The Journal of Arthroplasty Vol. 11 No. 7 1996

Demographic Biases Found in Scoring Instruments of Total Hip Arthroplasty

Mark R. Brinker, MD,*†‡ Peter J. Lund, MD,*
Dennis D. Cox, PhD,§ and Robert L. Barrack, MD*

Abstract: Four hip scoring systems were used in evaluating 200 adult subjects who had no prior history of injury, pathologic condition, or treatment of the hips, knees, lower extremities, or spine. All subjects were in the age range typical of a total hip arthroplasty candidate (average age, 65 years; range, 50–100 years). In addition to a physical examination, complete demographic data were collected on each subject. Data were recorded on standardized flow sheets so that hip scores could be calculated; scores were normalized by dividing the observed scores by the maximum possible score. The average normalized total hip scores were Harris hip score, 90.8%; modified Harris hip score, 91.9%; Merle D'Aubigne score, 93.9%; and Hospital for Special Surgery hip rating, 87.5%. Demographic variables that had a significant negative correlation with hip scores included advanced age (particularly past age 85), an income below the poverty level, and the presence of two or more major medical conditions. Differences in hip scores between different study groups that have not been matched for various clinically relevant factors ("case mix") are at least as likely to represent differences in the patient populations as differences in surgical technique or implant design. Hip scores may decline over the course of a 10- to 20-year follow-up period due to the change in a patient's age and/or medical condition rather than any factor relating to the hip arthroplasty.
Key words: hip scoring system, total hip arthroplasty, demographics biases.

Currently, approximately 120,000 primary total hip arthroplasties (THAs) are performed in North America each year [1]. With the increasing concerns related to the rising cost of health care (and the need for cost containment and cost-effectiveness), there has been a growing interest in outcome studies [2–4], and the need for validated scoring instruments as a means of measuring outcomes has been recognized [5–10].

Over the past four decades, numerous scoring systems have been used to assess the preoperative and postoperative status of individuals undergoing THA by assigning numerical scores based on a variety of factors, including pain, function, activities, and range of motion. These numerically based instruments have provided a basis for comparison of differing disease processes, patient populations, and types of treatment. Because these scoring systems are relatively convenient to use, they have enjoyed widespread approval throughout the orthopaedic community and have been the gold standard for clinical investigation of hip arthroplasty over the last 30 years.

A number of recent studies, however, have critically analyzed various commonly used hip scoring

*From the *Department of Orthopaedic Surgery, Tulane University School of Medicine, New Orleans, Louisiana, †Department of Orthopaedic Surgery, University of Texas, Health Science Center, Houston, ‡Fondren Orthopedic Group LLP, Texas Orthopedic Hospital, Houston, and §Department of Statistics, Rice University, Houston, Texas.*

Winner of the American Orthopaedic Association Residents Travel Award for Excellence in Orthopaedic Research, 1994.

Presented at the 61st Annual American Academy of Orthopaedic Surgeons Meeting.

Reprint requests: Mark R. Brinker, MD, Texas Orthopedic Hospital, Fondren Orthopedic Group LLP, 7401 S. Main, Houston, TX 77030.

instruments. Many of these studies have concluded that these instruments do not use standardized terminology and lack validation [7,8,11–18]. Gross defined the terms *internal validity* and *external validity* [8]. A scoring system is internally valid if there is small observer variability for a given population, such as the scoring system described by Johanson et al. [9]. A scoring system is externally valid only if it controls for differences between study populations.

Differences between study populations might include risk factors (confounding variables) which have an unrecognized effect on the scores of subjects enrolled in a study. Greenfield defined the term *case mix* as "the features that increase the risk of a bad outcome or influence the choice of treatment" [3]. The author further noted that "the purpose of case-mix adjustment is to separate the effects of the care given from those of the preexisting health status and other factors (such as age and socioeconomic status) that affect outcome measures" [3]. Although this phenomenon has been recognized [8,17], the effect of clinically relevant factors on numerically based hip scoring systems has not been previously delineated.

The purpose of this investigation was to define the effect, if any, of clinically relevant factors (such as age, sex, race, relative body weight, socioeconomic status, and number of major medical conditions) on numerically based hip scores. We chose to do this by examining a group of subjects who had not undergone hip arthroplasty, to eliminate the effects of care given (hip surgery) on the outcome measures to be studied. To the best of our knowledge, such a study has not previously been reported in the literature.

Materials and Methods

A total of 373 volunteers, age 50 or older, were interviewed and screened by one of us (M.R.B.) prior to inclusion into the study group. A total of 173 volunteers were excluded from the study because they (1) were unable to walk, (2) had sought treatment of hip or knee pain at any time in the past, (3) had a known disease of the spine, (4) had a musculoskeletal abnormality of the lower extremities, (5) had a history of surgery of the spine, hips, or knees, (6) had a prior injury or pathologic condition of the lower extremities (other than peripheral vascular disease). Volunteers were recruited through community groups, churches and synagogues, and bulk mailers distributed throughout Louisiana. All volunteers were recruited (by offering free examinations) without knowledge of the intent or purpose of the study or the type of history or physical examination to be performed. Every effort was made not to introduce bias into our recruitment to ensure that our study group of 200 subjects was representative of the population at large. With respect to age, sex distribution, and prevalence of major medical conditions, a thorough review of the medical literature produced only one published series [19] reporting all of these data in patients undergoing THA. Our study group was comparable to this series in terms of age and sex distribution and prevalence of hypertension; the prevalence of coronary artery disease and chronic obstructive pulmonary disease was less in our group as compared with a group of total hip patients reported by Koide et al. [19]. We would use this fact as evidence that our recruitment strategy was not more likely to have less healthy people respond. Two hundred of 373 volunteers met the eligibility criteria and underwent a detailed medical history and examination of their lower extremities.

A detailed medical history was performed for all 200 subjects by one of us (M.R.B.) and a physical examination was performed by a team of nine orthopaedic surgery residents; a prestudy workshop was attended by each of these residents where all terms and techniques and methods of examination were agreed on. Subjects were questioned regarding medical conditions, cigarette and alcohol use, prescription medications, and socioeconomic factors including family size and annual income. Subjects were also questioned about functional activities, walking ability, and hip pain.

The average age of all subjects was 65.5 years (range, 50–100 years); there were 57 men and 143 women. The distribution of race was 154 white and 43 black; this distribution approximates that of Louisiana. Two of our subjects were of Hispanic descent and one was of Asian descent. Because of the small numbers of Hispanic and Asian subjects in our study, no specific conclusion can be made on these ethnic groups. Men weighed an average of 86 kg (range, 65–124 kg) and women averaged 70 kg (range, 43–109 kg). The average heights of men and women were 177 cm (range, 152–193 cm) and 163 cm (range, 140–180 cm), respectively. Each subject's relative body weight was classified, based on height by the method of Stern and Insall [20], as underweight, normal weight, mildly obese, moderately obese, or severely obese as determined by the 1983 Metropolitan Life Insurance Company table. In this study, none of our subjects (0%) were underweight, 120 (60.0%) were of normal weight, 56 (28.0%) were mildly obese, 16 (8.0%) were moderately obese, and 3 (1.5%) were severely obese. Five of our subjects (4 men and 1 woman) could not be classified because of short stature to the extent that their height was not found on the 1983 Metropolitan Life Insurance Company table. An overview of the medical conditions reported by our subjects is shown in Table 1. A major medical condition was defined as any condition that could potentially limit daily activities or ambulation; these conditions were defined by the consensus of a committee of 10 physicians. Overall, 61 subjects (30.5%) had one major medical condition, 40 (20.0%) had two, 10 (5.0%) had three, 9 had four, and 3 had five major medical conditions. Eleven subjects had a history of steroid use related to a major medical condition. Thirty-five of 143 women (24.5%) had had a hysterectomy.

Overall, 147 subjects (73.5%) took prescription medications for medical conditions; 78 subjects (39.0%) were taking cardiovascular medications. Seventy-one subjects (35.5%) had a past history of smoking cigarettes. At the time of our examination, only 24 subjects (12.0%) continued to smoke. The overall pack-year history for all smokers averaged 35 pack-years (range, 1–130 pack-years). Regarding alcohol consumption, 52 subjects (26.0%) drank only socially, 14 (7.0%) had fewer than three drinks per day, and 1 (0.5%) had three or more drinks per day.

The average yearly family income of all responding subjects (193 of 200) was $23,800 (range, $0–$111,000). For the purpose of assessing whether a subject's family income was below the poverty level [21], complete data regarding family size and income were available for 181 subjects (90.5%). Of these 181 subjects, 46 (25.4%) had family incomes below the poverty level [21]. A total of 143 subjects (71.5%) had some form of medical insurance.

Table 1. Overview of Medical Conditions (n = 200 Subjects)

Medical Conditions	No. (%) of Subjects
Major	
Hypertension	70 (35.0)
Malignancy	27 (13.5)
Angina pectoris	24 (12.0)
Coronary artery disease	22 (11.0)
Diabetes (requiring daily medication)	20 (10.0)
Renal disease	13 (6.5)
Peripheral vascular disease	11 (5.5)
Asthma (requiring daily medication)	10 (5.0)
Chronic obstructive pulmonary disease	8 (4.0)
Cerebrovascular accident (in past)	7 (3.5)
Myocardial infarction (in past)	6 (3.0)
Liver disease	4 (2.0)
Sickle cell disease	0
Pulmonary embolism	0
Other	
Miscellaneous	107 (53.5)
Hypothyroidism	19 (9.5)
Deep venous thrombosis (in past)	7 (3.5)
Tuberculosis (in past, inactive)	3 (1.5)

Physical examinations were performed at three examination specific stations. Each of the nine residents (3 residents at each of the 3 examination-specific stations) performed the same unique specific tasks for each of the 200 subjects. For example, arc of passive motion measurements were performed on bilateral lower extremities, by a team of three residents, using a goniometer and were recorded to the closest 5°. One resident manipulated the lower extremity, one made measurements using a goiniometer, and one recorded the data. Hip flexor and abductor muscle strength were graded as normal, good, fair, poor, trace, or none [22].

Data were recorded on standardized flow sheets so that the following four hip scores could be calculated: Harris hip score [23], modified Harris hip score [24], Merle D'Aubigne score [25], and Hospital for Special Surgery hip rating [26]. These represent the most commonly used hip rating scales in the orthopaedic literature. A comparison of the maximum total and component scores possible for the four hip scoring systems studied is shown in Table 2. A hip score was calculated for the right and left hips of each subject. For each subject, we used the average of right and left hip scores as a single score for statistical analysis.

Statistical analysis was performed with Student's *t*-test, chi-square test, and analysis of variance using the SAS statistical package (SAS, Cary, NC). Results were judged significant if the *P* value was less than or equal to .05, and we have included *P* values or upper bounds thereon wherever available. When making multiple comparisons, we looked at the results from Tukey's studentized range (a variant of Newman–Keuls), Bonferroni, and Scheffe. In fact, all three methods gave identical conclusions in all cases at the .05 level of significance. For the purpose of statistical analysis, hip scores and component scores (pain, function/activities, range of motion) were normalized by dividing the observed score by the maximum possible score. For example, a subject with a total score of 12 and a pain score of 3, by the method of Merle D'Aubigne, would have a normalized total score of 67% (12 ÷ 18) and a normalized pain score of 50% (3 ÷ 6). The data were analyzed to determine the effect of clinically relevant factors (ie, age, sex, race, relative body weight, socioeconomic status, and number of major medical conditions) on nor-

Table 2. Comparison of the Maximum Total and Component Scores Possible for Four Hip Scoring Systems [23–26]

Hip Scoring System	Total Score	Pain Score	Function Activities Score*	Deformity Score	Range of Motion Score†
Harris hip score	100	44	47	4	5
Modified Harris hip score	100	44	46	5	5
Merle D'Aubigne score	18	6	6	0	6
Hospital for Special Surgery hip rating	40	10	20	0	10

*Function/activities for the Merle D'Aubigne score was taken to be the equivalent of "ability to walk" [25]. Function/activities for The Hospital for Special Surgery hip rating was taken to be the equivalent of the sum of "walking" plus "function" [26]. †Range of motion for the Merle D'Aubigne score was taken to be the equivalent of "mobility" [25]. Range of motion for The Hospital for Special Surgery hip rating was taken to be the equivalent of "muscle power and motion" [26].

malized total scores. A similar analysis was performed to determine the effect of clinically relevant factors on normalized component scores (pain, function/activities). We selected these clinically relevant factors because they were believed to be the most likely to have an effect on hip scores. We acknowledge that a variety of other variables could have been studied, and in this regard, our study should be considered preliminary.

Results

The average total and component scores and normalized total and component scores for all 200 subjects are shown in Table 3. Normalized total scores were significantly higher by the method of Merle D'Aubigne as compared with scores of The Hospital for Special Surgery hip rating ($P > .05$). Normalized pain scores were highest for the modified Harris hip score, whereas normalized function/activities and range of motion scores were highest for the Merle D'Aubigne score. The distribution of normalized total scores for all four hip scoring systems is shown in Table 4. Normalized total scores of 90 to 100 were seen in 81.5% of subjects by the method of Merle D'Aubigne, 73.0% using the modified Harris hip score, but in only 67.5% of subjects using the Harris hip score and Hospital for Special Surgery hip rating. Data regarding pain and function/activities are summarized in Table 5.

The averages and ranges for passive arcs of hip motion were flexion, 104° (range, 75°–130°); abduction, 42° (range, 15°–65°); adduction, 29° (range, 10°–50°); external rotation, 37° (range, 15°–60°); and internal rotation, 26° (range, 0°–60°).

A total of 46 subjects (23.0%) reported that their ambulation was limited by one or more factors. Ambulation was limited in 19 subjects (9.5%) because of a medical condition that manifests as shortness of breath, chest pain, or generalized weakness. Ambulation was limited in 23 subjects (11.5%) because of a musculoskeletal complaint such as hip or knee pain or stiffness, muscle weak-

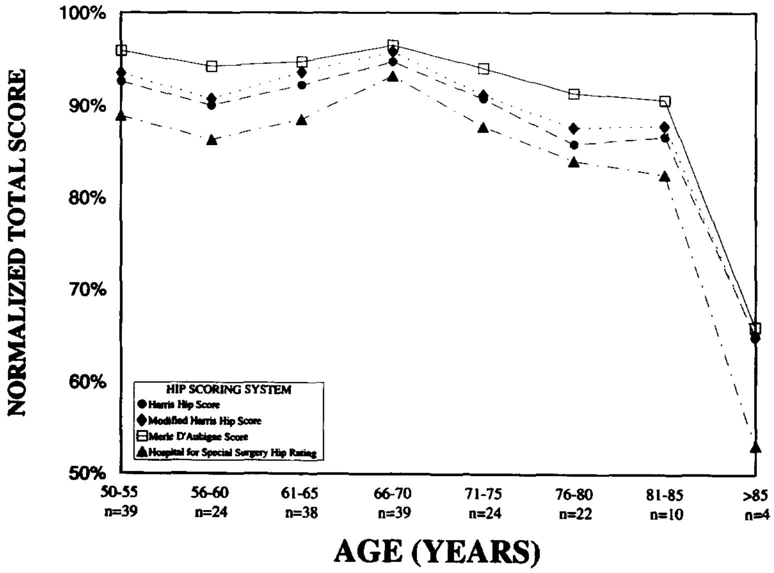

Fig. 1. Significant differences in normalized total scores were seen between age groups for each of the four hip scoring systems ($P < .0002$). Detailed analysis revealed that the over 85 age group had significantly lower scores than all other age groups; however, no significant differences were observed among the under 85 age groups ($P > .05$). This finding was seen for all four hip scoring systems.

Table 3. Averages and Ranges of Total Scores, Components Scores, and Normalized Total and Component Scores of Four Hip Scoring Systems in a Normal Population of Older Adults (n = 200 Subjects)

Hip Scoring System	Normalized Total Score	Total Score	Normalized Pain Score	Function/ Pain Score	Function/ Activities Score	Normalized Activities Score	Normalized Deformity Score	Range of Deformity Score	Range of Motion Score	Normalized Motion Score
Harris hip score	90.8 (32–100)	90.8%	39.0 (10–44)	88.6%	43.0 (9–47)	91.5%	4.0*	100%	4.8 (4.2–5)	96.0%
Modified Harris hip score	91.9 (30–100)	91.9%	39.9 (15–44)	90.7%	42.2 (8–46)	91.6%	5.0 (4–5)	100%	4.8 (3–5)	96.0%
Merle D'Aubigne score	16.9 (9–18)	93.9%	5.4 (2–6)	90.0%	5.7 (0–6)	95.0%	—†	—†	5.8 (4–6)	96.7%
Hospital for Special Surgery hip rating	35.0 (14–40)	87.5%	8.7 (2–10)	87.0%	18.2 (2–20)	91.0%	—†	—†	8.1 (4–10)	81.0%

Ranges shown in parentheses. *All 200 subjects received the maximum deformity component score possible. †A component score for deformity is not found in the Merle D'Aubigne score or The Hospital for Special Surgery hip rating.

Table 4. Distribution of Normalized Total Scores Four Hip Scoring Systems (n = 200 Subjects)

Normalized Total Hip Score	Harris Hip Score (%)	Modified Harris Hip Score (%)	Merle D'Aubigne Score (%)	Hospital for Special Surgery Hip Rating (%)
90–100	135 (67.5)	146 (73.0)	163 (81.5)	135 (67.5)
80–89	33 (16.5)	26 (13.0)	22 (11.0)	24 (12.0)
70–79	15 (7.5)	15 (7.5)	7 (3.5)	24 (12.0)
< 70	17 (8.5)	13 (6.5)	8 (4.0)	17 (8.5)

ness, or fatigue; ambulation was limited in four additional subjects (2.0%) because of both a medical condition and a musculoskeletal complaint.

Table 5. Data Regarding Pain and Function/Activities in a Normal Population of Older Adults

Category	Proportion	(No.)
Pain*		
None	65.0%	(260)
Slight	12.8%	(51)
Mild	11.0%	(44)
Moderate	8.2%	(33)
Marked	1.0%	(4)
Totally disabling	2.0%	(8)
Function/Activities†		
Limp		
None	82.5%	(165)
Slight	14.0%	(28)
Moderate	2.0%	(4)
Severe	1.5%	(3)
Support		
None	92.5%	(185)
Cane for long walks	1.5%	(3)
Cane most of the time	3.0%	(6)
One crutch	0.5%	(1)
Two canes	4.5%	(5)
Two crutches	0.0%	(0)
Unable to walk	0.0%	(0)
Ambulation		
Unlimited	77.0%	(154)
Six blocks	6.0%	(12)
Three blocks	11.5%	(23)
Indoors	0.5%	(1)
Bed and chair	5.0%	(10)
Stairs		
Normally	65.5%	(131)
With railing	30.5%	(61)
Any manner	2.5%	(5)
Unable	1.5%	(3)
Putting on shoes and socks		
With ease	91.5%	(183)
With difficulty	8.0%	(16)
Unable	0.5%	(1)
Sitting		
Any chair, 1 hour	87.0%	(174)
High chair	10.0%	(20)
Unable to sit comfortably	3.0%	(6)
Public transportation		
Able to use	95.5%	(191)
Unable to use	4.5%	(9)

*n = 400 hips. †n = 200 subjects.

An analysis of normalized total scores by subject age is shown in Figure 1. Significant differences in normalized total scores were seen between age groups for each of the four hip scoring systems, with older subjects having lower scores ($P < .0002$). Further analysis revealed that the over-85 age group had significantly lower scores than all other age groups; however, no significant differences were observed among the under-85 age groups ($P > .05$). These findings were seen for all four hip scoring systems. No significant differences in normalized component pain scores were seen between age groups for each of the four hip scoring systems ($P > .10$); however, significant differences in normalized component function/activities scores were seen between age groups, with the over-85 age group having lower scores for each of the four hip scoring systems ($P < .05$) (Fig. 2).

An analysis of normalized total scores by subject sex is shown in Figure 3. Overall, no significant differences in normalized total scores were seen between men and women for each of the four hip scoring systems ($P > .10$). A similar situation existed regarding normalized component scores and sex ($P > .10$). Similarly, no significant differences in normalized total or component scores were seen between the five classes of relative body weight [20] ($P > .37$).

When socioeconomic factors were analyzed, an interaction was present between race and income. An analysis of normalized total scores by race and income level revealed no significant differences in hip scores based on race alone for all four hip scoring systems ($P > .49$), but significantly lower scores in subjects below the poverty level for all four scoring systems ($P < .05$) (Fig. 4). An analysis of normalized component scores based on income level is shown in Figures 5 and 6. Subjects with family incomes below the poverty level tended to have significantly lower normalized component pain scores for all four scoring systems ($P < .03$) (Fig. 5) and significantly lower normalized component function/activities scores for the Harris hip score and the modified Harris hip score ($P < .02$) (Fig. 6).

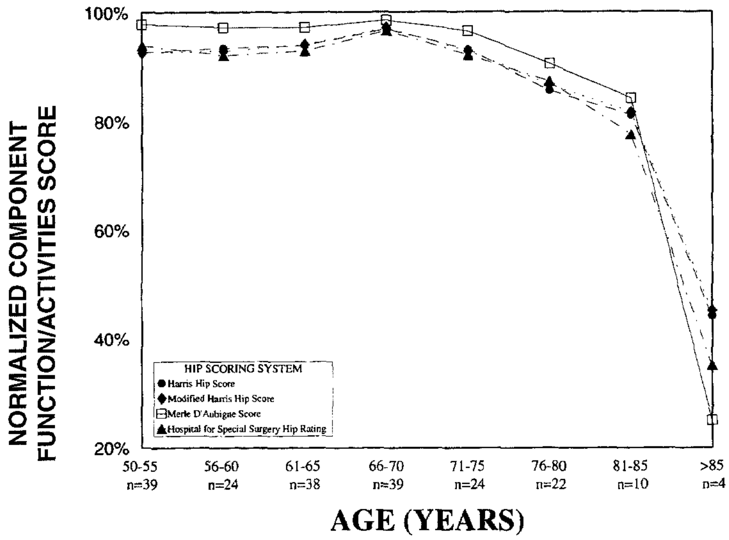

Fig. 2. Significant differences in normalized component function/activities scores were seen between age groups ($P < .0001$). Detailed analysis revealed that the over 85 age group had significantly lower scores than all other age groups; however, no significant differences were observed among the under 85 age groups ($P > .05$). This finding was seen for all four hip scoring systems.

An analysis of normalized total scores by medical conditions is shown in Figure 7. Significant differences in normalized total scores were observed between subjects based on the number of major medical conditions they had. Subjects with two or more major medical conditions had significantly lower normalized total scores for all four hip scoring systems ($P < .0002$). This relationship persisted when we accounted for income level ($P < .0003$) and age ($P < .003$), suggesting that the number of major medical conditions is an important predictor of hip scores. An analysis of normalized component scores based on medical conditions is shown in Figures 8 and 9. Subjects with two or more major medical conditions tended to have significantly lower normalized component pain scores for all four scoring systems ($P < .02$) (Fig. 8) and, likewise, significantly lower normalized component function/activities scores ($P < .03$) (Fig. 9).

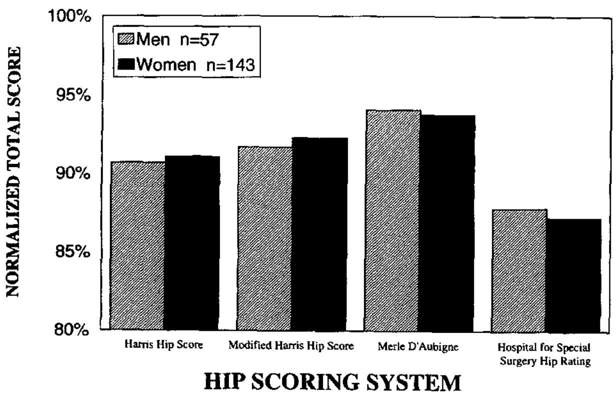

Fig. 3. No significant differences in normalized total scores were seen between men and women for each of the four hip scoring systems ($P = .57$).

Discussion

Results of this investigation suggest that several clinically relevant factors do indeed have a significant effect on the total and component scores of numerically based hip scoring systems. Factors that have a significant effect on hip scores include subject age, family income, and medical conditions. Factors that do not appear to have a significant effect on hip scores include subject sex, race, and relative body weight.

In this investigation, subject age had a significant effect on all four total hip scores and also component function/activities scores. Insler et al. [27] and Ilstrup et al. [28] have reported their similar findings of lower hip scores in older patients, and Constant has reported poorer shoulder function in

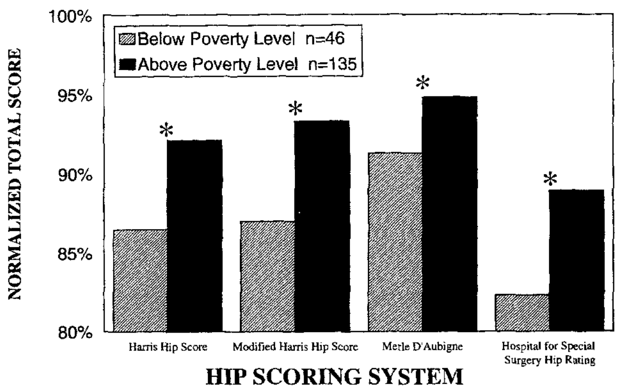

Fig. 4. Significant differences in normalized total scores were seen between impoverished and nonimpoverished subjects. Subjects with family incomes below the poverty level had significantly lower normalized total scores for the Harris hip score (*$P = .012$), modified Harris hip score (*$P = .004$), Merle D'Aubigne score (*$P = .041$), and Hospital for Special Surgery hip rating (*$P = .009$).

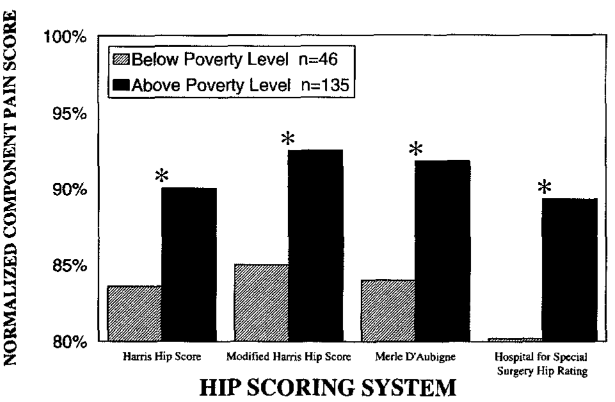

Fig. 5. Significant differences in normalized component pain scores were seen between impoverished and nonimpoverished subjects. Subjects with family incomes below the poverty level had significantly lower normalized component pain scores for the Harris hip score (*P = .025), modified Harris hip score (*P = .003), Merle D'Aubigne Score (*P = .001), and Hospital for Special Surgery hip rating (*P = .003).

his older patients recovering from injuries of the shoulder [6]. Although it is clear that none of the subjects in our study have been followed longitudinally to observe individual changes in total and function/activities scores over time, the significant trend of lower scores observed in our older subjects suggests a diminution in function and total hip scores with aging. In this investigation, normalized total Harris hip scores averaged 94.8% for subjects 66 to 70 years of age as compared with only 86.6% for subjects 81 to 85 years; normalized component

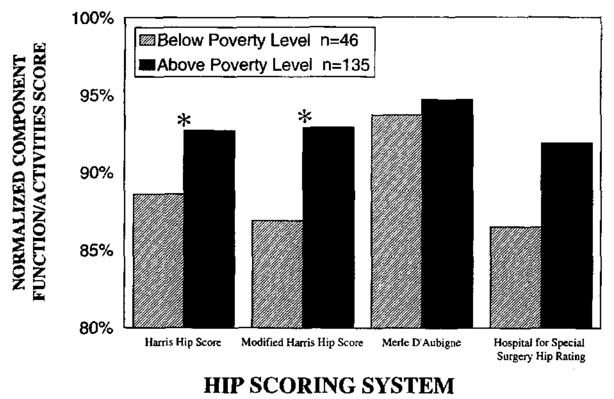

Fig. 6. Significant differences in normalized component function/activities scores were seen between impoverished and nonimpoverished subjects for the Harris hip score (*P = .014) and the modified Harris hip score (*P = .01). No significant differences in normalized function/activities scores were seen between impoverished and nonimpoverished subjects for the Merle D'Aubigne score (P = .66) or Hospital for Special Surgery hip rating (P = .068).

function/activities scores averaged 96.9% and 81.8% for these age groups, respectively. Conclusions drawn from 10- or 20-year follow-up outcome studies of patients with THAs must be tempered by the fact that an observed decline in hip scores may represent the natural morbidity of aging in a patient with an otherwise well-functioning arthroplasty.

No significant relationship was observed between sex and total or component hip scores or relative body weight and hip scores. Likewise, no significant relationship existed between race and hip scores; however, we did observe a significant relationship between race and socioeconomic status to the effect that a greater relative proportion of black subjects were impoverished as compared with whites. This observation has been reported by other authors [29,30]. When assessing the effect of clinical factors on hip scores, care must be taken so as not to draw false conclusions based on confounding variables such as race and socioeconomic status.

A significant relationship was observed in our group regarding socioeconomic status and hip scores. Impoverished subjects had significantly lower total and component pain scores for all four scoring systems and lower component function/activities scores for the Harris hip score and modified Harris hip score. A number of authors have addressed the complex relationship between socioeconomic status and wellness [29,31–37]. Adler et al. have noted that "socioeconomic status is a strong consistent predictor of morbidity and premature mortality" [29]. Syme and Berkman [38] have further noted that individuals in lower socioeconomic groups tend to have higher disability, morbidity, and mortality rates [36]. The three most recognizable components of socioeconomic status include income level, education, and occupational status [29]. In this investigation, we chose to use income level as our determinant of socioeconomic status because this information is quantifiable and was readily available for most of our subjects. Adler et al. have suggested that the poorer overall health seen in lower socioeconomic groups is related to health risk behaviors (smoking, alcohol, etc.), differential exposure to physical and social situations, stress, and control over work circumstances [29]. It is likely that many of these mechanisms had a role in affecting hip scores in our study group; in this investigation, subjects of lower socioeconomic status (impoverished subjects) tended to report more pain and less function. Although the complex relationship between socioeconomic factors and hip scores remains somewhat obscure, it is clear that income level has an effect on total scores and component pain and function/activities scores that must be

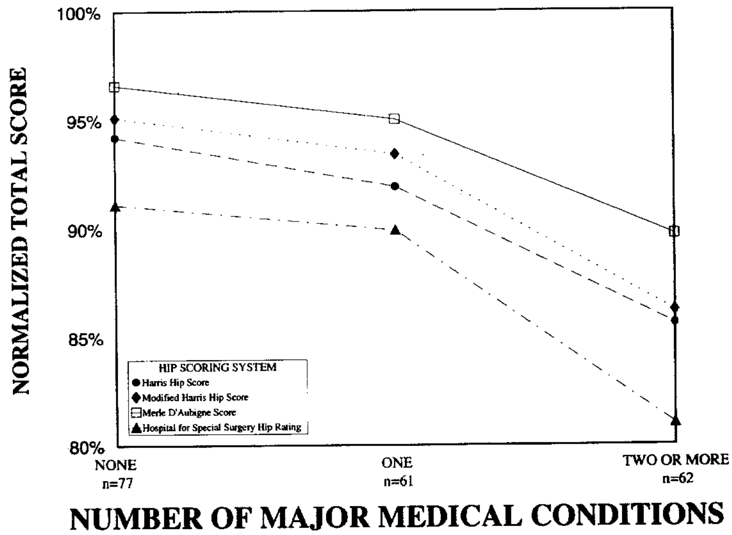

Fig. 7. Significant differences in normalized total scores were seen between subjects based on the number of major medical conditions they had. Subjects with two or more major medical conditions had significantly lower normalized total scores for all four hip scoring systems ($P < .0002$).

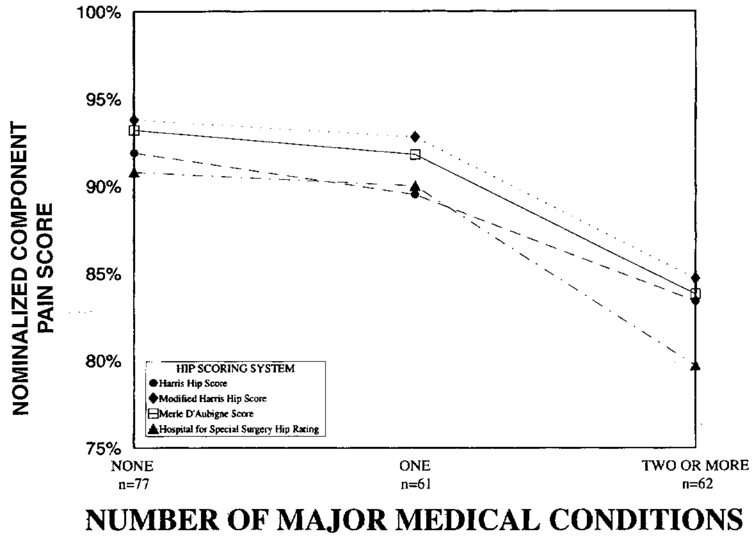

Fig. 8. Significant differences in normalized component pain scores were seen between subjects based on the number of major medical conditions they had. Subjects with two or more major medical conditions had significantly lower normalized component pain scores for the Harris hip scores ($P = .013$) and modified Harris hip score, Merle D'Aubigne score, and Hospital for Special Surgery hip rating ($P < .001$).

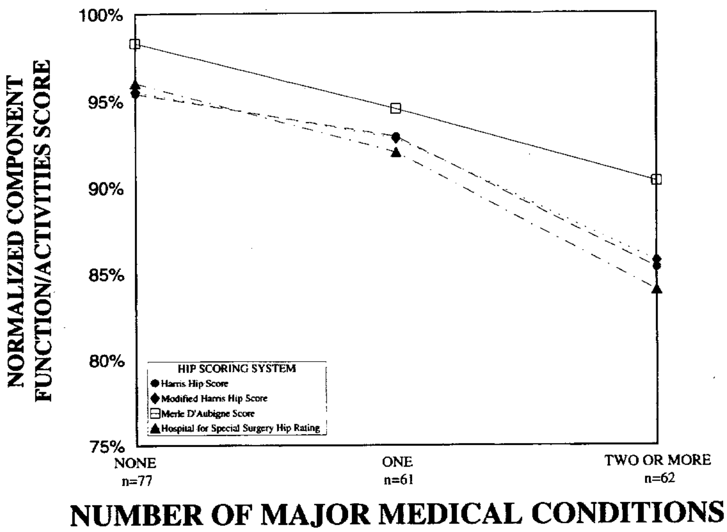

Fig. 9. Significant differences in normalized component function/activities scores were seen between subjects based on the number of major medical conditions they had. Subjects with two or more major medical conditions had significantly lower normalized component function/activities scores for the Harris hip score ($P < .0001$), modified Harris hip score ($P < .0002$), Merle D'Aubigne score ($P = .011$), and Hospital for Special Surgery hip rating ($P = .026$).

accounted for when reporting outcomes of patients with THAs.

We observed a significant relationship between hip scores and the number of major medical conditions subjects had. Subjects with two or more major medical conditions tended to have lower total scores as well as lower component pain and function/activities scores. Charnley recognized the importance of factoring medical conditions into hip evaluations when he described category C patients as those having "conditions directly impairing the act of walking," and noted that patients of different categories should not be compared [38]. Liang et al. have noted the importance of medical conditions as a comorbidity and have stated "concurrent active medical or operative problems may be associated with pain or with loss of function, potentially confounding the outcome of total hip arthroplasty" [17]. Results of this investigation support this thought insofar as our subjects with two or more major medical conditions reported more pain and poorer function and had lower hip scores.

Conclusion

The practice of blindly comparing series of THAs to draw conclusions about disease processes, surgical techniques, or implant designs based on hip scores should be reconsidered. Results of this investigation suggest that problems related to differences in patient populations represent confounding variables and that the case-mix must be accounted for (using a case-mix adjustment) [3] if comparisons between study groups are to be meaningful.

Acknowledgments

The authors thank Dr. Douglas Waldman, Dr. James Dunlap, Dr. Rolando Garcia, Dr. Randall Jennings, Dr. Andrew Palafox, Dr. Paul Pflueger, Dr. Bill Tejeiro, and Dr. Michael Wolfe for their assistance in completing this study. The authors also thank Becky Trahan and Michele LeBouef for secretarial assistance with the manuscript.

References

1. Harris WH, Sledge CB: Total hip and total knee replacement. N Engl J Med 323:725, 1990
2. Amadio PC: Outcomes measurements: more questions; some answers. (editorial) J Bone Joint Surg 75A:1583, 1993
3. Greenfield S: The state of outcome research: are we on target? N Engl J Med 320:1142, 1989
4. Keller RB, Rudicel SA, Liang MH: Outcomes research in orthopaedics. J Bone Joint Surg 75A:1562, 1993
5. Constant CR, Murley AHG: A clinical method of functional assessment of the shoulder. Clin Orthop 214:160, 1987
6. Constant CR: Age related recovery of shoulder function after injury. MA thesis, University College, Cork, Ireland, August 1986
7. Gartland J: Orthopaedic clinical research: deficiencies in experimental design and determinations of outcome. J Bone Joint Surg 70A:1357, 1988
8. Gross M: A critique of the methodologies used in clinical studies of hip-joint arthroplasty published in the English-language orthopaedic literature. J Bone Joint Surg 70A:1364, 1988
9. Johanson NA, Charlson ME, Szatrowski TP, Ranawat CS: A self-administered hip-rating questionnaire for the assessment of outcome after total hip replacement. J Bone Joint Surg 74A:587, 1992
10. Liang MH, Larson MG, Cullen KE, Schwartz JA: Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. Arthritis Rheum 28:542, 1985
11. Andersson G: Hip assessment: a comparison of nine different methods. J Bone Joint Surg 54B:621, 1982
12. Callaghan JJ, Dysart SH, Savory CF, Hopkinson WJ: Assessing the results of hip replacement: a comparison of five different rating systems. J Bone Joint Surg 72B:1008, 1990
13. Dorey F, Amstutz HC: Discrepancies in the orthopaedic literature: why? A statistical explanation. Orthopaedic research. vol. 42, p. 555. In Heckman JD (ed): Instructional course lecture. American Academy of Orthopaedic Surgeons, Rockville, IL, 1993
14. Galante J: Evaluation of results of total hip replacement. (editorial) J Bone Joint Surg 72A:159, 1990
15. Galante J: The need for a standardized system for evaluating results of total hip surgery. (editorial) J Bone Joint Surg 67A:511, 1985
16. Johnston RC, Fitzgerald RH Jr, Harris WH et al: Clinical and radiographic evaluation of total hip replacement. J Bone Joint Surg 72A:161, 1990
17. Liang MH, Katz JN, Phillips C et al: The total hip arthroplasty outcome evaluation form of the American Academy of Orthopaedic Surgeons. J Bone Joint Surg 73A:639, 1991
18. Ritter MA, Fechtman RW, Keating EM, Faris PM: The use of a hip score for evaluation of the results of total hip arthroplasty. J Arthroplasty 5:187, 1990
19. Koide M, Pilon RN, Vandam LD, Lowell JD: Anesthetic experience with total hip replacement. Clin Orthop 99:78, 1974
20. Stern SH, Insall JN: Total knee arthroplasty in obese patients. J Bone Joint Surg 72A:1400, 1990
21. Federal poverty income guidelines for fiscal year 1993. Federal Registry, p. 5455, Feb. 12, 1992

22. Hoppenfeld S: Physical exam of the spine and extremities. p. 26. Appleton-Century Crofts, Stamford, CT, 1984
23. Harris WH: Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. J Bone Joint Surg 51A:737, 1969
24. Gross AE, McDermott AGP, Lavoie MV et al: The use of allograft bone in revision hip arthroplasty. p. 47. In Brand A (ed): The hip: proceedings of the fourteenth open scientific meeting of The Hip Society. Mosby, St. Louis, 1986
25. D'Aubigne RM, Postel M: Functional results of hip arthroplasty with acrylic prosthesis. J Bone Joint Surg 36A:451, 1954
26. Wilson PD, Amstutz HC, Czerniecki A et al: Total hip replacement with fixation by acrylic cement. J Bone Joint Surg 54A:207, 1972
27. Insler H, Puri R: Functional evaluation of the hip in the elderly. Poster exhibit at American Academy of Orthopaedic Surgeons, San Francisco, CA, February 1993
28. Ilstrup DM, Nolan DR, Beckenbaugh RD, Coventry MB: Factors influencing the results in 2,012 total hip arthroplasties. Clin Orthop 95:250, 1973
29. Adler NE, Boyle WT, Chesney MA et al: Socioeconomic inequalities in health: no easy solution. JAMA 269:3140, 1993
30. Egbert LD, Rothman JL: Relation between the race and economic status of patients and who performs their surgery. N Engl J Med 297:90, 1977
31. Dutton DB, Levine S: Overview, methodological critique, and reformulation. p. 29. In Bunker JP, Gomby DS, Kehrer BH (eds): Pathways to health. Henry J. Kaiser Family Foundation, Menlo Park, CA, 1989
32. Kitagawa EM, Hauser PM: Differential mortality in the United States: a study of socioeconomic epidemiology. Harvard University Press, Cambridge, MA, 1973
33. McLeod JD, Kessler RC: Socioeconomic status differences in vulnerability to undesirable life events. J Health Soc Behav 31:162, 1990
34. Otten MW Jr, Teutsch SM, Williamson DF, Marks JS: The effect of known risk factors on the excess mortality of black adults in the United States. JAMA 263:845, 1990
35. Stokols D: Establishing and maintaining health environments. Am Psychol 47:6, 1992
36. Syme SL, Berkman LF: Social class, susceptibility and sickness. Am J Epidemiol 104:1, 1976
37. Yergan J, Flood AB, LoGerfo J, Diehr P: Relationship between patient race and the intensity of hospital services. Med Care 25:592, 1987
38. Charnley J: Numerical grading of clinical results (grading of hips). p. 20. In Low friction arthroplasty of the hip. Springer Verlag, New York, 1979