Research Report

Using Clinical Outcomes to Identify Expert Physical Therapists

Background and Purpose. Previous studies of expert physical therapists have sampled therapists based on years of clinical experience or reputation, not on their patients' clinical outcomes. The purposes of this study were to identify expert physical therapists by using patient self-reported outcomes and to describe the characteristics of clinicians whose patients with lumbar spine syndromes reported higher health-related quality of life (HRQL) following rehabilitation. **Methods.** Retrospective data were analyzed on 24,276 patients (mean age=47.8 years, SD=16, range=14–97) with lumbar spine syndromes treated by 930 physical therapists participating in the Focus On Therapeutic Outcomes database in 1999–2000. Physical therapists and staff answered questions concerning years of experience and practice setting when starting their participation in the outcomes system. Patient self-report HRQL data were collected at intake and discharge from outpatient rehabilitation. Discharge HRQL data were risk adjusted using patient characteristics. Data were aggregated by physical therapist. Risk-adjusted discharge HRQL scores were used to classify physical therapists whose patients reported mean HRQL improvement above the 90th percentile as experts and physical therapists whose patients reported mean HRQL improvement between the 45th and 55th percentiles as average. **Results.** Therapists classified as expert had fewer patients in the database than did therapists classified as average ($\overline{X}\pm SD$) (19±17 versus 29±22). Mean treatment duration was different between groups (32±11 days for the expert group versus 31±8 days for the average group). **Discussion and Conclusion.** The results challenge assumptions that extensive clinical experience is necessary to achieve superior patient outcomes, and they provide information about the relationship between therapist characteristics and patient outcomes. [Resnik L, Hart DL. Using clinical outcomes to identify expert physical therapists. *Phys Ther*. 2003;83:990–1002.]

Key Words: *Clinical competence, Expert clinicians, Low back pain, Outcome measurement, Physical therapist practice.*

Linda Resnik, Dennis L Hart

Promotion of expert practice in the health care professions is of interest to those who have a stake in improving practice and enhancing patient outcomes.[1] This interest has fueled research into the nature and acquisition of clinical expertise in medicine, nursing, and physical therapy.[2–11] As a result, theories describing dimensions of physical therapist expertise have been developed and refined,[1,10–12] and lessons learned through the study of expertise are beginning to be disseminated to educators, clinicians, and administrators.[1]

One limitation of current theoretical models of physical therapist expertise is they have been developed through research on therapists sampled solely on the basis of years of experience or reputation.[10–14] Although Rothstein[15] defined *expert therapists* as those who achieve the best clinical outcomes and it is generally presumed that there is a relationship between a practitioner's level of expertise and patient outcomes, this relationship has not been examined.[1,10–12] Although reputation is a recognized facet of being acknowledged as an expert,[16] no research has explored the relationship between reputation and measures of treatment outcome. Although the relationship between years of clinical experience and better patient outcomes is often hypothesized,[11,14,17,18] our literature search yielded no research articles supporting this hypothesis.

To our knowledge, expert clinicians selected on the basis of patient outcomes data have not been previously studied, nor have the outcomes data of experts identified through experience or reputation been analyzed. Past researchers did not evaluate whether patients managed by experts obtained better outcomes than patients managed by nonexperts, although this type of analysis was recommended.[10,11] One impediment to studies on clinical outcomes of physical therapy experts has been the lack of credible and widely accepted operational definitions of improvement following intervention[19] and the lack of a "gold standard" for patient self-report of health status[20] to assess the outcome of intervention.

Over the past decade there have been advances in the measurement of treatment outcomes using health-related quality-of-life (HRQL) instruments. Health-related quality-of-life instruments allow assessment of patients' perceptions of activities they can do, how often they do them, and the level of functional difficulty they have performing them. These instruments quantify physical, psychological, and social dimensions reflecting HRQL. As a result, HRQL data have been recommended as outcomes measures for physical therapists[21] and have been used to assess intervention outcomes in patients with a wide variety of health conditions.[22] We believed that the use of patient HRQL outcomes would provide an alternative to identifying expert physical therapists on

L Resnik, PT, PhD, OCS, is Postdoctoral Fellow, Center for Gerontology and Health Care Research, Brown University, 2 Stimson Ave, Providence, RI 02912 (USA) (linda_resnik@brown.edu). Address all correspondence to Dr Resnik.

DL Hart, PT, PhD, is Director of Consulting and Research, Focus On Therapeutic Outcomes Inc, White Stone, Va. He is also an investor in the company.

Dr Resnik provided concept/idea/research design and project management. Both authors contributed writing and data collection and analysis. Dr Hart contributed subjects and consultation (including review of manuscript before submission). The authors thank Leah Nof, PT, PhD, and Melissa Tovin, PT, PhD, for their dissertation committee work; Susan Allen, PhD, for her constructive critique of this manuscript; and Jason Roy, PhD, for his assistance with statistical analyses.

This study was approved by the Institutional Review Board for the Protection of Human Subjects, Nova Southeastern University, Fort Lauderdale, Fla.

This article was submitted November 14, 2002, and was accepted June 27, 2003.

Table 1.
Patient Characteristics (N=24,276)

Characteristic	$\overline{X}$	SD	Range	%	Frequency
Age (y)	47.8	16	14–97		
Visits	8.6	5	2–32		
Duration of intervention (d)	31	21	3–141		
Sex					
Male				42	10,120
Onset[a]					
0–7 d				9	2,100
8–14 d				11	2,589
15–21 d				10	2,403
22–90 d				25	6,015
91 d to 6 mo				11	2,798
Over 6 mo				24	8,371
No. of surgeries					
None				83	20,113
1				11	2,718
2				3	852
3				1	344
4 or more				1	249
Exercise history					
At least 3×/wk				34	8,229
1–2×/wk				26	6,225
Seldom/never				40	9,668
Region					
Middle Atlantic				11	2,654
Mountain				10	2,365
New England				6	1,478
North Central				36	8,797
Pacific				3	808
South Atlantic				15	3,711
South Central				18	4,463
Employment status					
Full duty				36	8,748
Modified work[b]				14	3,355
Employed, not working				15	3,737
Previously employed, receiving disability				2	568
Unemployed				8	2,041
Retired				20	4,872
Student				3	684
Reimbursement source					
Indemnity				10	2,557
Litigation				1	284
Medicaid				2	612
Medicare[c]				16	3,855
Patient				2	419
HMO/PPO[d]				43	10,572
Workers' compensation				19	4,648
Other				5	1,223

[a] Onset=days from onset of condition to initial evaluation.
[b] Modified work=full-time employment at a different job, restricted duty same job and restricted duty different job.
[c] Medicare=Medicare A and Medicare B.
[d] HMO=health maintenance organization, PPO=preferred provider organization.

the basis of reputation or years of experience. Thus, if patients of some therapists reported much higher gains in HRQL compared with similar patients managed by other therapists, data would support our belief that the former therapists could be considered expert. The purposes of our study were to identify expert physical therapists by using patient self-report of HRQL and to compare the characteristics of therapists whose patients reported high levels of HRQL at discharge with the characteristics of therapists whose patients reported average levels of HRQL at discharge.

Method

Design

We conducted a retrospective study by using secondary analyses of an existing database containing information on patient HRQL data as well as physical therapist years of experience, educational degree, specialty certification, sex, and practice setting, which provided an opportunity to investigate the association of these factors.[23]

Patients

The study sample was a convenience sample of 24,276 patients (Tab. 1) who received rehabilitation for low back pain syndromes in 1999 and 2000. These patients were managed by 930 therapists in 354 outpatient rehabilitation clinics in the United States (Tabs. 2 and 3). These patients were selected from a larger dataset of 57,917 patients with low back pain syndromes who were entered into the Focus On Therapeutic Outcomes Inc (FOTO) database[23,*] in the 24 months of 1999 and 2000.

The study sample was selected from the larger dataset of patients with lumbar syndromes because they were treated by physical therapists, and they completed health status surveys at both intake and discharge. To assess poten-

* Focus On Therapeutic Outcomes Inc, PO Box 11444, Knoxville, TN 37939-1444.

Table 2.
Therapist Characteristics (N=930)

	Frequency	%	X̄	SD	Range
Sex					
Male	370	40			
Educational level					
Missing	16	2			
AS	1	<1			
BS	530	57			
MS	330	36			
DPT	5	1			
PhD	4	<1			
Other	44	5			
Advanced certification[a]					
None	890	96			
OCS	26	3			
AAOMPT	5	<1			
MTC	7	<1			
AAOMPT/MTC	2	<1			
Years of clinical experience			10	8	0–43
No. of patients per therapist			26	23	2–224

[a] OCS=Orthopaedic Certified Specialist, AAOMPT=graduate of American Academy of Orthopaedic Manual Therapy Residency Program, MTC=miscellaneous certification in manual therapy, AAOMPT/MTC=therapists with both MTC and AAOMPT certification.

Table 3.
Practice Settings (N=354)

	Frequency	%	X̄	SD	Range
Practice setting					
Missing	36	10			
Payer owned	6	2			
Hospital	181	51			
Physician's office	1	<1			
Private practice	2	2			
Corporate owned	62	17			
Other	13	4			
No. of patients per practice (n=354)			69	81	8–769

tial bias introduced by incomplete data, we used chi-square and Student t tests to assess differences between patients completing only intake surveys and patients completing both intake and discharge surveys. Patients who completed intake and discharge surveys were older, had higher initial scores on the FOTO overall health status measure (OHS) and the physical functioning scale (PF-10) of the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36) questionnaire[24] (see below for measure descriptions), and had more visits over a longer duration than patients who completed intake forms only ($P<.001$). Patients with completed intake and discharge forms were more likely to exercise regularly and to receive Medicare and were less likely to receive Medicaid ($P<.001$) (Tab. 4).

The larger dataset was further reduced by limiting the selection to patients whose physical therapists had managed a minimum of 8 patients with lumbar syndromes over the 24 months. Distribution of intervention duration and number of visits was evaluated to determine if there were possible errors (eg, reporting dates of intervention incorrectly). Cases were deleted when the number of visits exceeded duration (ie, number of calendar days) of intervention because these were assumed to be data-entry errors. Extreme (1st and 99th percentiles) patient outliers of duration and visits were deleted. This eliminated patients with intervention durations of less than 2 days and greater than 141 days and patients with fewer than 2 visits and more than 32 visits, leaving a final study sample of 24,276 subjects. There are varying philosophies concerning deletion or retention of outliers. We chose to eliminate outliers for duration and visits for 2 reasons. First, we did not expect patients with only 1 visit or whose intervention duration was only 1 day to demonstrate change in HRQL, and they would not have had the opportunity to complete intake and discharge forms in only one visit. Therefore, we believed that such cases had data-entry errors. Second, we deleted outliers beyond the 99th percentile (ie, patients who had extremely long episodes of intervention or large numbers of visits) because we believed they were patients who were receiving long-term intervention (ie, in several cases up to 2 years), not short-term intervention, or these cases contained data-entry errors.

Therapists

Nine hundred thirty physical therapists employed in 354 outpatient rehabilitation facilities in 39 states managed the patients. The majority of therapists held bachelor's degrees in physical therapy (57%). Fewer held master's degrees (36%) or doctor of physical therapy (DPT) degrees (1%) (Tab. 2). Most therapists (96%) did not report any type of advanced orthopedic certification, such as an Orthopaedic Certified Specialist (OCS), certification in manual therapy (MTC), or completion of an approved manual therapy residency program (AAOMPT).

Table 4.
Comparisons of Patients Who Completed Intake And Discharge Surveys Versus Patients Who Completed Intake Surveys Only[a]

	Intake Surveys Only				Intake and Discharge Surveys				
Variable	n	%	$\overline{X}$	SD	n	%	$\overline{X}$	SD	P
Age (y)	19,413		45	16	38,503		48	17	<.001
Visits	19,413		6	5	38,503		9	6	<.001
Duration	19,413		24	32	38,503		34	34	<.001
Intake OHS	19,366		49	19	38,503		51	19	<.001
Intake PCS	19,366		33	9	38,403		33	9	.001
Intake PF-10	19,413		48	27	38,483		49	26	<.001
Sex									.124
Missing	78	0			138	0			
Male	8,225	42			16,007	42			
Female	11,110	57			22,359	58			
Onset									.001
0–7 d	1,542	8			3,237	8			
8–14 d	2,002	10			4,058	10			
15–21 d	1,836	9			3,711	10			
22–90 d	4,845	25			9,665	25			
91 d to 6 mo	2,105	11			4,461	12			
Over 6 mo	7,063	36			13,372	35			
Surgery									.15
None	15,756	81			31,235	81			
1	2,405	12			4,902	13			
2	694	4			1,379	4			
3	320	2			543	1			
4 or more	238	1			445	1			
Exercise history									<.001
Missing	132	1			247	1			
At least 3×/wk	6,272	32			13,041	34			
1–2×/wk	4,846	25			9,759	25			
Seldom/never	8,163	42			15,457	40			
Reimbursement source									<.001
Data missing	119	1			209	1			
Indemnity	2,157	11			4,075	11			
Litigation	304	2			631	2			
Medicaid	926	5			1,091	3			
Medicare A	951	5			2,385	6			
Medicare B	1,449	8			3,996	10			
Patient	530	3			814	2			
HMO	4,172	22			8,685	23			
PPO	4,208	22			7,327	19			
Workers' compensation	3,643	19			7,215	19			
Other	954	5			2,076	5			
Treating therapist[a]									<.001
No certification	16,529	97			32,598	96			
OCS only	331	2			832	3.4			
AAOMPT only	66	0.4			185	0.5			
MTC	88	0.5			272	0.8			
AAOMPT and OCS	5	0			106	0.3			
Missing	2,394	12			4,511	11.7			
Expert group	685	27			1,865	73			.001
Average group	1,477	34			2,837	66			

[a] Age=patient age, visits=number of visits in treatment episode, HMO=health maintenance organization, PPO=preferred provider organization, OCS=Orthopaedic Certified Specialist, AAOMPT=graduate of American Academy of Orthopaedic Manual Therapy Residency Program, MTC=miscellaneous certification in manual therapy.

Clinics

All interventions were performed at *outpatient rehabilitation facilities*, operationally defined as clinics where patients with neuromusculoskeletal impairments not requiring hospitalization or nursing services are managed (Tab. 3). Outpatient rehabilitation clinics commonly provide single-discipline or multidisciplinary services (eg, physical therapy or occupational therapy) and may employ a variety of licensed and unlicensed providers, including physical therapists, occupational therapists, nurses, athletic trainers, physical therapist assistants, occupational therapist assistants, or aides. The majority of clinics were hospital based (51%), with fewer being corporately owned (17%). Only a small percentage of clinics were identified as private practices, payer owned, or physicians' offices (Tab. 3).

Data Collection

FOTO data collection procedures have been described previously.[19,25] Briefly, patients completed self-report health status surveys before their initial examination and following discharge from their rehabilitation episode of care. Patient demographic data were collected at intake. Clinicians entered number of visits and intervention dates at discharge. When a clinic began collecting FOTO data, clinicians provided information on their education and years of experience using a clinician registration survey, and staff provided information describing the clinic. Clinic staff were trained in the data-collection process and were instructed to survey all patients older than 14 years of age who could communicate in English[24] at intake and discharge from rehabilitation. Data from patients and staff were entered on paper survey forms, which were submitted to FOTO, where data were checked manually for completeness. Data were entered into a computer database where computer programs checked again to ensure that data were complete and were within appropriate ranges for each variable. Data identified as incomplete or inappropriate were returned to the clinic for correction. Corrected data underwent the same data quality screening upon resubmission, improving database quality for analyses. Duration of the episode of care was calculated electronically from dates of intervention. Sixty-six percent of all patients in the FOTO dataset completed both intake and discharge surveys.

Outcome Measures

Three HRQL measures were utilized in this study: the OHS, the physical component summary (PCS) of 12-Item Short-Form Health Survey (SF-12) questionnaire,[26] and the PF-10.[24] These measures were calculated from 24 HRQL items, which have been described previously,[19,27] that were asked at intake and discharge. Items included questions from the acute version (1 week recall) of the SF-36.[22] The SF-36 has been studied in populations of patients with low back pain, and measurements of reliability and validity have been published.[22,28] Other studies have demonstrated good reliability and validity statistics of the individual scale measures of the SF-36[22,29,30] and the SF-12 summary measures.[26,31]

The OHS was chosen because it measures both mental and physical dimensions of health. Internal consistency of items in the OHS constructs with 2 or more items has been reported (α=.57–.91).[19,32] Internal consistency reliability statistics of the items of the OHS constructs[19,32] are comparable to the internal consistency reliability statistics calculated from the same items embedded in the SF-36[22] and the SF-12.[19,32] Test-retest reliability of data obtained with the OHS was good (intraclass correlation coefficient [ICC(2,1)]=.92).[27] Validity of data obtained with the OHS has not been examined, but there is evidence that an overall HRQL measure with similar items is responsive for patients receiving outpatient therapy.[33]

Overall health status scores are calculated by averaging scores from the 8 embedded HRQL constructs: general health (1 item from the SF-12),[26] physical functioning (10 items from the PF-10),[22] role physical (2 items from the SF-12),[26] bodily pain (2 items from the SF-36),[22] vitality (1 item from the SF-12),[26] mental health (2 items from the SF-12),[26] role emotional (2 items from the SF-12),[26] and social functioning (1 item from the SF-12).[26] The OHS physical functioning construct also includes 3 new questions pertinent to clients with upper-extremity impairments.[34] Scoring of item responses followed published algorithms.[22,34] Raw ordinal scores were transformed to interval scores varying from 0 to 100 for each question.[22,34] Transformed item scores were grouped by construct and averaged to obtain the score for each of the 8 OHS functional scales.

The PCS[26,31] is a summary measure representing the construct of physical functioning. Items included in the PCS were selected from the SF-36 subscales[22] of physical functioning, bodily pain, role functioning, and general health. The PCS[26] was originally designed to reduce respondent burden while maintaining measure precision of the original instrument, the SF-36.[22] The PCS score is transformed to have a mean of 50 and standard deviation of 10 in the general population.[31] Test-retest reliability and validity of data obtained with PCS measures are good.[26,31] Test-retest reliability (r=.89) was assessed using previously collected data from repeat administrations of the SF-36, which was completed by people from the general US population as part of the National Survey of Functional Health Status.[26] Test-retest reliability (ICC[2,1]=.82) for the PCS[26,31] was assessed using prospectively collected data from repeat administrations of the 24 OHS items from patients receiving outpatient rehabilitation.[27] Known group construct validity was assessed using previously collected data from the Medical Outcomes Study, an observational study of adult patients with chronic medical conditions.[26] The PCS differentiated groups of patients on seriousness of physical condition, frequency of acute symptoms, and self-reported change in physical condition.[26] The PCS was analyzed because scores from the PCS were sensitive to change over the physical therapy episode of care for patients with low back pain syndromes and PCS change scores differentiated groups of patients with low back pain syndromes who met their treatment goals compared with patients who did not meet their goals,[35] and it discriminates change in HRQL for patients receiving outpatient acute work rehabilitation compared with patients receiving work conditioning or work hardening.[19]

The PF-10 is a subscale of the SF-36 that measures physical functioning.[22] Studies support internal consistency[22,30] and test-retest reliability[22] as well as construct validity[19,22,29] and content validity[22,29] of data obtained with the PF-10 measure. Internal consistency of the PF-10 calculated from the FOTO instrument has been reported (α=.90 at intake and .91 at discharge).[32] The PF-10 was analyzed because it is responsive to change for patients with spinal syndromes receiving outpatient therapy[36,37] and discriminates change in HRQL for patients receiving outpatient therapy.[32,38] The PF-10 scores vary from 0 to 100. For all scales, the higher the score, the higher the client perceives his or her physical functioning. Each HRQL measure represents an estimate of the current level of HRQL related to physical functioning (PF-10 and PCS) or health status (OHS).

Health-related quality-of-life measures have been recommended as reliable, valid, and responsive measures for patients with low back pain.[36,38–42] Research comparing the responsiveness of the SF-36 and condition-specific instruments for patients with low back pain has shown that data obtained with the PF-10 correlate well with data obtained with condition-specific instruments.[38,43] Taylor et al[28] reported that most scales of the SF-36 were able to detect change in patients who had improved after intervention and that the PF-10 and PCS were consistently responsive in a group of patients with low back pain and sciatica. Of the 8 SF-36 scales, Patrick et al[38] reported that the physical functioning scale was the most responsive to change in a population of patients with sciatica.

Data Analysis

Risk adjustment. Risk adjustment, also called "case-mix adjustment," is a statistical process used to control effects of confounding variables seen in patient populations.[44–46] Risk adjustment considers factors other than the health care intervention or processes of care that help explain variation in patient outcomes.[44] Previous studies have demonstrated effects of certain patient characteristics on rehabilitation outcomes. For example, age,[36] depression,[36] comorbidity,[36] acuity of symptoms,[19] intake functional status,[33,36] history of prior exercise,[32] and history of surgery[36] have been associated with change in functional outcomes in patients receiving rehabilitation.

Although control of variables that could affect dependent variables is necessary before comparing patient outcomes across groups of clinicians, no consensus exists describing the optimal risk-adjustment method.[45,46] For this study, we used an approach to risk adjustment similar to the one used by Jette and Jette.[36] First, univariate analyses were used to identify possible confounding variables among patient characteristics[36] available in the 1999–2000 FOTO data. All variables shown in Table 1 were analyzed. Next, a general linear model (GLM) was developed for each of the 3 outcome measurements—the OHS,[19,27] PCS,[26,31] and PF-10.[22,24] Each patient characteristic found to be significant in univariate analyses was included in the models using a backward-deletion process. General linear models allow simultaneous control of continuous and categorical variables in the risk-adjustment process.[36,47]

The following independent variables were included in the GLMs: age, severity, sex, onset of condition, number of surgeries for condition, reimbursement, exercise history, and employment status. Age (in years) was entered as a continuous variable. Severity of the condition was a continuous variable represented by the intake OHS score. Onset of condition represents the number of days from onset of symptoms until beginning intervention (coded as 0–7 days, 8–14 days, 15–21 days, 22–90 days, 91 days to 6 months, and over 6 months). Number of surgeries represents number of surgeries for the low back (coded as none, 1, 2, 3, or greater than 3). Reimbursement was the primary source of the payment for the patient's physical therapy (coded as indemnity, litigation, Medicaid, Medicare, patient private pay, health maintenance organization [HMO] or preferred provider organization [PPO], workers' compensation, or other). Exercise history was a measurement of the patient's self-reported exercise frequency prior to the episode of physical therapy care. Exercise history was coded as at least 3 times a week, 1 to 2 times a week, or seldom/never. Employment status represents the patient's employment status at intake to physical therapy, including full-duty full-time work, modified work, employed but not working, previously employed and currently receiving disability, unemployed, retired, or student.

Table 5.
General Linear Models of Discharge Outcomes Scores[a]

Source	OHS		PCS		PF-10	
	P	Partial R^2	P	Partial R^2	P	Partial R^2
Age	<.001	.007	<.05	.013	<.05	.018
Employment	<.001	.010	<.05	.017	<.05	.013
Exercise history	<.001	.002	<.001	.002	<.01	.005
Sex	.507	<.001	<.001	.001	<.01	.003
Intake score	<.001	.286	<.001	.196	<.05	.249
Onset	<.001	.040	<.05	.035	<.05	.038
Reimbursement	<.001	.015	<.001	.013	<.01	.008
Surgery	<.001	.003	<.01	.004	<.01	.006
Corrected model	<.001	.416	<.001	.345	<.001	.421

[a] OHS=FOTO overall health status measure; PCS=SF-12 physical component summary scale; PF-10=SF-36 physical functioning scale; employment=full duty, modified work, employed, but not working, previously employed and receiving disability, unemployed, retired, student; exercise history=at least 3×/wk, 1–2×/wk, seldom/never; intake score=intake score of outcomes measure for each model (OHS, PCS, or PF-10); onset=0–7 d, 8–14 d, 15–21 d, 22–90 d, 91 d to 6 mo, over 6 mo; reimbursement=indemnity, litigation, Medicaid, Medicare A, Medicare B, patient, health maintenance organization, preferred provider organization, workers' compensation, or other; surgery=number of surgeries for this condition (none, 1, 2, 3, or 4 or more).

Responsiveness. Effect sizes (ESs)[48] and standardized response means (SRMs)[48] of the 3 outcomes measures were calculated. All measures had moderate to large ESs[49] and SRMs. Effect sizes were .83 for the OHS, .86 for the PCS, and .69 for the PF-10. Standardized response means were .87 for the OHS, .79 for the PCS, and .75 for the PF-10. Based on these calculations and the fact that the OHS, unlike the other measures, measures both mental and physical dimensions of health, we chose the OHS as the measurement for use in classifying therapists by their patients' outcomes.

Risk-adjusted outcomes. Residual scores for each discharge outcome measure were calculated after general linear modeling and saved for use in the final group analyses. *Residual scores* are the difference between actual discharge scores and the predicted scores after risk-adjusted modeling. Units of residual scores are in original scale points. We then calculated mean residual patient discharge scores for the OHS for each therapist.

The aggregated scores were used to classify therapists by their patients' outcomes. We operationally defined an *expert therapist* as a physical therapist whose mean risk-adjusted patient outcomes were above the 90th percentile and an *average therapist* as one whose patient outcomes were between the 45th to 55th percentiles. Because existing literature does not, to our knowledge, include research on using clinical outcomes to profile therapists, our selection method was based on the assumption that therapists managing patients with outcomes in the top 10% would reasonably represent expert therapists, and that therapists managing patients with outcomes in the middle 10% would reasonably represent average clinicians.

Therapist comparisons. Differences in mean number of visits per patient and mean intervention duration per patient were assessed for therapists in the average and expert groups using *t* tests. Differences between therapist years of experience, type of professional (entry-level) degree, record of advanced orthopedic certification, region of the country, and type of practice setting were assessed using chi-square tests for categorical measures and *t* tests for continuous measures.

Results

Risk Adjustment

General linear models predicted 42% of the variance in discharge OHS scores, 34% of the variance in discharge PCS scores, and 42% of the variance in PF-10 scores. In all cases, intake score and onset of condition were the largest predictors of discharge score (Tab. 5).

Group Comparisons

Table 6 presents the risk-adjusted patient outcomes data ($\overline{X}\pm SD$) for therapist groups. For each outcome measure, patients of therapists classified as expert reported higher risk-adjusted discharge scores (OHS=78±6, PCS=47±3, PF-10=80±8) compared with patients of therapists classified as average (OHS=67±5, PCS=41±3, PF-10=68±8). Therapists classified as expert or average had similar numbers of years of clinical experience (8±8 years).

Therapists classified as expert had fewer patients entered into the database, on average, than did therapists classified as average (19±17 versus 29±22, *P*=.001). Number of visits (8±2) was not different between groups. Mean treatment duration was different between groups (32±11 days for the expert group versus 31±8 days for

Table 6.
Patient Health Status Scores Per Therapist Group[a]

		OHS			PCS			PF-10		
Therapists	n	Intake	Discharge	Risk-Adjusted Discharge	Intake	Discharge	Risk-Adjusted Discharge	Intake	Discharge	Risk-Adjusted Discharge
Expert	94									
$\bar{X}$		52	78	10	34	47	5	52	80	11
SD		7	6	3	4	3	2	10	8	5
Average	94									
$\bar{X}$		52	67	0	34	41	0	51	68	−1
SD		6	5	0	3	3	2	8	8	3

[a] OHS=FOTO overall health status measure, PCS=SF-12 physical component summary scale, PF-10=SF-36 physical functioning scale. Risk-adjusted=residual scores after general linear modeling. Residuals are the differences between actual discharge scores and predicted scores after controlling independent variables. Units are actual scale points.

the average group, $P<.05$) (Tab. 7). There were no differences between groups in sex, practice setting, region of the country, or professional physical therapy degree (Tab. 7). Six therapists from the group classified as expert but no therapists in the group classified as average had advanced orthopedic clinical certification ($P<.05$) (Tab. 8). Examination of the total sample of therapists demonstrated, however, that 15 of the 38 therapists with advanced certification had mean risk-adjusted outcomes that were below the 50th percentile (Tab. 8), and 87 of the 94 therapists in the expert group did not hold advanced certification (Tab. 7).

Discussion

In contrast to previous studies that have used years of clinical experience and clinician reputation to identify clinical experts, we used risk-adjusted patient HRQL outcomes to classify clinicians as expert. Clinicians were classified as experts because their patients reported high health status on discharge from rehabilitation compared with similar patients reporting average health status and whose therapists were classified as average. To our knowledge, this is the first time risk-adjusted patient self-report health status has been used to classify clinical experts in physical therapy.

Years of Experience

The assumption that expert therapists would have many years of clinical experience has guided sampling of therapists in prior studies of expertise.[10–12] Our finding that there was no difference in years of clinical experience between groups classified as expert or average challenges the assumption that experience is necessary to achieve superior patient outcomes.[50] Although our finding was somewhat surprising, it confirms the findings of Constance,[51] who found no effect of therapists' years of experience on patient outcomes.

Advanced Certification

Results of this retrospective analysis demonstrated clinical outcomes measured by patient self-report of health status were related to practitioners' advanced orthopedic certification. Therapists in the group classified as expert were more likely than therapists in the group classified as average to hold 1 of 3 types of advanced orthopedic clinical certification—OCS, AAOMPT, or MTC. However, we believe that our findings were influenced by our analytical methods (ie, the examination of advanced certification and patient outcomes for only 20% of the therapists, rather than the entire sample). Further analysis of data for the entire sample of therapists (shown in Tab. 8) revealed that almost 40% of therapists with advanced certification had outcomes at or below the 50th percentile. This finding suggests that there was no clear relationship between advanced certification and patient outcomes.

Our finding was also influenced by the small sample size of therapists participating with the FOTO database who held clinical certification in 1999 and 2000. Some therapists might have possessed the knowledge and skills of a certified specialist but had not received their certification. The clinician registration form used by FOTO allowed therapists to indicate if they were formally certified but afforded no method of tracking therapists who have continuing education but no formal certification. Although clinician credentials are periodically updated, the difference between formal certification and possession of clinical skills might have confounded the results. Further inquiry into the effects of therapist training and experience and factors contributing to therapist effectiveness is needed.

There is a paucity of research examining the effectiveness of practitioners with advanced training. Our review of the literature yielded only one study that analyzed clinical outcomes of therapists with OCS. Hart et al[33] reported that patients of therapists with OCS had higher PF-10 scores as compared with patients of therapists without OCS. We found no research on clinical outcomes of residency graduates or graduates of miscellaneous manual therapy programs. More information is

Table 7.
Comparison of Expert and Average Therapist Groups

	Expert Therapists					Average Therapists					
	$\bar{X}$	SD	95% CI[a]	Frequency	%	$\bar{X}$	SD	95% CI	Frequency	%	P
Mean duration of patient's rehabilitation (calendar days)	32	11	29.9, 34.4			31	8	29.1, 32.6			<.05
No. of patients per therapist	19	17	15.5, 22.6			29	22	24.7, 34.7			.001
Mean no. of visits per patient	8	2	7.9, 8.9			8	2	7.9, 8.9			1
Years of clinical experience	8	8				8	8				.53
Professional education											.730
BS				58	61.7				52	55.3	
MS				32	34.0				35	37.2	
Sex											.304
Male				38	40.4				45	47.8	
Female				56	59.6				49	52.1	
Region											.765
Middle Atlantic				5	0.05				5	0.05	
Mountain				12	12.7				7	7.4	
New England				5	5.3				5	5.3	
North Central				36	38.3				40	42.6	
Pacific				3	3.2				3	3.2	
South Atlantic				19	20.2				14	14.9	
South Central				14	14.9				20	21.3	
Advanced orthopedic clinical certification[b]											<.05
OCS				3	4.3				0	0	
AAOMPT and OCS				1	1.1				0	0	
MTC				2	2.1				0	0	
Practice setting											.383
Missing				6	6.4				8	8.5	
Payer owned				4	4.3				0	0	
Hospital				60	63.8				55	58.5	
Physician's office				1	1.1				0	0	
Private practice				10	9.4				8	8.5	
Corporate owned				13	13.8				15	16.0	
Other				3	4.3				5	5.3	

[a] 95% CI=95% confidence interval.
[b] OCS=Orthopaedic Certified Specialist, AAOMPT=graduate of American Academy of Orthopaedic Manual Therapy Residency Program, MTC=miscellaneous certification in manual therapy.

needed on the outcomes of practitioners with advanced orthopedic certification.

Risk Adjustment

Our large sample size allowed analysis of many levels of categorical variables. This may have improved our explanatory models, which explained more variance than previous models using the same commercial database.[36] Intake scores were the greatest predictor of discharge scores, explaining 29% of the variance in OHS scores. The 2 other largest predictor variables in our model of OHS were onset of condition (explaining 4%) and reimbursement source (explaining 1.5%). Inclusion of additional predictor variables, such as socioeconomic status,[36] fear and avoidance beliefs,[52] and patient education level,[36] might have enhanced the explanatory power of our study. Further research to identify characteristics that predict outcome for patients with low back pain and other common diagnostic categories is recommended.

Patients Per Therapist

Therapists from the group classified as expert had fewer patients in the database than therapists from the group classified as average. One explanation for this might be a smaller caseload for these expert therapists. A testable hypothesis is that therapists who manage fewer patients per day may spend more time with each patient, a factor that may be associated with better

Table 8.
Number of Therapists by Percentile Ranking of Mean Risk-Adjusted Overall Health Status Measure Scores With and Without Orthopedic Certification[a]

Percentile	No Certification	OCS Only	AAOMPT Only	MTC	AAOMPT and OCS
0	89	2			
10	89	3	1	1	
30	89	3		1	
40	89	2	1	1	1
50	89	3			
60	89	3	1	1	
70	89	2	1	1	1
80	89	3		1	
90	89	3	1	1	
100	89	2			
Total no. of therapists	890	26	5	7	2

[a] OCS=Orthopaedic Certified Specialist, AAOMPT=graduate of American Academy of Orthopaedic Manual Therapy Residency Program, MTC=miscellaneous certification in manual therapy.

patient outcomes. We found no prior research to support or refute this hypothesis.

Limitations

Others have discussed the limitations of using retrospective data to answer questions that have not been determined a priori, including problems with missing observations, data control, and patient selection bias that might threaten external validity of commercial databases.[36,53] Researchers cannot exercise control of data-collection quality with retrospective data. There is a possibility of clinician and patient error in completing items on HRQL forms in retrospective and prospective studies. Because data were retrospective, there was no method to monitor the training process of data collection or assess level of adherence to FOTO guidelines for data quality. FOTO checks data quality manually prior to entry and via computer once data are entered into the database. Because of this data auditing, entered data were complete and values were within expected ranges for each variable necessary for calculation of HRQL estimates. However, there was no process to monitor which patients were selected for data collection in the clinic and no method to assess the reason for loss to follow-up.

We included only patients who had completed intake and discharge forms in our analyses because we were studying patient outcomes. The percentage of completed FOTO forms (66%) surpassed previous reports of 64% (1996 database)[37] and 28% (1993–1994 database).[36] Differences between patients with complete forms and those with incomplete forms were evaluated, and potential effects of missing follow-up data on our analysis were considered. Because the rate of form completion at discharge was lower for patients of physical therapists in the group classified as average, differential rates of inclusion in data analysis may have biased results toward the null (ie, minimizing differences in patient outcomes between average and expert groups). However, direction of potential bias cannot be accurately predicted. Although some patients who did not complete an episode of care may have left physical therapy because of failure to improve or because their condition was worsening, others may have ended therapy early because they were feeling better. The latter supposition is supported by our data because patients failing to complete data collection were younger and had higher initial OHS and PF-10 scores. Patients without discharge surveys simply may have returned to their physician for a follow-up visit and not kept a subsequent physical therapy appointment. This problem illustrates clinical reality and the need for quality control and data auditing for a clinical database used for research.

Another limitation of utilizing previously collected data is the researcher has no control over the choice of data elements. We therefore had no method to obtain information that might have allowed classification of patients according to movement signs and symptoms or specific diagnosis thought to affect outcome.[54–58] Classification schemes improve homogeneity of samples for more powerful comparisons.[54,55,59]

There may be limitations in the exclusive use of intake and discharge HRQL measurements to measure benefits of treatment as these measures may not include all areas of importance to the clinician and patient.[1] It is possible that aspects of physical therapy intervention, such as patient education, have lifelong health effects that cannot be captured with HRQL measurements or assessed at the time of discharge.[1] In our study, there was no method to track long-term HRQL outcomes within the existing database. Although these measurements may not reflect the actual long-term effect of physical therapy intervention, other research has shown that discharge scores of the SF-36 are good indicators of long-term outcomes for patients with low back pain.[60]

There are limits to the generalizability of the conclusions from our study to a broader population of physical therapists or patients. We do not know if our sample is representative of physical therapy practices in the United States.

Our sample included only patients of physical therapists who participated in the FOTO database. No effort was made to contact clinicians who do not participate in this database. Because practices elect to participate with FOTO for a variety of reasons,[19,36,61] there are threats to external validity of the findings.

Although we considered our selection criteria (90th percentile outcomes and above) to be a reasonable way to classify therapists as experts, we recognize other researchers might have chosen a different sample to represent expert therapists (such as the top quartile) or limited selection to a smaller group (such as the top 5%). Others might have used assessment of clinically important change concerning HRQL measures to select groups for comparison. Others might have used assessment of clinically important change concerning HRQL measures to select groups for comparison. There is debate in the literature[62,63] as to what represents clinically important change in HRQL scores. Effects of different selection techniques await future analyses.

We classified and analyzed only 2 groups of therapists in this study: expert therapists and average therapists. We did not include a third group of below-average therapists. Selection of only 2 groups of therapists limited information available for analysis and minimized information about characteristics of therapists whose patients had below-average outcomes.

Because of these limitations, our findings should be considered as hypothesis generating rather than hypothesis testing.[36] Our study provides new information concerning the relationship between therapist characteristics and patient outcomes and introduces a method for using outcomes data to gauge therapist expertise. Although an observational study of this nature does not provide the type of rigor afforded by a clinical trial, our design provides data-based hypotheses that can be tested in future clinical trials.

Conclusion

To our knowledge, our study is the first to classify therapists as expert or average by aggregating risk-adjusted patient self-report HRQL scores for their patients. The group of therapists classified as expert could not be distinguished from therapists classified as average based on their years of clinical experience, sex, or professional degree. Although expert therapists in our sample were more likely than average physical therapists to possess an advanced orthopedic certification, the majority of physical therapists in this group did not have advanced clinical certification.

References

1 Jensen GM, Gwyer J, Hack LM, Shepard KF. *Expertise in Physical Therapy Practice.* Boston, Mass: Butterworth-Heinemann; 1999.

2 Benner PE. *From Novice to Expert: Excellence and Power in Clinical Nursing Practice.* Menlo Park, Calif: Addison-Wesley Publishing; 1984.

3 Benner PE, Tanner CA, Chesla CA, et al. *Expertise in Nursing Practice: Caring, Clinical Judgment, and Ethics.* New York, NY: Springer Publishing; 1996.

4 Norman GR, Rosenthal D, Brooks LR, et al. The development of expertise in dermatology. *Arch Dermatol.* 1989;125:1063–1068.

5 Muzzin LJ, Norman GR, Feightner JW, et al. Expertise in recall of clinical protocols in two specialty areas. *Proc Annu Conf Res Med Educ.* 1983;22:122–127.

6 Orme L, Maggs C. Decision-making in clinical practice: how do expert nurses, midwives and health visitors make decisions? *Nurse Educ Today.* 1993;13:270–276.

7 Embrey DG, Guthrie MR, White OR, Dietz J. Clinical decision making by experienced and inexperienced pediatric physical therapists for children with diplegic cerebral palsy. *Phys Ther.* 1996;76:20–33.

8 Patel VL, Groen GJ, Arocha JF. Medical expertise as a function of task difficulty. *Mem Cognit.* 1990;18:394–406.

9 Chi MT, Glaser R, Farr MJ. *The Nature of Expertise.* Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

10 Jensen GM, Shepard KF, Hack LM. The novice versus the experienced clinician: insights into the work of the physical therapist. *Phys Ther.* 1990;70:314–323.

11 Jensen GM, Shepard KF, Gwyer J, Hack LM. Attribute dimensions that distinguish master and novice physical therapy clinicians in orthopedic settings. *Phys Ther.* 1992;72:711–722.

12 Jensen GM, Gwyer J, Shepard KF, Hack LM. Expert practice in physical therapy. *Phys Ther.* 2000;80:28–43; discussion 44–52.

13 Embrey DG, Yates L. Clinical applications of self-monitoring by experienced and novice pediatric physical therapists. *Pediatric Physical Therapy.* 1996;8:156–164.

14 Shepard KF, Hack LM, Gwyer J, Jensen GM. Describing expert practice in physical therapy. *Qual Health Res.* 1999;9:746–758.

15 Rothstein JM. Foreword II. In: Jensen GM, Gwyer J, Hack LM, Shepard KF. *Expertise in Physical Therapy Practice.* Boston, Mass: Butterworth-Heinemann; 1999:xviii.

16 Sternberg R, Horvath J. A prototype view of expert teaching. *Educational Researchers.* 1995;24:9–17.

17 Rivett D, Higgs J. Experience and expertise in clinical reasoning. *New Zealand Journal of Physiotherapy.* 1995;23(1):16–21.

18 Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: theory and implication. *Acad Med.* 1990;65:611–621.

19 Hart DL. The power of outcomes: FOTO Industrial Outcomes Tool, initial assessment. *Work.* 2001;16:39–51.

20 Binkley JM, Stratford PW, Lott SA, et al. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. *Phys Ther.* 1999;79:371–383.

21 Jette AM. Using health-related quality of life measures in physical therapy outcomes research. *Phys Ther.* 1993;73:528–537.

22 Ware JE Jr, Snow KK, Kosinksi M, Gandek B. *SF-36 Health Survey: Manual and Interpretation Guide.* Boston, Mass: The Health Institute, New England Medical Center; 1993.

23 Dobrzykowski EA, Nance T. The Focus On Therapeutic Outcomes (FOTO) Outpatient Orthopedic Rehabilitation Database: results of 1994–1996. *Journal of Rehabilitation Outcomes Measurements.* 1997;1(1):56–60.

24 Ware JE Jr, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36), I: conceptual framework and item selection. *Med Care.* 1992;30:473–483.

25 Hart DL, Wright BD. Development of an index of physical functional health status in rehabilitation. *Arch Phys Med Rehabil.* 2002;83: 655–665.

26 Ware JE Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care.* 1996;34:220–233.

27 Hart DL. Test-retest reliability of an abbreviated self-report overall health status measure. *J Orthop Sports Phys Ther.* In press.

28 Taylor SJ, Taylor AE, Foy MA, Fogg AJ. Responsiveness of common outcome measures for patients with low back pain. *Spine.* 1999;24: 1805–1812.

29 McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36), II: psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care.* 1993;31:247–263.

30 McHorney CA, Ware JE Jr, Lu JF, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36), III: tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care.* 1994;32:40–66.

31 Ware JE Jr. *How to Score the SF-12 Physical and Mental Health Summary Scales.* Boston, Mass: The Health Institute, New England Medical Center; 1995.

32 Hart DL, Dobrzykowski EA. Impact of exercise history on health status outcomes in patients with musculoskeletal impairments. *Orthopaedic Physical Therapy Clinics of North America.* 2000;9(1):1–16.

33 Hart DL, Dobrzykowski EA. Influence of orthopaedic clinical specialist certification on clinical outcomes. *J Orthop Sports Phys Ther.* 2000;30:183–193.

34 Hart DL. Assessment of unidimensionality of physical functioning in patients receiving therapy in acute, orthopedic outpatient centers. *J Outcome Meas.* 2000;4:413–430.

35 Riddle DL, Lee KT, Stratford PW. Use of SF-36 and SF-12 health status measures: a quantitative comparison for groups versus individual patients. *Med Care.* 2001;39:867–878.

36 Jette DU, Jette AM. Physical therapy and health outcomes in patients with spinal impairments. *Phys Ther.* 1996;76:930–941; discussion 942–945.

37 Di Fabio RP, Boissonnault W. Physical therapy and health-related outcomes for patients with common orthopaedic diagnoses. *J Orthop Sports Phys Ther.* 1998;27:219–230.

38 Patrick DL, Deyo RA, Atlas SJ, et al. Assessing health-related quality of life in patients with sciatica. *Spine.* 1995;20:1899–1908.

39 Jette AM. Outcomes research: shifting the dominant research paradigm in physical therapy. *Phys Ther.* 1995;75:965–970.

40 Delitto A. Are measures of function and disability important in low back care? *Phys Ther.* 1994;74:452–462.

41 Enebo BA. Outcome measures for low back pain: pain inventories and functional disability questionnaires. *Chiropractic Technique.* 1998; 10(1):27–33.

42 Di Fabio RP, Mackey G, Holte JB. Physical therapy outcomes for patients receiving workers' compensation following treatment for herniated lumbar disc and mechanical low back pain syndrome. *J Orthop Sports Phys Ther.* 1996;23:180–187.

43 Bronfort G, Bouter LM. Responsiveness of general health status in chronic low back pain: a comparison of the COOP charts and the SF-36. *Pain.* 1999;83:201–209.

44 Iezzoni LI, ed. *Risk Adjustment for Measuring Health Care Outcomes.* Ann Arbor, Mich: Health Administration Press; 1994.

45 Iezzoni LI. The risks of risk adjustment. *JAMA.* 1997;278: 1600–1607.

46 Iezzoni LI. An introduction to risk adjustment. *Am J Med Qual.* 1996;11:S8–S11.

47 Nelder J, Wedderburn W. Generalized linear models. *Journal of the Royal Statistical Society.* 1972;A(135):370–384.

48 Stratford PW, Binkley J, Solomon P, et al. Assessing change over time in patients with low back pain. *Phys Ther.* 1994;74:528–533.

49 Cohen J. *Power Analysis for the Behavior Sciences.* 2nd ed. New York, NY: Academic Press Inc; 1988.

50 Noll E, Key A, Jensen GM. Clinical reasoning of an experienced physiotherapist: insight into clinician decision-making regarding low back pain. *Physiother Res Int.* 2001;6:40–51.

51 Constance DP. *Effect of Experience on Physical Therapist Functional Outcomes* [master's thesis]. St Augustine, Fla: University of St Augustine; 2000.

52 Fritz JM, George SZ. Identifying psychosocial variables in patients with acute work-related low back pain: the importance of fear-avoidance beliefs. *Phys Ther.* 2002;82:973–983.

53 Pryor DB, Lee KL. Methods for the analysis and assessment of clinical databases: the clinician's perspective. *Stat Med.* 1991;10: 617–628.

54 Werneke M, Hart DL, Cook D. A descriptive study of the centralization phenomenon: a prospective analysis. *Spine.* 1999;24:676–683.

55 Delitto A, Erhard RE, Bowling RW. A treatment-based classification approach to low back syndrome: identifying and staging patients for conservative treatment. *Phys Ther.* 1995;75:470–485; discussion 485–489.

56 Delitto A, Cibulka MT, Erhard RE, et al. Evidence for use of an extension-mobilization category in acute low back syndrome: a prescriptive validation pilot study. *Phys Ther.* 1993;73:216–222; discussion 223–228.

57 Werneke M, Hart DL. Centralization phenomenon as a prognostic factor for chronic low back pain and disability. *Spine.* 2001;26:758–764.

58 Long AL. The centralization phenomenon: its usefulness as a predictor or outcome in conservative treatment of chronic law back pain (a pilot study). *Spine.* 1995;20:2513–2520.

59 Werneke M, Hart DL. Discriminant validity and relative precision for classifying patients with nonspecific neck and back pain by anatomic pain patterns. *Spine.* 2003;28:161–166.

60 Gatchel RJ, Mayer T, Dersh J, et al. The association of the SF-36 health status survey with 1-year socioeconomic outcomes in a chronically disabled spinal disorder population. *Spine.* 1999;24:2162–2170.

61 Jette DU, Jette AM. Professional uncertainty and treatment choices by physical therapists. *Arch Phys Med Rehabil.* 1997;78:1346–1351.

62 Testa MA. Interpretation of quality-of-life outcomes: issues that affect magnitude and meaning. *Med Care.* 2000;38(suppl 9):II166–II174.

63 Guyatt GH. Making sense of quality-of-life data. *Med Care.* 2000; 38(suppl 9):II175–II179.