data reported were retrieved from the first administration. A ceiling effect was found for the SMAF-Social (59–90% according to items). Data were missing for fewer than 10% of the subjects for the TUG (9%) and the GWBS (4.5%).

### CONCLUSION

Assembling measurement instruments into a toolkit for outcome measurement of geriatric rehabilitation was an important task to undertake. Very little is known about either the short- or long-term status of older adults in the context of reintegration into the home environment following rehabilitation. This research offers a system-
atic selection of widely used measurement tools based on a solid conceptual foundation. Complementary measurement tools are needed to optimize the applicability and value of outcome results and also to provide a comprehensive and integrated description of the older adult. Putting the new toolkit to work has the potential to produce the data needed to follow up patients discharged from rehabilitation programs and to demonstrate the effectiveness of rehabilitation services. For instance, the toolkit could be administered in conjunction with a comorbidity index to study how medical stability or complexity uniquely impacts on geriatric rehabilitation.

Assembling the toolkit was also found to be a complex task, constrained by the characteristics of existing tools. The procedure that was used to assess the instruments relied on predetermined criteria that were considered relevant for this purpose. Andresen recommended similar characteristics for criteria for assessing tools in disability outcomes research, including respondent and administration burden, hereby included within the pragmatic aspects of administration.33 Andresen also suggests that the conceptual basis is the first desirable characteristic of an outcome measure for people with disabilities, which is consistent with the approach used in this study. The strict adherence to the criteria had the effect of excluding well-validated tools such as the Barthel Index34, the FIM cognition and ADL subscales 35, or the Geriatric Depression Scale36, because they lacked specificity for the target population, or did not address all facets of the domain definition. The fact that some tools are mandated in several organizations in North America for rehabilitation reporting systems was not considered as an additional criterion in favor of inclusion. This could be considered a limitation of the methodology used in this study.

On the positive side, using well-established tools such as the Timed Up and Go15 will have the advantage of allowing easy acceptance by clinicians and researchers. It will also facilitate communication of outcomes across settings. Moreover, this strategy was more efficient than developing measurement instruments from scratch, which would have consumed a great deal of time and effort. To comply with authors’ instructions, most tools were retained as they were, with no item or scale modifications. There were two violations of this rule. The first violation was the interviewer-assisted administration of questionnaires to some frail older subjects to enhance better acceptability and reduce missing data. The second violation was the use of validated subscales, rather than complete instruments, for assessing the activity domains of mobility, basic activities of daily living, independent activities of daily living, and leisure.

On the negative side, a limitation of the methodology was the fact that existing instruments did not cover all the aspects of the outcome domains outlined by the conceptual framework. The domain related to the caregiver status and use of available resources was the most difficult to constitute, and thus had a narrower range of content assessment. Indeed, although the Montgomery Borgatta Caregiver Burden Scale31 is a valuable tool for the assessment of caregiver burden, it needs to be supplemented by another measure addressing the use and availability of home services. Such an instrument was not found at this stage of the study. Therefore, a future challenge will be to stay abreast

<table>
<thead>
<tr>
<th>Tool</th>
<th>Possible Range</th>
<th>Actual Range</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Percentage of Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMAF-ADL9</td>
<td>-21 to 0</td>
<td>-10 to 0</td>
<td>-3.68</td>
<td>2.39</td>
<td>0%</td>
</tr>
<tr>
<td>SMAF-IADL9</td>
<td>-24 to 0</td>
<td>-18 to -2</td>
<td>-10.39</td>
<td>5.46</td>
<td>0%</td>
</tr>
<tr>
<td>SMAF-mobility9</td>
<td>-18 to 0</td>
<td>-7 to 0</td>
<td>-2.05</td>
<td>1.81</td>
<td>0%</td>
</tr>
<tr>
<td>SMAF-social10</td>
<td>-18 to 0</td>
<td>-5 to 0</td>
<td>-1.00</td>
<td>1.45</td>
<td>0%</td>
</tr>
<tr>
<td>ILP11, satisfaction</td>
<td>0 to 3</td>
<td>0 to 3</td>
<td>1.48</td>
<td>0.81</td>
<td>0%</td>
</tr>
<tr>
<td>Amount of activities</td>
<td>0.40 to 3.00</td>
<td>2.15</td>
<td>0.78</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>Satisfaction with needs and expectations</td>
<td>0.79 to 2.71</td>
<td>1.95</td>
<td>0.58</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>Use of spare time</td>
<td>8.5 to 59.0</td>
<td>22.0</td>
<td>11.9</td>
<td>9%</td>
<td></td>
</tr>
<tr>
<td>TUG13</td>
<td>No limits</td>
<td>9 to 64</td>
<td>46</td>
<td>12</td>
<td>0%</td>
</tr>
<tr>
<td>BBT19</td>
<td>No limits</td>
<td>1 to 10</td>
<td>5.84</td>
<td>2.53</td>
<td>0%</td>
</tr>
<tr>
<td>NPRS22</td>
<td>0 to 10</td>
<td>35.5 to 84.6</td>
<td>61.5</td>
<td>14.0</td>
<td>4.5%</td>
</tr>
<tr>
<td>GWBS24</td>
<td>0 to 110</td>
<td>77 to 100</td>
<td>88.4</td>
<td>7.1</td>
<td>0%</td>
</tr>
<tr>
<td>3MS27</td>
<td>0 to 110</td>
<td>36 to 78</td>
<td>56.6</td>
<td>14.2</td>
<td>27.8%</td>
</tr>
<tr>
<td>Montgomery Borgatta Caregiver Burden Scale29</td>
<td>0 to 110</td>
<td>35 to 80</td>
<td>61.5</td>
<td>14.0</td>
<td>4.5%</td>
</tr>
</tbody>
</table>
of new developments in measurement instruments, especially those whose inclusion would help to fully reflect the outcomes domains important to geriatric rehabilitation. Careful monitoring of the impact of adding upcoming tools will be essential. Indeed, acceptability is a key issue if the toolkit is to be widely used; the assessments will not be administered at all if the burden of assessment is too high for the participants. It is therefore crucial to keep a minimal number of items and instruments, and a reasonable length of time to complete the toolkit, to prevent the occurrence of missing data and non valid responses.

Further steps with the toolkit will address some issues that surfaced in the pretest. One of them is to further scrutinize the appropriateness of two measurement instruments. The first instrument, the Individual Leisure Profile-Satisfaction, is challenged on the ground of acceptability issues. Indeed, the time required to complete the scale was considered disproportionate, compared with the time required for the toolkit as a whole. Moreover, several items raised uneasiness and discomfort for the respondents who did not fully comprehend the questions. It is important to note that this particular instrument had not previously been used in the context of rehabilitation outcomes. The paucity of scales targeting satisfaction and participation with leisure activities is a major obstacle, and warrants further study.

Another instrument that performed below expectation was the SMAF-Social, due to lack of variability and a ceiling effect. These problems could be explained partly by a bias inherent to the study population that was selected. All subjects were recruited from a single day center, where users systematically benefit from professional services and social support. Thus, the results would likely be different among a community-based group of frail elderly with no ongoing support services. A study is underway to verify the appropriateness of the SMAF-Social over the Social Support Questionnaire that ranked second in the ratings. It can also be argued that a smaller sample of cognitively unimpaired and well motivated participants. This limits the potential of generalizing the results. The next step of validation will examine the longitudinal applicability of the toolkit with older adults discharged from in- and outpatient rehabilitation programs (geriatric rehabilitation units and day hospitals). This study will require repeated measurements at discharge and 3 mos later. The contribution of individual tools within a multiscale outcome score will also require further study, for instance, by testing hypotheses related to the strength of association between tools related to activity and performance domains. Finally, the responsiveness to a change of tools across a longer time period needs to be researched, for instance, by contrasting a group of stable older adults in the community with a group of patients undergoing active rehabilitation treatments.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support provided by the McGill University–Université de Montréal Research Group on Integrated Services for the Older Adult (SOLIDAGE). Special thanks are extended to Marie Chevalier and the personnel of the day center of the Institut universitaire de gériatrie de Montréal, and to the study participants.

REFERENCES

14. Carbonneau H, Hélie C, Ouellet G: Évaluation d’un pro-
gramme d'éducation au loisir. Trois-Rivières: Université du Québec à Trois-Rivières, 1994


25. Stratford, Spadoni: Seven points to consider when investigating a measure’s ability to detect change. *Physiother Can* 2002:16–24


Lumbar Stabilization
Core Concepts and Current Literature, Part 1

ABSTRACT

The factors that affect lumbar stability have been an area of extensive research. The clinical application of this research in the form of lumbar stabilization exercise programs has become a common treatment of low back pain and is also increasingly used by athletes to improve performance and by the general public for health and the prevention of injury. This article includes a review of the key concepts behind lumbar stabilization. The literature regarding how those with low back pain differ in their ability to stabilize the spine from those without low back pain is discussed, and an overview of current research that assesses the benefits of a lumbar stabilization program to treat low back pain is provided.

Key Words: Low Back Pain, Spine, Exercise, Rehabilitation

The study of factors that affect lumbar stability and the treatment of low back pain (LBP) by increasing lumbar stabilization has been an area of research and investigation for >30 yrs. Today, exercise programs designed to improve lumbar stability and core strengthening are popular both to increase athletic performance and to treat pain. The purpose of this article is to review the concepts of lumbar stabilization and how instability can lead to injury and pain. We will also describe how those with LBP differ biomechanically from those without pain, and discuss the literature regarding the effectiveness of these types of programs. Because of the vast amount of literature in this area, this article is limited to a brief overview of lumbar stabilization theory and a review of clinically based research of LBP and lumbar stabilization exercises. It does not include a thorough review of spine biomechanics. A PubMed search of English language articles from 1985 to November 2004 with the key words “physical therapy,” “LBP,” “lumbar stability,” and “core strengthening” was performed. Case reports, various lumbar surgeries and stability, and articles that did not specifically describe a lumbar stabilization program were excluded. The remaining articles were reviewed by the first author and included if relevant.

Theory of Lumbar Stability
In the 1970s, researchers began to describe the concept of spinal stability. They theorized that back injury and therefore pain could be caused by the gradual degeneration of joints and soft tissue over time from repetitive micro-trauma, which was caused by poor control of spinal structures.1 As this theory
has been developed over time, it has become clear that stability is a dynamic process that includes both static positions and controlled movement. This model describes the biomechanics of the spine as similar to the biomechanics of other systems in that longevity of the components and efficiency of the system depend on the precise function of each segment. This includes both an alignment in sustained postures and movement patterns that reduce tissue strain, avoids causing trauma to the joints or soft tissue, and allows for efficient muscle action. They theorized that movement patterns that were altered by faulty strength and flexibility, fatigue from poor endurance, or abnormal neural control would eventually cause tissue damage. Tissue damage would lead to decreased stability of spinal structures, increased challenges to the already inefficient muscles, and the perpetuation of a degenerative cascade. With increasing advances in the understanding of pain and the involvement of the extensive peripheral and central processing, it is clear that the physical and emotional experience of pain is not purely a biomechanical phenomenon. However, biomechanics still play a major role in spinal pathology and pain. By understanding spine biomechanics and function and how spinal stability is altered in those with LBP, a rational approach to treatment of this condition can be developed.

Panjabi first described a model for spine stability that consists of three components. The first component is the bone and ligamentous structures that contribute to the stability of the spine. These structures provide the most stability by passive restraint toward the end of the range of motion. The interdependent ability of this system to control intersegmental stability was highlighted in a fascinating study by Cholewicki and McGill. Experienced power lifters were filmed with fluoroscopy while lifting very heavy weights. Although at first it appeared that the spine was fully flexed during lifting, the power lifters actually maintained each segment at 2–3 degrees from full flexion as an unconscious protective measure, as the spine is more easily injured in full flexion because of the excessive strain on the passive stabilizing system such as the ligaments, disk, and joint capsule. However, in one subject lifting a very heavy weight, the L2-L3 segment momentarily reached full flexion and then surpassed it by a half a degree. The spine buckled and the lifter suffered a back injury and pain. This theoretically could be the mechanism of injury when LBP is caused by much lesser loads. Spinal stability could be compromised by motor control errors or poor muscular endurance of intersegmental muscles and allow for overloading of passive tissues.

The focus of this model is the creation of the spinal stiffness and stability, but movement is just as important to the spine as stiffness. Movement of

A cadaver spine in which the bones and ligaments are intact but the muscles have been removed will buckle under about 20 pounds.

They do not provide as much support to the spine when it is in the neutral position. A cadaver spine in which the bones and ligaments are intact but the muscles have been removed will buckle under about 20 pounds. Muscles provide the support and stiffness at the intervertebral level to sustain forces commonly encountered in life. Therefore, the second component of spinal stability is the muscles that surround the spine. The greater the stiffness at each segment, the greater the stability. Very modest levels of muscle activity can create sufficiently stiff and stable joints. In usual situations, only a small amount of muscular coactivation, about 10% of maximal contraction, is needed to provide segmental stability. In a segment damaged by ligamentous laxity or disk disease, slightly more may be needed. Therefore, endurance is much more important than absolute muscle strength in most patients, although a strength reserve is needed for unpredictable activities such as a fall, a sudden load to the spine, or quick movements. In sports and heavy physical work, there are increased demands on both strength and endurance. For example, in rapid breathing caused by exertion, there is rhythmic contraction and relaxation of the abdominal wall. A fit person can support the spine with abdominal wall muscles and meet this demand at the same time, but a less fit person may not have the reserve necessary or could become injured or have pain. Muscular strength and endurance is often diminished in those with LBP.

The third component of spinal stability is the neural control system that coordinates muscle activity to respond to both expected and unexpected forces. This system must activate the correct muscles at the right time by the right amount to protect the spine from injury and also allow the desired movement. Stiffness is achieved with specific patterns of muscle activity, which differ depending on the position of the joint and the load on the spine.

Panjabi saw these three components as interdependent, and one system could compensate for deficits in another. Instability could be a result of tissue damage, making the segment more difficult to stabilize, insufficient muscular strength or endurance, or poor muscular control, and instability is usually a combination of all three.

The interdependent ability of this system to control intersegmental stability was highlighted in a fascinating study by Cholewicki and McGill. Experienced power lifters were filmed with fluoroscopy while lifting very heavy weights. Although at first it appeared that the spine was fully flexed during lifting, the power lifters actually maintained each segment at 2–3 degrees from full flexion as an unconscious protective measure, as the spine is more easily injured in full flexion because of the excessive strain on the passive stabilizing system such as the ligaments, disk, and joint capsule. However, in one subject lifting a very heavy weight, the L2-L3 segment momentarily reached full flexion and then surpassed it by a half a degree. The spine buckled and the lifter suffered a back injury and pain. This theoretically could be the mechanism of injury when LBP is caused by much lesser loads. Spinal stability could be compromised by motor control errors or poor muscular endurance of intersegmental muscles and allow for overloading of passive tissues.
the spine is required to dissipate forces and minimize energy expenditure, and a stiff and rigid spine is not the ideal, which is why surgical fusion of the lumbar spine does not cure all LBP. The neuromuscular system modulates stiffness and movement to match the demands of internal and external forces. Too much stiffness causes unnecessary energy expenditure and increased loading of spinal segments. These concepts are theoretically intriguing, and research is ongoing that attempts to quantify spinal stability so that it can better be determined what affects it and to determine how clinically significant differences in stability are.\(^9,10\)

**Muscle Function and Lumbar Stability**

A large number of muscles cross the spine, and all contribute to the modulation of lumbar stability and movement to some extent. This is a complex system consisting of deep muscles that have their origin or insertion on the lumbar vertebrae, which theoretically are responsible for the control of stiffness and intervertebral relationships, and the global muscle system that encompasses the large superficial muscles of the trunk that are the torque generators for spinal motion and handle external loads applied to the spine.\(^11\)

The focus of many lumbar stabilization programs is the deep local muscle system. The muscles most mentioned in the research and clinical literature are the multifidi, which have short intervertebral attachments and control vertebral movement during posture and spinal movement to protect the articular structures, disks, and ligaments from excessive bending strains and injury,\(^12\) and the transversus abdominis (TA), which attaches to the vertebrae through the thoracolumbar fascia and seems to stiffen the spine by increasing intraabdominal pressure.\(^13\) There is evidence that in patients with LBP, this deep stabilizing system is often very dysfunctional. In addition, more superficial muscles such as the latissimus dorsi and more superficial paraspinals and abdominal musculature have been shown to affect lumbar stiffness and stability, particularly in direction-specific movements and in carrying weights, so these muscles are also addressed in lumbar stabilization exercise programs.\(^10\)

The multifidi have deep and superficial fibers. In a study of normal subjects without LBP, it seems that the deep fibers of the multifidi, along with the TA, are the first muscles to become active when a limb is moved in response to a visual stimulus and fire independent of limb movement direction to control intervertebral movement. These fibers were also found to be active in quiet standing and subtle postural tasks such as neck flexion. The superficial fibers are also activated before the muscles that move the limb, but the timing of this

seems to be dependent on the direction the limb is moved to assist with control of spinal orientation.\(^14\) Because the fibers that contribute the most to spinal stabilization are the deepest, research studies in which surface electromyography is used to measure activity are often flawed by an inability to fully measure multifidi activation.

Studies of the multifidi have found several abnormalities in patients with LBP. Multiple imaging studies have demonstrated multifidi atrophy in patients with chronic LBP. Laasonen\(^15\) studied postoperative patients with unilateral LBP and found that paraspinals were 10–30% smaller on the affected side as compared with the unaffected side. In a study using ultrasound to compare multifidi size in patients with unilateral acute or subacute LBP vs. a control group without LBP, the side-to-side difference in normal subjects was 3% ± 4% and was 31% ± 8% in the LBP patients. This atrophy was found on the same side as symptoms and was usually confined to one vertebral level.\(^16\)

Biopsies of multifidi in patients with LBP also show abnormalities. Atrophy of type II muscle fi-

### Multiple imaging studies have demonstrated multifidi atrophy in patients with chronic LBP.

bers and internal structural changes of type I fibers, giving them a "moth eaten" appearance, is seen.\(^17\) These multifidi changes do not seem to resolve spontaneously without specific treatment, even when the pain has abated. In a study of patients undergoing surgery for lumbar disk disease, multifidi biopsies collected at time of surgery revealed type II fiber atrophy and type I fiber structural changes. These biopsies were repeated 5 yrs postoperatively, and type II fiber atrophy was still found in all patients, both those who had shown clinical improvement and those who had not. However, there was a higher percentage of type I fibers with abnormal structure in the negative outcome group as compared with the positive outcome group that showed a decrease in the percentage of type I fibers with abnormal structure.\(^17\) There is some evidence that with specific exercise training, multifidi atrophy can be reversed.\(^15,16\)

The TA is the second deep stabilizer that does not seem to function normally in patients with LBP. In an elegant study in which patients with LBP were matched with pain-free controls, the response of the abdominal muscles, lumbar multifidi, and deltoid were measured when subjects rapidly performed shoulder flexion, abduction, and extension in response to a visual stimulus. As in earlier studies that showed that the TA is the first
The quadratus lumborum is an important lateral stabilizer of the spine.

back pain. Further research will be needed to either confirm or disprove these theories.

Spinal stability does not depend on only the multifidi and TA. A cylinder of deep muscles surround the spine to provide stability, and the function of these muscles is an area of increasing research.

The quadratus lumborum is an important lateral stabilizer of the spine. It is attached to the transverse processes of the lumbar spine through the thoracolumbar fascia and therefore increases lumbar stiffness. It is a key muscle targeted in physical therapy for lumbar stabilization.

The pelvic floor also has an important role in proper muscular activation for lumbar stabilization. The pelvic floor forms the base of the abdominal cavity, so pelvic floor muscles must contract during tasks that elevate intraabdominal pressure to maintain continence and contribute to pressure increases. In subjects without LBP, strong voluntary abdominal muscle contraction caused pelvic floor muscle activity at the same intensity as maximal pelvic floor muscle effort. The pelvic floor does not simply respond to increases in intraabdominal pressure; instead, the pelvic floor muscles contract before the abdominal muscles.

As the roof of the cylinder of muscles that surround the spine and assist with stability, the diaphragm is a major contributor to intraabdominal pressure and therefore lumbar stability. For the TA to increase tension in the thoracolumbar fascia, diaphragmatic activity is required to prevent displacement of the abdominal viscera. The diaphragm contributes to IAP before the initiation of large limb movements to assist with spinal stability, and this occurs independent of the respiratory phase.

Many other trunk muscles contribute to spinal stability. These include the other abdominal muscles (internal and external obliques, rectus abdominus), the other paraspinal muscles, and the iliopsoas muscle. These muscles seem to be activated to assist with stability by direction and load-specific activity. They prevent unwanted trunk movement caused by limb movement and the acceptance of heavy loads to the trunk. For example, in the study in which rapid shoulder movement was done in response to a visual stimulus, the rectus abdominis was activated before arm extension, presumably to prevent trunk extension from occurring with arm extension. The internal oblique contracted before the deltoid in arm abduction, but neither the rectus abdominis nor the internal oblique contracted before the deltoid in arm forward flexion. It seems that the central nervous system predicts the effect movement will have on the body and plans muscle activity accordingly. The local muscles such as the deep fibers of the multifidi and the TA are activated to provide a general increase in intervertebral stiffness, and the superficial global muscles are activated in a direction-specific response to control spinal orientation. The central nervous system is able to quickly accomplish this after learning from a lifetime of movement experience.

It seems that the activity of these superficial muscles may become dysfunctional in LBP patients as well. Researchers have hypothesized that when there is dysfunction in the passive stabilizing system, global muscle may try to compensate by co-activation. Although global coactivation increases spine stability and stiffness, this comes at the price of increasing compressive load on lumbar segments and can lead to spinal pain. Also, global
muscles cannot provide control over individual spinal segments, and they have a limited ability to control shear forces compared with the deep stabilizers. They restrict spinal motion, which compromises spinal function. They may not be able to perform other intended motions and functions if they are used to try to achieve spinal stability, and they may perform additional, unwanted action as they are activated as stabilizers. For example, subjects who contract the superficial abdominal muscles to support the spine also bring into play the role of these muscles to depress the rib cage, which may lead to compromised respiratory function.24

Several studies have shown that patients with LBP have weaker back extensor muscles when measured isometrically and isokinetically compared with asymptomatic controls. LBP patients also have decreased endurance of the extensors compared with controls.12,26 They tend to have abnormal trunk flexor–to–extensor strength ratios as well.17,28 Weak lumbar spine extensions are also a risk factor for the development of LBP. This was even seen in a study of 14- to 16-yr-old children, in which those with weak spine extenders were at increased risk of having LBP at 3-yr follow-up, despite the fact that there would be relatively little degenerative spine changes expected in this young population.28

Besides intervertebral control and control of spinal orientation, lumbopelvic stability also requires control of whole-body equilibrium.24 There is a close link between lumbar stabilization and posture, balance, and proprioception of the spine. Postural control has repeatedly been found to be altered in patients with chronic LBP compared with healthy controls. Patients with LBP do more poorly than controls on one-footed–stand balance and postural stability tests.12 For example, patients with LBP had failure rates more than four times that of controls in a task that involved bilateral stance balance after periods of relaxed full lumbar flexion, the group with LBP had significantly more repositioning errors than the control group.31 Another study that compared asymptomatic controls with patients with lumbar disk herniations tested postural control and rotational proprioception. They found decreased postural control and proprioception in the group with disk herniation. This group was then treated with microdiscectomy. They found no correlation with pain relief and improvement with proprioception or postural control after surgery. Position sense improved postoperatively, but postural control did not.32 The mechanism of these deficits in LBP is thought to be secondary to multifidus dysfunction, as the multifidi have a segmental nerve supply and are highly rich in muscle spindles. The deep fibers attach to the lumbar zygapophyseal joint capsules that are also rich in proprioceptive organs.33 Bogduk34 has proposed that the predominant function of the multifidi are proprioception and kinesthetic sense. This has important treatment implications, as those with LBP may need extensive training in posture and exercise positioning because their ability to reproduce precise movements reliably is reduced.

### Ability of Exercise to Affect Lumbar Stabilization

The deficits that have been defined in lumbar stabilization in patients with LBP seem to be mostly related to muscular and neurologic function. The third component of the stabilization system, spinal structure, also plays a role, but in spinal segments with structural damage, proper muscular function seems to be able to compensate for structural deficits. That is why exercise training is the mainstay of treatment to improve stabilization.

Research in this area is often difficult. Many of the muscles tested are deep and require invasive measurement to accurately determine muscular activity. It is often unclear what degree of difference is clinically significant and what is “normal” as applied to strength, flexibility, and movement patterns. Long-term change that occurs because of exercise depends on the subject’s motivation, effort, and compliance with the program. It may be difficult to design adequate placebo controls. The natural history of back pain further complicates this research. Each episode of back pain generally has a good prognosis, and studies that look at reoccurrence need to have frequent, long-term follow-up to capture differences between groups. If only patients with persistent back pain are studied, central and peripheral pain processing problems may have a major contribution to the persistence of the pain, and symptoms may not change as muscle function improves. Despite these difficul-
ties, there is a growing body of literature that addresses the deficits outlined above and the effect of a stability program on clinical outcome. We will discuss some of the major contributions in this area below.

In regard to lumbar stabilization exercises, several issues need to be addressed. These include:

1. Can exercise reverse the changes seen in muscle mass, fiber type, strength, and endurance?
2. Can exercise change neural firing patterns so that patients with LBP can recruit their muscles in the same way as patients without a history of back problems?
3. Can exercise improve the proprioceptive and balance deficits seen in patients with persistent back pain?
4. Can patients who are suffering from pain and, in some cases, spinal damage participate in this type of exercise program?
5. If these changes do occur, does it affect the clinical outcome of patients with back pain?

It is clear that exercise can cause changes in muscle mass and increase strength and endurance. Hides et al.\textsuperscript{35} showed that lumbar stabilization exercises designed to target the multifidi can increase their mass in patients with LBP and multifidi atrophy. In a randomized, controlled trial of 39 subjects with acute first episode of unilateral LBP with multifidus atrophy, subjects were randomized to a control group that received education and regular care and a treatment group that received specific exercise training for multifidus activation and strengthening. Both groups had near resolution of LBP and return to baseline function at 4 wks. However in the control group, the multifidi remained almost unchanged at 4 and 10 wks, whereas in the treatment group, the multifidus cross-sectional area was restored to normal within 4 wks of treatment.\textsuperscript{35} This seemed to have a marked clinical affect as well. Long-term follow-up revealed that 84% of those in the control group had recurrence of LBP in a year vs. 30% of the treatment group. After 3 yrs, the control group subjects were nine times more likely to have further episodes of pain than the multifidi exercise group.\textsuperscript{36}

Other studies have shown mixed results in increasing strength and muscle mass with stabilization programs. For example, one study that compared multifidi strengthening exercises with trunk extension exercises only saw paraspinal muscle hypertrophy in the latter group. They hypothesized that this was because the stabilization exercises did not provide enough resistance to affect type II fibers that contribute most to muscle hypertrophy.\textsuperscript{37}

The second question is whether exercises can normalize neural firing patterns. Research in this area has been flawed because researchers have used surface electromyography to measure the effect of stabilization programs rather than invasive electrodes into the deeper muscles.\textsuperscript{38} Some questions regarding this have been answered, however. It is clear that subjects can learn to activate their deeper stabilizing muscles rather than more superficial muscles during exercise based on verbal and tactile cues from a physical therapist and that they can remember this for at least a week between physical therapy sessions.\textsuperscript{39} Subjects who received 5 mins of instruction to maintain neutral spine position by contracting their abdominals and trunk extensors showed less segmental spinal motion with hip flexion, extension, and biceps curl than before this instruction.\textsuperscript{40} Exercise training can also be used to change lumbar posture during standing, sitting, and walking so that it is in the neutral zone rather than excessively lordotic or kyphotic.\textsuperscript{41}

It is unclear whether the balance and proprioceptive deficits seen in patients with LBP improve with a lumbar stabilization program. Studies to date have been inconclusive or not shown changes.\textsuperscript{42} Although improvement is often seen clinically, randomized, controlled trials are still needed in this area. These exercises are also often used to improve sports performance, and improvement is thought at least in part to be secondary to this mechanism. This has not yet been proven either.

The problem of poor exercise tolerance in patients with LBP is well known to clinicians. Research has begun to study tissue loads and spinal forces with a variety of exercise to determine which stabilization exercises impose the least loading of painful tissue. In an elegant study by Kavcic et al.,\textsuperscript{42} compression forces across the L4-L5 segment in a variety of common stability exercises, such as bridging, trunk curl, quadruped exercises, and sitting on a physio-ball, were compared with the stabilizing effects of the exercises and challenges to muscles. They then ranked the exercises by stability vs. compression and abdominal vs. extensor training. This should assist clinicians in choosing exercises that impose low forces on the spine if this causes pain, score well in creating stability, and that forcefully activate the muscles that have been found too weak.\textsuperscript{10}

**Clinical Outcomes After Lumbar Stabilization Programs**

The final question is whether stability exercises change the clinical course in patients with LBP. This is an area in which much more research is needed. As discussed earlier, some research supports the theory that stability exercises prevent recurrences of back pain, with those participating in the stabilization training having nine-times-less back pain recurrences at 3 yrs than the group that did not participate.\textsuperscript{35} However, this was a small
study. A study of patients with radiculopathy showed improvement in patients who participated in stabilization exercises, although results were confounded by other treatments occurring at the same time, including injections, medication, and back school. O’Sullivan et al. showed that a lumbar stabilization program in subjects with spondyloysis and spondylolisthesis decreased pain and increased function compared with the control group at both 3- and 30-mo follow-up. Although a great deal of research has shown that exercises in general are an effective treatment of LBP, much more research is needed that specifically addresses if lumbar stabilization exercises are more effective than other types of exercise in treating back pain.

Conclusions

The factors that lead to spinal stability and the motor control of trunk muscles in people with and without back pain have been extensively studied. Much of this research is clinically applicable for those who treat back pain. The use of thoughtful exercise programs that address common deficits in patients with back pain and those at risk for recurrent episodes of LBP has been successful in small studies. In addition, research has shown that this can be accomplished by imposing low loads to the spine so that exercise is better tolerated, the risk of injury is low, and compliance is increased. It is hoped that with continued research in this area, our ability to use specific physical therapy to create individual programs to improve dynamic stability will result in better outcomes for the treatment of LBP.

REFERENCES


A 47-yr-old physiatrist was demonstrating right ulnar nerve conduction studies to residents when the responses displayed in Figure 1 were recorded.¹ Review of systems was remarkable only for intermittent tingling of the small finger, worsened by prolonged bicycle rides. Physical examination revealed no weakness in any of the upper limb muscles. Muscle stretch reflexes were 2+ and symmetrical. The tracings demonstrate recordings from the abductor digiti minimi (top four traces) and first dorsal interosseous (bottom four traces), while stimulating the ulnar nerve at the wrist, below elbow, above elbow, and axilla. Concern was raised by the residents that this might demonstrate an ulnar neuropathy in the distal forearm because the recording from the first dorsal interosseous dropped from the wrist to below elbow by almost 50% in amplitude and area. The residents were ready to make a surgical referral.

Differential diagnosis included ulnar neuropathy in the forearm (which is very rare) or Martin-Gruber anastomosis (which is quite common). Subsequent nerve conduction studies stimulating the median nerve at the elbow while recording at the first dorsal interosseous demonstrated an initially negative response, with an amplitude of 3.8 mV, consistent with a crossover in the forearm. The larger amplitudes seen with stimulation in the axilla likely represents co-stimulation of the median nerve, which is very close to the ulnar nerve in the axilla. Studies on the contralateral limb revealed essentially the same results.

The physician/patient quickly recognized that this was consistent with Martin-Gruber anastomosis and fortunately saved himself an operation for ulnar neuropathy in the forearm. His numbness has since resolved on its own.

REFERENCE