Baseline Function and Comorbidity Predict Outcome in Total Hip Arthroplasty

Norman A. Johanson, MD; Joshua M. Abzug, MD; Christopher A. Gallus, DO; Mary E. Charlson, MD

Abstract

The objective of this study is to show that patients with lower baseline hip scores will obtain the same or increased improvement from total hip arthroplasty when compared to patients with higher baseline scores. We assessed 300 patients undergoing total hip arthroplasty with a baseline Hip Rating Questionnaire, 6-minute walking distance test, instrumental activities of daily living, and medical comorbidities. Hip Rating Questionnaire data were collected at 1 and 5 years post-operatively. We stratified patients into groups of high and low pre-operative functional performance and health status, then compared post-operative mean Hip Rating Questionnaire scores and change from baseline. Patients with severe pre-operative impairments and higher levels of comorbidity had the same or better improvement, although their absolute hip scores were lower.

Introduction

Total hip arthroplasty (THA) is widely accepted as a leading surgical procedure in terms of cost-effectiveness and cost savings. THA produces substantial improvements in health-related quality of life.[1,2] During the evolution of the procedure, a variety of fixation modes, implant materials, bearing surfaces, and surgical techniques have been developed and used with great anticipation for improvement of the results.

The need to carefully measure the impact of orthopaedic procedures such as THA on overall health status, quality of life, and function has led to the development of validated measurements of function and health status.[1,3] It is possible to prospectively assess patient improvement by comparing the within-patient change from baseline rather than focusing only on the final score. Measurement of changes from baseline after a treatment is an important method of assessing its effectiveness in both large populations and in the clinical evaluation of an individual patient. Ethgen et al, after an extensive literature review of outcomes of THA using valid measures of health status, stated "patients who had poorer pre-operative health-related quality of life were more likely to experience greater improvement."[1] This observation underscores the need for prospective analysis of considerable functional and health-related covariates as they relate to joint-specific scores to avoid bias in assessing the true benefit of surgery.

To confirm the conclusion of Ethgen et al,[1] we prospectively assessed the relationships between the outcome of THA as measured by a validated Hip Rating Questionnaire (HRQ) and baseline patient characteristics including the 6-minute walking distance test, instrumental activities of daily living (IADL) assessment, and comorbidity.

Materials and Methods

We screened 1241 consecutive patients undergoing THA by participating physicians at a single institution between December 1988 and March 1992. From the total, 968 patients were eligible for participation. Various ineligibilities influenced the flow of participants through the study (Fig 1). The primary reason eligible patients were not entered was logistic, because a concurrent parent study required

Address for Correspondence:
Christopher A. Gallus, DO
Philadelphia College of Osteopathic Medicine
135 Atherton Drive
Exton, PA 19341
Phone: 303-478-2460
Fax: 215-762-6134
E-mail: christopherga@pcom.edu.

Christopher Gallus is an Intern at Philadelphia College of Osteopathic Medicine, Dept. of Orthopaedic Surgery, Philadelphia, PA
Norman Johanson is Chairman, Dept. of Orthopaedic Surgery, Drexel University College of Medicine, Philadelphia, PA
Joshua Abzug is a Chief Resident at Drexel University College of Medicine, Dept. of Orthopaedic Surgery, Philadelphia, PA.
Mary Charlson is a Professor of Medicine, Weil College of Medicine, Cornell University, NY, NY.

all patients to have preoperative dual-energy x-ray absorptiometry (DEXA) to assess bone quality of the femoral neck and its relationship to the outcome of THA.[4, 5, 6] Thus, patients identified as eligible at preadmission testing had to be willing either to wait that day for DEXA or to return to the hospital for DEXA before admission. Three hundred patients were finally enrolled, yielding 31% of eligible patients enrolled in the study. We obtained informed consent from all patients for participation in this study. Full Institutional Review Board approval was granted at the original time of the study and again for the purposes of preparing this report. All patients completed the HRQ at baseline and 1 and 5 years postoperatively. The HRQ is a self-administered questionnaire developed to assess the outcome of THA.[4] This scale has four domains: pain, walking ability, daily function, and overall impact of hip disease on quality of life. Each domain has a maximum score of 25, and the scores for the four domains are added for a total hip rating score in which 100 is the best and 16 indicates the worst hip function. Use of the HRQ is beneficial because it is specific for evaluation of THA. Additionally, the HRQ has demonstrated reproducibility and validity and provides excellent responsiveness in small and moderately sized patient samples.[4]

Of the 300 enrolled patients, 129 were men and 171 women with an overall mean age of 64.5 ± 10.2 years. The population was mostly white (97.7%) with only 1.3% black, 0.7% Latino or Hispanic, and 0.3% Asian. Two-thirds were married and living with their spouse. Twenty-six percent of the patients completed college, 21% completed high school and attended some college, 23% completed high school, and 30% did not complete high school.

The majority (92%) of patients had osteoarthritis, 3% had rheumatoid arthritis or lupus, and 5% had osteonecrosis. Most of the patients (72%) had experienced the onset of symptoms of hip arthritis 1 to 5 years before surgery, and 15% had symptoms for 5 to 10 years. Only 5% had symptoms for less than 1 year, whereas 7% had symptoms for more than 10 years. Overall, the mean duration of symptoms was 9.2 ± 8.8 years. The course of symptoms over the year before surgery was described as “rapidly progressive” or “much worse” by 47% of patients or as “slowly progressive” and only “a little worse” by 9% of the patients. Almost half (45%) characterized their arthritis as worse than 1 year ago because the flares were more frequent and more intense. Although 57% of patients used no assistive device, 35% used a cane, 4% used a walker, and 4% used crutches.

All baseline and follow-up evaluations were conducted according to a standard protocol with a specific script for the questionnaires. The 6-minute walking distance was performed in a standardized fashion at baseline only.[7, 8] We asked all patients to walk as far as they could in a level corridor during a 6-minute interval and with encouragement at prespecified intervals. The 6-minute walk test is suitable for elderly, frail, and severely limited patients and is also useful in patients with comorbid conditions such as heart failure and lung disease.[8] In addition to corresponding closely with the demands of everyday activities, this test is simple, inexpensive, and safe. The 6-minute walk test yields objective and reproducible results when compared with other measures of functional status.[8] The 6-minute walk test was not performed at follow-up because most patient evaluations were conducted by telephone or mail.

We used the IADL elements of the Arthritis Impact Measurement Scales (AIMS) to assess the impact of hip arthritis in activities related to independent daily living.[9] We administered the AIMS IADL items preoperatively and at follow-up intervals of 1 and 5 years. This includes three of the classic IADL items: (1) using public transportation, (2) shopping, and (3) housework. The IADL items were based on self-rating and reported as completely dependent, requiring assistance to perform, or independent. The development of the IADL scale has been previously validated against other functional measures of overall health status.[9]

Comorbidity was classified by use of the Comorbidity Scale of Charlson et al.[10] This scale provides a prospective method of classification for comorbid conditions that alter the risk of mortality. The scale uses a weighted index that accounts for the number and seriousness of comorbid diseases.[10] The scale predicts the impact of individual diseases in patients.

To assess the impact of 6-minute walk test scores on patient outcome, we stratified patients into two groups based on the ability to walk less than 125 meters or 125 meters or greater in a 6-minute period. Patients were also stratified according to the level of baseline functional impairment using the IADL assessment in the following categories: no impairments (n = 76), one impairment (n = 72), and two or more impairments (n = 90).

Of the 300 patients who were eligible for follow-up, 238 were contacted and completed follow-up. The minimum follow-up was 5 years with a median of 6 years, a mean of 5.6 years, and a range of 5 to 9 years. There was no difference in age, gender, diagnosis, comorbidity, type of fixation, bone density, or baseline hip function between patients who were lost versus those who were not. We performed the final assessment between 5 and 9 years

postoperatively: 22% at 5 years, 29% at 6 years, 28% at 7 years, 18% at 8 years, and 3% at 9 years. The results did not differ according to the year of final follow-up.

The patients who died were older (72 versus 63 years, $p < 0.0003$) and had more ($p = 0.005$) comorbidity but otherwise did not differ from the patients who were alive at follow-up. Thus, data were available on 238 patients who had completed more than 5 years of follow-up. The 238 patients represent 79% of the enrolled population of 300 patients.

We analyzed the data in two ways. In one set of analyses, the last follow-up was used for all patients and 96% of the original cohort was included. We excluded only the 13 patients without any follow-up. The second set of analyses focused on only the 238 patients who had 5-year follow-up or more and 79% of the cohort was included. There were no differences in the results.

We used general linear models to evaluate the differences in HRQ scores controlling for age and gender.

The principal outcome is total HRQ score at 5 years. However, within-patient change between baseline and 5 years was also analyzed to provide additional insights with regard to patients who have the least or most benefit. Regression analysis was performed using SAS GLM (8.02; SAS Institute, Cary, NC). Variables that were significant predictors of outcome in univariate regression analysis were entered into the multivariate analysis. The multivariate analysis included age, gender, walking distance, comorbidity, and impairment in IADL.

Results

The mean baseline 6-minute walking distance was 246 meters and the interquartile range was 126 meters (25^{th} percentile) to 332 meters (75^{th} percentile). Sixty percent of the patients were able to complete the test without stopping. Of those who were forced to stop, 83% stopped because of hip pain, 7% because they were tired, 5% because they were short of breath, and another 5% for other reasons. HRQ total scores and four subscales (pain, functioning, walking, overall impact of hip arthritis) and within-patient differences between baseline and 5 years for the entire patient population were collected to determine change over time; however, scores at 24 months and 36 months were virtually identical to those seen at 12 months (Table 1). Outcomes as measured by the HRQ at 5-year follow-up were similar to the assessments at 1 year postoperatively. Improvements in pain, walking, function, and overall impact of hip arthritis that were initially achieved by THA were sustained over time. The greatest improvements from baseline were in the HRQ pain and overall impact score. The improvements in the mean pain and overall impact subscales were 10.7 and 10.5, respectively. The mean total HRQ score rose from 57.2 to 87, an improvement of 29.8 points. Twelve points on the HRQ scale reflects a clinically important difference.[4]

Measure	Baseline	1 Year Postoperatively	5 Years Postoperatively	Within-patient Difference: Baseline To 5 Years Postoperatively
Total HRQ	57.9 ± 13.7	89.2 ± 10.9	87.0 ± 13.8	29.1 ± 15.6
Pain	10.3 ± 4.1	21.8 ± 4.5	21.0 ± 4.7	10.7 ± 5.4
Function	20.7 ± 2.7	23.2 ± 2.2	23.3 ± 2.3	2.5 ± 2.6
Walk ability	16.2 ± 4.0	21.9 ± 3.5	21.9 ± 4.4	5.7 ± 4.5
Overall impact	10.8 ± 6.4	21.9 ± 4.4	20.9 ± 5.2	10.1 ± 7.9

Table 1: HRQ = Hip Rating Questionnaire Hip Rating Questionnaire Scores at Baseline, 1 and 5 Years Postoperatively for the Entire Patient Population (n = 238)

HRQ scores were analyzed according to 6-minute walking distance, IADL, and comorbidity (Table 2). The majority (77%) of patients had a Charlson comorbidity scale score of 0, signifying a relatively healthy population. Fifteen percent had a score of 1 and 5% had a score of 2. Three percent had a score of 3 or more. At baseline, in a multivariate analysis controlling for age and gender, patients with lower walking distance had lower ($p < 0.0001$) HRQ scores and patients with greater IADL impairments had lower ($p < 0.0001$) baseline HRQ scores. Patients with more comorbid disease also had slightly, but not significantly, lower baseline HRQ scores.

Preoperative Walking Distance (meters in 6 minutes)	Number	Baseline HRQ Score	5-year HRQ Score	Within-patient Change Baseline to 5 Years
125 meters or greater*	184	60.8 ± 12.1	88.8 ± 12.0	28.0 ± 13.5
Less than 125 meters†	54	48.2 ± 13.9	80.9 ± 17.3	32.7 ± 20.9
IADL impairment (number of impairments)				
0	76	68.6 ± 9.6	92.9 ± 9.5	24.3 ± 12.1
1	72	60.3 ± 10.0	88 ± 13.0	27.7 ± 24.3
2 or more	90	47.0 ± 10.7	81.2 ± 15.1	34.2 ± 17.6
Comorbidity				
1 or less	209	58.8 ± 13.4	88.2 ± 12.4	29.4 ± 14.9
2 or more	29	51.9 ± 13.9	78.5 ± 19.5	26.5 ± 20.2

Table 2: Total HRQ Scores Related to Preoperative 6-minute Walking Distance, IADL, and Comorbidity Means ± standard deviation; *patients with walking distances 125 meters or more in 6 minutes; †patients with walking distances less than 125 meters in 6 minutes; HRQ = Hip Rating Questionnaire; IADL = instrumental activities of daily living

The principal outcome of interest was the 5-year HRQ score. Controlling for age and gender, patients with greater IADL impairment at baseline had worse ($p < 0.0002$) HRQ scores at 5 years. At 5 years, patients with the greatest IADL impairment at baseline were 10 HRQ points below the patients without IADL impairment. In the same analysis, patients with more comorbidity had worse ($p = 0.005$) HRQ scores at 5 years. In contrast, preoperative walking distance did not predict 5-year HRQ scores. The overall R^2 for the model was 0.20.

A secondary outcome was within-patient change in HRQ score between baseline and 5 years. In a multivariate analysis controlling for age and gender, the only predictor of greater within-patient improvement in HRQ was greater ($p = 0.01$) IADL impairment at baseline. The patient population with two or more baseline IADL impairments improved 10 HRQ points more than those without IADL impairment. Neither preoperative walking distance nor comorbidity was related to within-patient change.

Patients who were the most impaired in IADL gained the most, but their overall status at 5 years was lower than patients who did not have more than one IADL impairment at baseline. Thus, although THA provided substantial improvement in IADL over a sustained period, those who were the most impaired were at higher risk to remain impaired at 5-year follow-up. For example, of the 90 patients who had two or more impairments at baseline, 25% continued at the same level of impairment at 5 years (Table 3). Patients with comorbid disease also had less favorable 5-year outcomes.

Discussion

The 31% recruitment rate among eligible patients may be seen as a limitation of this study. Charlson and Horwitz reported the average proportion of eligible patients enrolled for National Institutes of Health-funded trials is 50%.[11] Although the proportion of patients enrolled in our study is lower than the National Institutes of Health average, the patients who were not entered were similar to those who were entered with respect to age, gender, and underlying disease. The same was true for entered versus non-entered patients within disease categories. A weakness in this study is a substantial gap between the time of recruitment and follow-up of this population and the analysis and reporting of the results. Although the population was prospectively recruited according to a standardized protocol, the database was "mined" with a hypothesis that is different from the original, which dealt with bone quality and surgical outcome. The baseline and outcome variables, however, are legitimate for the

Baseline IADL	Number	Number of IADL Impairments (5 years)	One IADL Impairment (5 years)	Two or More IADL Impairments (5 years)	
No impairment	76	95%	4%	1%	100%
1 impairment	72	65%	21%	14%	100%
2 or more impairments	90	59%	23%	25%	100%

Table 3. Relationship of Baseline IADL to IADL at 5 Years follow-up
IADL = instrumental activities of daily living

purposes of this study. The HRQ and the AIMS were the only two outcome questionnaires utilized. We acknowledge the validity of other published instruments, and believe that similar results would be achieved with their use.

Patient-reported outcomes have become the gold standard for assessing the effectiveness of treatments given to individuals and populations. The HRQ was the first validated patient-administered questionnaire designed exclusively for hip surgery. Since the validation of the HRQ, the American Academy of Orthopaedic Surgeons (AAOS) has developed several valid outcomes questionnaires dealing with all body regions, including a Lower Limb Core questionnaire, which is similar to the HRQ.[3] Because of the reliability of the AAOS instruments, they may be used to compare the effectiveness of multiple surgical procedures and the differences that result when applied to various populations. When measured prospectively, the change in scores from baseline to follow-up is a powerful indication of the effectiveness and hence the value of a procedure. Although various populations may achieve distinctively different final outcome scores, the change from baseline may not be considerably different as was demonstrated in this report when comparing the low and high comorbidity groups. The comorbidity scale was designed to measure the impact of comorbid disease on mortality but has proved helpful in explaining differences in other outcomes. Overall, the patients undergoing THA had a low burden of comorbidity, although 8% of patients had a comorbidity score of 2 or more. For example, patients with a high burden of comorbidity may be unable to walk longer distances or to regain independence in activities of daily living because of problems other than hip arthritis. In fact, this was true. Patients with higher comorbidity had worse total HRQ scores at 5 years. In the case of IADL, however, the more severely impaired patients actually had considerably larger improvement than less affected patients. Although the measure of patient satisfaction may have provided an additional dimension of interest, we did not include it in this study.

When the emphasis is placed only on the mean final scores for determination of the value of a particular procedure, patient populations with lower mean baseline and follow-up scores may be perceived as having inferior results, although the actual benefit of surgery may have been higher. On the other hand, if the less severely impaired patients are preferentially selected to test a particular surgical procedure, the higher follow-up scores may obscure the reality that there was less of a net benefit achieved by performing THA. Ultimately, a procedure's effectiveness should be viewed from the standpoint of its ability to generate a clinically important improvement from the baseline state rather than focusing on an isolated final outcome score. This has more than merely philosophic implications. When applying a risk-benefit analysis, the sicker and more impaired patients may have a higher surgical risk, but there may also be a higher potential benefit. If the potential benefit is overlooked in the consideration of surgical indications, it is theoretically possible that patients who might be good surgical candidates, but are becoming progressively dependent with deteriorating function, will be denied access to THA. At the other extreme, ignoring the substantially higher baseline scores in patients with less comorbidity and better function, and focusing on the higher expected outcome scores, will lead to the selective recruitment of this group of patients for THA. Studies designed to demonstrate superior outcomes for any treatment modality or new

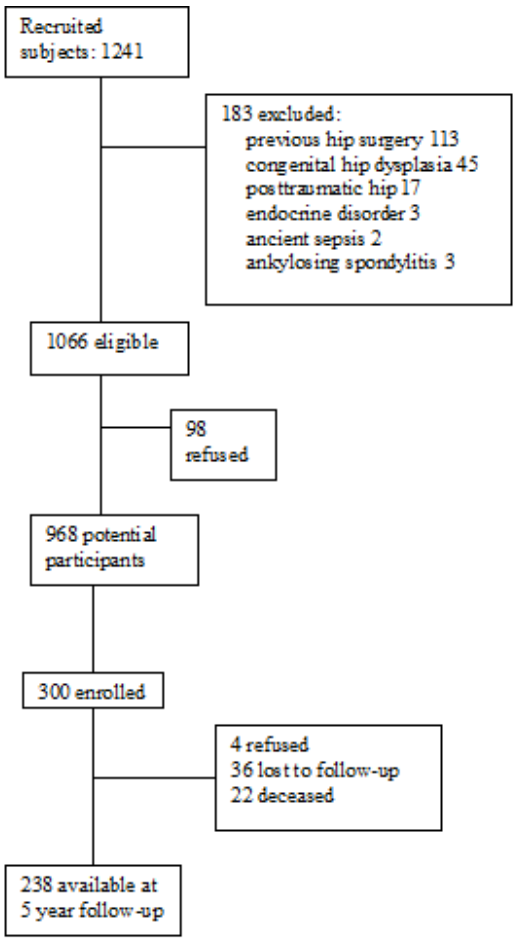

Figure 1. CONSORT diagram: participant flow through study

technology would naturally benefit by including this healthier and functionally less impaired population in the study. The current competitive healthcare environment often places extreme pressure on physicians to market their services. In addition, report cards are becoming an accepted industry standard in the marketplace, in many locations being publicly issued for the purpose of demonstrating differences between expected and observed patient outcomes. Given the potential power of these forces on the physician's decision-making process, there is reason to have concern that sicker and more functionally impaired patients will have an increasing problem gaining access to THA unless more sophisticated risk adjustment strategies are widely adopted. The foundation for developing an evidence-based risk adjustment system includes a commonly accepted system of baseline patient characteristics that may have impact on the surgical outcome. Further study is needed to identify the risks and benefits of THR in patients with high baseline comorbidity and functional impairment. Ethgen's literature review suggested that greater benefits from THA were realized by those having poorer health-related quality of life. The results of this study are in agreement with that observation.

Acknowledgment

We thank the following surgeons from the Hospital for Special Surgery, New York, NY, for facilitating the recruitment of their private patients into this study: Chitranjan S. Ranawat, MD, Eduardo A. Salvati, MD, Thomas P. Sculco, MD, and Philip D. Wilson, Jr, MD.

References

1. Ethgen O, Bruyere O, Richy F, Dardennes C, Reginster JY. Health-related quality of life in total hip and total knee arthroplasty. A qualitative and systematic review of the literature. J Bone Joint Surg Am. 2004; 86:963-974.
2. Laupacis A, Bourne R, Rorabeck C, Feeny D, Wong C, Tugwell P, Leslie K, Bullas R. The effect of elective total hip replacement on health-related quality of life. J Bone Joint Surg Am. 1993; 75:1619-1626.
3. Johanson NA, Liang MH, Daltroy L, Rudicel S, Richmond J. American Academy of Orthopaedic Surgeons lower limb outcomes assessment instruments. J Bone Joint Surg Am. 2004; 86:902-909.
4. 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 Am. 1992; 74:587-597.
5. Johanson NA, Charlson ME, Cutignola L, Neves M, DiCarlo EF, Bullogh PG. Femoral neck bone density. Direct measurement and histomorphometric validation. J Arthroplasty. 1993; 8:641-652.
6. Johanson NA, Neves MC, Bansal M, DiCarlo EF, Bullough PG. Cross sectional densitometric and histomorphometric analysis of the femoral neck. Transactions of the 35th Annual Meeting, Orthopaedic Research Society. 1989; 14:455.
7. Guyatt GH, Pugsley SO, Sullivan MJ, Thompson PJ, Berman L, Jones NL, Fallen EL, Taylor DW. Effect of encouragement on walking test performance. Thorax. 1984; 39:818-822.
8. Guyatt GH, Sullivan MJ, Thompson PJ, Fallen EL, Pugsley SO, Taylor DW, Berman LB. The 6-minute walk: a new measure of exercise capacity in patients with chronic heart failure. Can Med Assoc J. 1985; 132:919-923.
9. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist. 1969; 9:179-186.

10. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987; 40:373-383.
11. Charlson ME, Horwitz RI. Applying results of randomized trials to clinical practice: impact of losses before randomization. BMJ. 1984; 289:1281-1284.