International Journal of Technology Assessment in Health Care, 16:1 (2000), 165–177.

EVALUATION OF THE APPROPRIATENESS OF HIP JOINT REPLACEMENT TECHNIQUES

José M. Quintana
Hospital de Galdakao

Jesús Azkarate
Hospital del Bajo Deba

J. Ignacio Goenaga
Hospital de Santiago

Inmaculada Aróstegui
Universidad del País Vasco

Ignacio Beldarrain
Hospital de Aránzazu

Jose M. Villar
Hospital de Crúces

Abstract

Objective: To evaluate the appropriateness of the use of hip joint replacements (HJRs) using explicit criteria developed by an expert panel.
Methods: Observational study. Nine hundred ninety-seven patients from five hospitals with a diagnosis of osteoarthritis, avascular necrosis, hip fracture, or revision who were undergoing HJR were consecutively included in the study during a 1-year period. The appropriateness of the indication was judged by explicit criteria. Complications were recorded at the time of the intervention and 3 months postoperatively.

We thank Dr. Pablo Lazaro from the Unidad de Investigación de Servicios Sanitarios of the Instituto de Salud Carlos III, Madrid, for his assistance in the development of the algorithm and the panel debate and Dr. James Kahan for reviewing the manuscript. We also thank the following individuals for their contributions to this study: Drs. Jose M. Aranburu, Andoni Arcelay, Jesús Azkoaga, Pedro Armendariz, Enrique Cáceres, Xabier Elexpe, Bengoña Goicoetxea, Jon Letona, Manuel Martínez-Grande, Enrique Queipo de Llano, Ramón Tobio, and Ignacio Vidaurreta, and staff members from the Quality Units from the participating hospitals.

This work was supported by a grant from the Fondo de Investigación Sanitaria (96/0020-05) and the Department of Health of the Basque Government.

Preliminary results of this study were presented at the 13th Annual Meeting of the International Society of Technology Assessment in Health Care, Barcelona, 1997.

Results: Of the 1,030 interventions, 604 were for osteoarthritis, 31 avascular necrosis, 191 fractures, and 204 revisions. No differences were found among the hospitals for the main clinical and patient variables. Indications for surgery were considered appropriate in 59% of cases, uncertain in 32%, and inappropriate in 8%, mainly in the osteoarthritis group. Differences were found in the rates of appropriateness among some centers. The complication rate did not differ among the groups based on the level of appropriateness of the procedure.
Conclusions: The appropriate use of HJR, as measured by the criteria established by the panel, identified a moderate percentage of inappropriate indications. Those equivocal and inappropriate cases demand further studies to identify patients with an adequate risk-to-benefit ratio from this procedure.

Keywords: Hip prosthesis, appropriateness, Utilization review

Inexplicable variations in surgery rates, identification of inappropriate care, and escalating healthcare costs raise questions about potential underuse or overuse of many medical and surgical procedures (42). Healthcare systems should function such that appropriate care increases and inappropriate care decreases. Reducing overuse should enhance quality of care and decrease medical costs (1;35).

Central to this investigation is the determination of what is considered an appropriate indication for any given procedure. Unfortunately, rigorous scientific data are lacking on the efficacy and effectiveness to justify medical practice. For most conditions, something other than rigorous data on efficacy or effectiveness must be used to determine the criteria of appropriateness (7). The results of surveys disagreed when orthopedic surgeons were asked about indications for some orthopedic procedures (38).

A method that combines expert opinion with available scientific evidence was developed by the RAND–University of California at Los Angeles (UCLA) group. This method has been used to evaluate the appropriateness of a variety of medical and surgical interventions (6;28).

Hip joint replacement (HJR) can produce dramatic improvements in patients' functional status and health-related quality of life (18;20). Its use is still increasing in developed countries (13;23). At the same time, surgical rates continue to vary across regions of different countries (5;29), a finding that cannot be explained solely by differences in the prevalence of hip disease.

The goal of this study was to apply explicit criteria, developed using a multidisciplinary approach, to examine the appropriateness of the indications for HJR in various hospitals and diagnostic groups.

METHODS

Explicit Criteria Development

Criteria to measure the appropriate use of HJR were developed according to a previously described explicit method (2), consisting of the following steps.

First, an extensive literature review was conducted to summarize existing knowledge about efficacy, effectiveness, risks, costs, and opinions concerning the use of HJR on patients with a diagnosis of osteoarthritis, avascular necrosis, hip fracture, and revision.

Second, from this review, a comprehensive and detailed list of clinical scenarios (indications), each mutually exclusive and clinically specific, was developed in which HJR might be performed on those patients. The list contained 696 indications, and each was described in sufficient detail that patients within a given indication were

reasonably homogeneous. These indications included the following variables: age, pain (categorized as minor, moderate, or severe based on need for medication and the effect on pain; relation to rest, sleep, or night disturbance; rhythm; and intensity [21;24]), functional limitations assessment (categorized as mild, moderate, or severe based on the American College of Rheumatology [ACR] classification [11], and the need for a mobility aid); bone quality, measured by x-ray according to the classification of Singh et al. (37); surgical risk (based on the American Society of Anesthesiologists [ASA] criteria [34]); and whether previous nonsurgical treatments were performed based on the current protocol (32). The avascular necrosis diagnosis group was classified according to the criteria of Ficat and Arlet (9). The variable "previous nonsurgical treatments" was not considered in this group or in the revision group. Except for the previous considerations in the necrosis and revision groups, the same variables of the osteoarthritis group applied to both. Only the fracture group included the type of fracture (divided into displaced or nondisplaced femoral neck fracture), age, and life expectancy (Figure 1).

Third, we selected a national panel that included nine orthopedic surgeons. Panelists were provided with the literature review and list of indications, and they rated each indication for its appropriateness, considering the average patient and average physician in 1997. Appropriateness was defined as indicating that the expected health benefits exceeded the expected negative consequences by a sufficiently wide margin such that HJR was worth performing.

Ratings were made on a 9-point scale. The use of HJR for a specific indication was considered appropriate if the panel's median rating was between 7 and 9 without disagreement, inappropriate if the value was between 1 and 3 without disagreement, or uncertain if the median rating was between 4 and 6 or if the members of the panel disagreed. Disagreement was defined as occurring when at least three panelists rated an indication from 1 to 3 and at least three others rated it from 7 to 9. This method did not attempt to force panelists to reach agreement on appropriateness. The ratings were confidential and took place in two rounds using a modified Delphi process.

The first round of ratings was performed privately and anonymously. During a subsequent meeting of the panel members and after extensive discussion, the panelists revised the indications according to the previously described definition. Two more scenarios were added. Each panelist rated 698 separate indications. All indications considered appropriate were rated by the panelists on a 9-point scale to determine those that were considered necessary. Necessary was defined as meaning that a procedure is not only appropriate but also crucial and that it would be improper care not to recommend it in a given clinical situation, as defined by previous authors (15). Indications with a median necessity rating from 7 to 9 and with no disagreement were considered necessary. The other appropriate indications were considered elective.

Data Collection

This prospective observational study was conducted in five large public hospitals (four university-affiliated and one community-based), starting in December 1996 and ending in December 1997. For those patients with a diagnosis of hip fracture, the design was retrospective because of the urgent nature of the process. The identities of the hospitals and surgeons were not revealed in the research reports. Physicians at each hospital were blinded to the study goals.

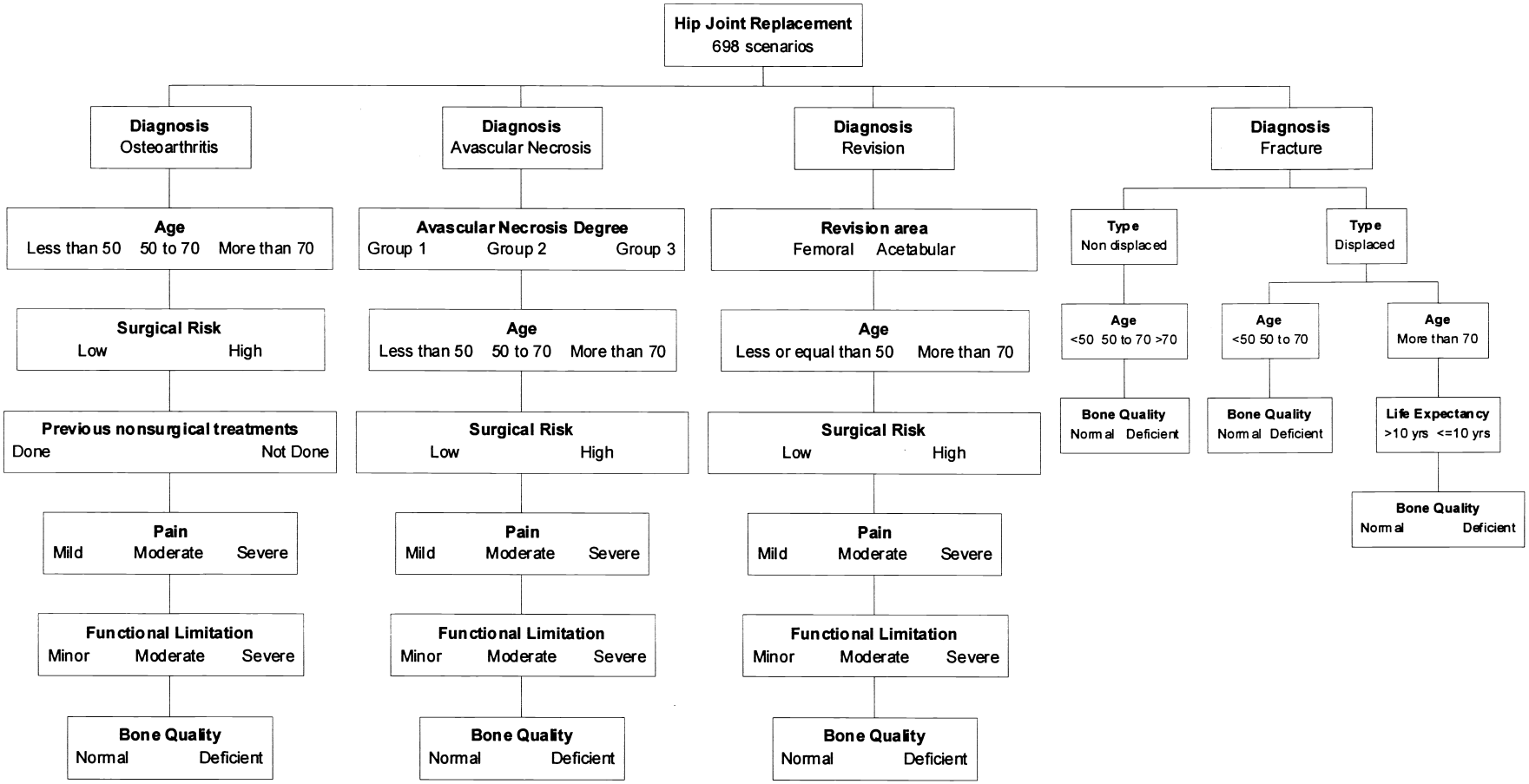

Figure 1. Variables of the algorithm and their categories.

All consecutive patients with a diagnosis of osteoarthritis, avascular necrosis, or revision who were to undergo HJR and who were followed in any of the five hospitals were included in the study. Patients with malignant, severe, or psychiatric diseases who were unable to communicate or who refused to participate were excluded. Of 850 patients from the prospective study who fulfilled the selection criteria, 806 (94.6%) patients agreed to participate, and 191 patients with hip fractures who were undergoing HJR during the same period of time also were included.

To collect data and determine appropriateness, we developed a computerized algorithm based on the panel results. We also developed data collection questionnaires that included variables previous to the intervention, admission, and discharge, including the intervention, and complications 3 months after discharge. In addition to those belonging to the appropriateness algorithm, other variables included sociodemographic data, patient's family support, height, weight, primary complaint, 12 comorbidities, other joints affected, previous interventions, intervention characteristics, local and general complications peri- and post-intervention, length of hospital stay, and death.

One trained reviewer, who was unaware of the study goals, collected the data via a standardized questionnaire. He located all patients who fulfilled the selection criteria and interviewed them preoperatively, at which time permission for inclusion in the study was also obtained from all patients. A standardized questionnaire also was completed by a surgeon at each center. At discharge, data related to the admission and intervention were gathered by reviewing the medical records. Three months after discharge, all medical records were again reviewed to ascertain if patients had been readmitted, died, or had any complication resulting from the intervention.

Statistical Analysis

The unit of study was the number of interventions. Descriptive statistics, frequency tables, and mean and standard deviations were used. The chi square and Fisher's exact test were used to test for statistical significance. For continuous variables such as age, the ANOVA test was performed in the univariate analysis.

Multivariate logistic regression models were used to study the relation between two dependent variables: the development of any complication during the intervention or after discharge (yes/no), and the main independent variable, the appropriateness judgment (12). Models were adjusted by other covariates, such as the hospital and the diagnosis group. General linear models were used to study the relationship between the length of stay and the appropriateness judgment, adjusted by hospital and diagnosis group (41).

All effects were considered statistically significant when $p < .05$, unless otherwise noted. All statistical analyses were performed using the SAS for Windows, version 6.12, statistical software (33).

RESULTS

During the 1-year recruitment period, 997 patients were included (1,030 interventions) who were to undergo HJR at any of the five participating hospitals. The mean patient age was 68.8 years, and 54.6% were females. Based on diagnosis, 58.6% of the patients had osteoarthritis and 3.1%, avascular necrosis. Revisions comprised 19.8% and fractures, 18.5%; use of a partial prosthesis made up 87.0%

of the latter. Total HJRs were performed in 61.7%, mainly in patients with osteoarthritis.

There were no differences among the five hospitals regarding the main patient sociodemographic variables (sex, marital status, family support), other joints affected, previous hip interventions, or any of the 12 comorbidities recorded. More revisions (32.8%) were performed in hospital 2, and there were more patients with osteoarthritis in hospital 3. Hospitals 3 and 4 had the highest proportion of fractures (26.0% and 28.4%, respectively), and hospital 2 the lowest (2.7%). As a result, the mean patient age was higher in hospitals 3 and 4 (70 and 70.9 years, respectively) compared with hospital 2 (65.6 years). Surgical risk, assessed by the ASA classification system, differed among centers (Table 1).

After applying the explicit criteria to the 1,030 interventions, we found that 610 (59.3%) of the cases were considered appropriate, and 93.4% of these were necessary. The panel was uncertain about the appropriateness of 334 (32.4%) of the interventions, and they considered 86 (8.3%) to be inappropriate. Most cases deemed uncertain were in the osteoarthritis group (46.2%). There were no inappropriate cases in the fracture and revision groups. The cases deemed inappropriate were concentrated in the osteoarthritis (13.6%) and avascular necrosis (12.9%) groups (Table 2). The most frequently found scenarios are displayed in Table 3.

Regarding the differences among centers (Table 4), hospital 4 had the lowest number of interventions considered inappropriate, and one of the highest number of appropriate interventions. Hospital 2 had one of the lowest rates of appropriateness in all groups, except for revisions. Apart from the minor avascular necrosis group, statistically significant ($p < .05$) differences among centers were concentrated in the osteoarthritis group. In this group, hospitals 2 and 5, respectively, had 15.8% and 16.3% of interventions that were considered inappropriate, while hospital 4 had 6.7%. No differences were found among hospitals in the other diagnostic groups. All cases deemed appropriate were found to be necessary interventions in the revision group and most of the fracture group.

Of all recorded complications, no specific complication that resulted from the intervention or developed 3 months after discharge occurred more frequently in any appropriateness category. Though the univariate analysis showed a statistically significant difference, there was a higher probability of developing a complication for those interventions considered necessary, either during the intervention or 3 months after discharge. However, the multivariate analysis, which included an adjustment by hospital and diagnostic group, no longer showed those differences (Table 5). In the univariate analysis, the mean length of hospital stay differed among appropriateness categories, i.e., shorter for those considered inappropriate, uncertain, or elective than for those considered necessary. Those differences disappeared after controlling for the previous covariates.

DISCUSSION

Although the benefit of hip prosthesis intervention has been proven in patients with severe pain or functional limitation, there are no conclusive data from clinical trials regarding the risk-to-benefit ratio in different patient groups (39). Although variations in the use of HJR were not as high as with other procedures (4), the use of HJR does vary, which justifies the study of its appropriate use.

We based our study on methodology broadly used to create explicit criteria that allowed us to judge the indication of hip prosthesis in the most common

Table 1. Patient Statistics by Hospital

Variable	Hospital 1	Hospital 2	Hospital 3	Hospital 4	Hospital 5	*p* Value[a]
n (%)	184 (17.9)	183 (17.8)	177 (17.2)	183 (17.8)	303 (29.4)	
Age $\bar{x}$ (SD)[a]	68.6 (11.7)	65.6 (11.2)	70.0 (10.6)	70.9 (10.9)	68.9 (10.9)	<.01
Diagnosis						<.01
Osteoarthritis	118 (64.1)	114 (62.3)	116 (65.5)	90 (49.2)	166 (54.8)	
Avascular necrosis	4 (2.2)	4 (2.2)	5 (2.8)	4 (2.2)	14 (4.6)	
Fracture	33 (17.9)	5 (2.7)	46 (26.0)	52 (28.4)	55 (18.2)	
Revision	29 (15.8)	60 (32.8)	10 (5.6)	37 (20.2)	68 (22.4)	
Surgical risk						<.01
ASA I, II, or III	178 (96.7)	170 (92.9)	160 (90.4)	176 (96.2)	266 (87.8)	
ASA IV	6 (3.3)	13 (7.1)	17 (9.6)	7 (3.8)	37 (12.2)	

Abbreviations: ASA = American Society of Anesthesiologists; SD = standard deviation.
[a] Chi square and Fisher's exact test, except for age (ANOVA test).

Table 2. Ratings of Appropriateness by Diagnostic Group[a]

	Appropriate[a]				
	Necessary	Elective	Uncertain	Inappropriate	Total
Osteoarthritis	211 (34.9)	32 (5.3)	279 (46.2)	82 (13.6)	604
Avascular necrosis	16 (51.6)	3 (19.7)	8 (25.8)	4 (12.9)	31
Fracture	158 (82.7)	5 (2.6)	28 (14.7)	0 (0)	191
Revision	185 (90.7)	0 (0)	19 (9.3)	0 (0)	204
Total	570 (55.4)	40 (3.9)	334 (32.4)	86 (8.3)	1,030

[a] The group with appropriate indications was divided based on the panel scores on those indications that were considered necessary interventions or appropriate but unnecessary (elective).

diagnostic groups in our area. Applied to a group of patients with a variety of common diagnoses at five large acute care hospitals, a relevant percentage of indications were judged inappropriate. The results also call attention to the percentage of cases classified as uncertain, especially in the osteoarthritis group.

No cases were found to be inappropriate in the fracture or the revision groups, and the percentage of cases judged to be uncertain was low. In the case of fractures, the variability in its use was reported to be minor in some studies (17). Some reasons that might explain why revisions were considered appropriate indications

Table 3. Description of the Most Frequent Indications Encountered in the Study in Each Ratings Category of Appropriateness

Description of the indications (scenarios)	Rating	Number of patients (%)[a]
Osteoarthritis[b]		
Age 50–70, moderate/severe pain and severe functional limitation	Appropriate	54 (22.2)
Age 50–70, moderate pain and moderate functional limitation	Uncertain	64 (22.9)
Age 50–70, minor pain and minor functional limitation	Inappropriate	12 (14.6)
Avascular necrosis[c]		
Age $<$ 50, severe pain and functional limitation, grade 3	Appropriate	3 (15.7)
Age $<$ 50, moderate pain and severe functional limitation, grade 3	Uncertain	3 (37.5)
Age 50–70, moderate pain and functional limitation, grade 1	Inappropriate	3 (75.0)
Fracture[c]		
Age $>$ 70, displaced fracture and bone quality deficient	Appropriate	125 (76.6)
Age 50–70, displaced fracture and bone quality deficient	Uncertain	16 (57.1)
Revision[c]		
Age $\leqslant$ 70, moderate pain and functional limitation	Appropriate	42 (22.7)
Age $>$ 70, minor pain and functional limitation	Uncertain	7 (36.8)

[a] Numbers in brackets are the percentages of each indication from its appropriateness and diagnostic group.
[b] Surgical risk was low and Singh class $>$ 3 in all patients. Appropriate category had previous nonsurgical treatments correctly done; the other categories did not.
[c] Surgical risk was low in all cases.

Table 4. Ratings of Appropriateness by Diagnostic Group and Hospital[a]

	Hospital 1	Hospital 2	Hospital 3	Hospital 4	Hospital 5	*p* Value[b]
Osteoarthritis						
Necessary	35 (29.7)	39 (34.2)	44 (37.8)	33 (36.7)	60 (36.1)	NS[c]
Elective	9 (7.6)	5 (4.4)	3 (2.6)	6 (6.7)	9 (5.4)	
Uncertain	58 (49.2)	52 (45.6)	54 (46.6)	45 (50.0)	70 (42.3)	
Inappropriate	16 (13.6)	18 (15.8)	15 (12.9)	6 (6.7)	27 (16.3)	
Avascular necrosis						
Necessary	2 (50.0)	3 (75.0)	3 (60.0)	2 (50.0)	6 (42.9)	NS
Elective	0 (0.0)	0 (0.0)	2 (40.0)	1 (25.0)	0 (0.0)	
Uncertain	2 (50.0)	0 (0.0)	0 (0.0)	1 (25.0)	5 (35.7)	
Inappropriate	0 (0)	1 (25.0)	0 (0)	0 (0)	3 (21.4)	
Fracture						
Necessary	27 (81.8)	3 (60.0)	39 (84.8)	47 (90.4)	42 (76.4)	NS
Elective	2 (6.1)	0 (0.0)	1 (2.2)	0 (0.0)	2 (3.6)	
Uncertain	4 (12.1)	2 (40.0)	6 (13.0)	5 (9.6)	11 (20.0)	
Revision						
Necessary	28 (96.6)	55 (91.7)	9 (90.0)	30 (81.1)	63 (92.7)	NS
Uncertain	1 (3.4)	5 (8.3)	1 (10.0)	7 (18.9)	5 (7.3)	

[a] There were no inappropriate cases in the fracture and revision groups in any hospital; there were no elective cases in the revision group.
[b] Chi square test; NS = not statistically significant.
[c] Differences statistically significant at $p < .05$, comparing the inappropriate rates of all hospitals to the one with the lowest (hospital 4).

Table 5. Relationship of Appropriateness Evaluation and Complications as a Result of the Intervention[a]

	Complications at surgery		Complications at 3 months	
	OR	95% CI[b]	OR	95% CI[b]
Appropriateness evaluation				
Necessary	1.7	0.54–7.50	1.39	0.70–2.96
Elective	1.7	0.20–9.90	0.73	0.19–2.33
Uncertain	0.92	0.28–4.16	1.02	0.52–2.18
Inappropriate	1	—	1	—

Abbreviations: OR = odds ratio; CI = confidence interval.
[a] Logistic regression models were adjusted by other variables such as diagnostic group and hospital.
[b] Odds ratio 95% confidence intervals.

in most cases were: a) the panelists rated most of the theoretical indications as appropriate, and b) when prosthesis mobilization is diagnosed, a revision is performed. The variability in the use of a hip prosthesis in those two groups is therefore irrelevant. The avascular necrosis group had a small number of patients and, although no conclusions could be drawn from our data, the results are similar to that of the ostheoarthritis group.

The main group of interest is the osteoarthritis group, because osteoarthritis is the most common diagnosis and the diagnosis in which most of the total hip prosthesis interventions were concentrated. Most of the interventions judged inappropriate and uncertain were in this group, which warrants careful investigation of the criteria used to indicate the intervention and other characteristics of these patients. It seems necessary to ascertain with more detail who undergoes the intervention and their results, which would require knowing the patients' perception of their quality of life and measuring their quality-of-life improvement and their satisfaction with the intervention.

Our results also showed that there were differences among hospitals. Because all the hospitals belong to the same public health system, share similar technologic and human resources, and the primary patient sociodemographic and clinical variables were similar, further study is needed to determine why the inappropriateness rates for some diagnoses ranged from 7% to 16%, depending on the center. It is also relevant to study what happens in other smaller centers, both private or those associated with the public health system.

The only previous study in which explicit criteria have been developed to assess HJR oriented its 120 indications to general practitioner referral to the specialist (27). Osteoarthritis was the only diagnostic group, and the algorithm used by those investigators did not include surgical risk or bone-quality assessment. The retrospective field study carried out by this group (40), using those criteria, presented some important methodologic problems. Based on a review of medical records, the researchers lacked relevant data from almost 50% of the 329 cases recruited. When they applied their algorithm, the inappropriateness rates ranged from 4.4% to 15.8%, similar to those found in the osteoarthritis group in the present study. The prospective design of the present study allowed us to capture and follow all patients who fulfilled the selection criteria, thus minimizing losses. When compared with other studies in which similar methodology was applied to other procedures

(3;19;28), our results indicated that a moderate percentage of potentially inappropriate procedures was performed.

Compared with another study of a different procedure, which examined the relation between the appropriateness evaluation and rate of complications (43), we did not find significant differences between the level of appropriateness and the peri- or post-intervention complication rate or other quality indicators, such as the death rate or length of hospital stay, after controlling for other possible confounders. However, even though the univariate analysis initially showed differences among appropriateness groups, this was a result of the fact that patients in the fracture group had confounders such as a higher complication rate, longer hospital stays, and higher surgical necessity rate. Nevertheless, the fact that the patients who were considered to have undergone inappropriate indications had complication rates similar to the other categories implies an important risk to the patient. If the indication was inappropriate, the risk was unnecessary.

This study has certain limitations. The limitations and criticisms related to the explicit criteria created by the RAND/UCLA method have been reported in previous publications (10;16;30;36). Even so, the method has been accepted as an important tool for evaluating the care provided and for studying variations (31). When developing our criteria, we followed recommendations made for this kind of study (26). However, in the absence of scientific evidence, the created criteria seem to be based more on opinions than on results (25;36).

The data collection process also had some limitations. The single blinded reviewer was a physician trained to assess and record the main variables of the algorithm in a standardized manner to reduce the chance of bias, but a certain degree of bias is inherent. In all diagnostic groups, except the fracture group, we based our criteria to classify a patient according to the level of pain and functional limitation on classifications developed by reputed medical societies (ACR) (11) or other investigators (21;24). However, information bias that may influence the final judgment cannot be disregarded. Our algorithm, used as the current clinical decision-making tool in this case, is based fundamentally on pain and the functional limitations assessment of the patient. However, only prospective studies can accomplish this assessment, because these subjective data (level of pain or functional limitation) are not available on the medical record in a standardized manner.

A question that may arise is why patients with moderate pain or functional limitations undergo the procedure. Discrepancies between physicians' and patients' evaluations of their disease status have been reported for several diseases, including osteoarthritis (22). Other variables may have influenced physicians' decision making in relation to specific patients.

Finally, this study presents the overuse results of a medical procedure but did not consider underuse (15), which may exist and has been recognized as an equally important factor in determining quality of care.

As suggested by some authors (25;36), this method may be useful when comparing levels of appropriate procedures among populations but not to direct care for individual patients. When used as a utilization review tool, interventions considered inappropriate underwent an individualized and in-depth revision before being considered inappropriate (8).

Policy Implications

The percentages of uncertain and inappropriate cases demand an in-depth long-term investigation to determine the consequences of not following the panel's

recommendations. As different authors and organizations have suggested (14;29), it might be relevant to investigate which treatment is most beneficial in certain patient groups, e.g., those younger than 70 years or those with minor or moderate pain and functional limitations.

Besides all the limitations of our study, the results show that there is some variability in the appropriateness of a commonly performed orthopedic procedure. More research is needed to delineate the risk-to-benefit ratio for different patients.

REFERENCES

1. Brook, R. H. Appropriateness: The next frontier (editorial). *British Medical Journal*, 1994, 308, 218–19.
2. Brook, R. H., Chassin, M. R., Fink, A., et al. A method for the detailed assessment of the appropriateness of medical technologies. *International Journal of Technology Assessment in Health Care*, 1986, 2, 53–63.
3. Casparie, A. F. The ambiguous relationship between practice variation and the appropriateness of care: An agenda for further research. *Health Policy*, 1996, 35, 247–65.
4. Chassin, M. R., Brook, R. H., Park, R. E., et al. Variation in the use of medical and surgical services by the Medicare population. *New England Journal of Medicine*, 1986, 314, 285–90.
5. Cohen, M. M., de Boer, D., & Young, W. Total hip replacement. In C. D. Naylor, G. M. Anderson, & V. Goel (eds.), *Patterns of health care in Ontario*. Ottawa: Canadian Medical Association, 1994, 72–76.
6. Coulter, I. Manipulation and mobilization of the cervical spine: The results of a literature survey and consensus panel. *Journal of Musculoskeletal Pain*, 1996, 4, 113–23.
7. Dubinsky, M., & Ferguson, J. H. Analysis of the National Institutes of Health Medicare coverage assessment. *International Journal of Technology Assessment in Health Care*, 1990, 6, 480–88.
8. Dubois, R. W. Appropriateness studies. *New England Journal of Medicine*, 1994, 330, 433.
9. Ficat, P., & Arlet, J. Coxopathies ischémiques. *Revue de Chirurgie Orthopédique et Réparatrice de l'Appareil Moteur*, 1972, 58, 543–61.
10. Hicks, N. R. Some observations on attempts to measure appropriateness of care. *British Medical Journal*, 1994, 309, 730–33.
11. Hochberg, M. C., Chang, R. W., Dwosh, I., et al. The American College of Rheumatology 1991 revised criteria for the classification of global functional status in rheumatoid arthritis. *Arthritis and Rheumatism*, 1992, 35, 498–502.
12. Hosmer, D. W., & Lemeshow, S. *Applied logistic regression*. New York: Wiley, 1989.
13. *Hospital Episode System Data. England 1989/90*. London: OPCS, DHSS, 1992.
14. Imamura, K., Gair, R., McKee, M., & Black, N. Appropriateness of total hip joint replacement in the United Kingdom. *World Hospitals and Health Services*, 1997, 32, 10–14.
15. Kahan, J. P., Bernstein, S. J., Leape, L. L., et al. Measuring the necessity of medical procedures. *Medical Care*, 1994, 32, 357–65.
16. Kahan, J. P., Park, R. E., Leape, L. L., et al. Variations by specialty in physician ratings of the appropriateness and necessity of indications for procedures. *Medical Care*, 1996, 34, 512–23.
17. Keller, R. B., Soule, D. N., Wennberg, J. E., & Hanley, D. F. Dealing with geographic variations in the use of hospitals. The experience of the Maine Medical Assessment Foundation Orthopaedic Study Group. *Journal of Bone and Joint Surgery: American Volume,* 1990, 72, 1286–93.
18. Kramer, A. K., Steiner, J. F., Schlenker, R. E., et al. Outcomes and cost after hip fracture and stroke. *JAMA*, 1997, 277, 396–404.
19. Larequi-Lauber, T., Vader, J. P., Burnand, B., et al. Appropriateness of indications for surgery of lumbar disc hernia and spinal stenosis. *Spine*, 1997, 22, 203–09.
20. Laupacis, A., Bourne, R., Rorabech, C., et al. The effect of elective total hip replacement on health-related quality of life. *Journal of Bone and Joint Surgery*, 1993, 75-A, 1619–26.

21. Lequesne, M. Indices of severity and disease activity for osteoarthritis. *Seminars in Arthritis and Rheumatism*, 1991, 20, 48–54.
22. Lieberman, J. R., Dorey, F., Shekelle, P., et al. Differences between patients' and physicians' evaluations of outcome after total hip arthroplasty. *Journal of Bone and Joint Surgery*, 1996, 78-A, 835–38.
23. Madhok, R., Lewallen, D. G., Wallrichs, S. L., et al. Trends in the utilization of primary total hip arthroplasty, 1969 through 1990: A population-based study in Olmsted County, Minnesota. *Mayo Clinic Proceedings*, 1993, 68, 11–18.
24. Merle, D. Cotation chiffre de la fonction de la hanche. *Revue de Chirurgie Orthopedique et Reparatrice de l'Appareil Moteur,* 1970, 56, 481–86.
25. Naylor, C. D. What is appropriate care? *New England Journal of Medicine*, 1998, 338, 1918–20.
26. Naylor, C. D., & Guyatt, G. Users' guide to the medical literature, XI: How to use an article about a clinical utilization review. *JAMA*, 1996, 275, 1435–39.
27. Naylor, C. D., & Williams, J. Primary hip and knee replacement surgery: Ontario criteria for case selection and surgical priority. *Quality in Health Care*, 1996, 5, 20–30.
28. Park, R. E., Fink, A., Brook, R. H., et al. Physician ratings of appropriate indications for three procedures: Theoretical indications vs indications used in practice. *American Journal of Public Health*, 1989, 79, 445–47.
29. Peterson, M. G. E., Hollenberg, J. P., Szatrowski, P. et al. Geographic variations in the rates of elective total hip and knee arthroplasties among Medicare beneficiaries in the United States. *Journl of Bone and Joint Surgery*, 1992, 74-A, 1530–39.
30. Phelps, C. E. The methodologic foundations of studies of the appropriateness of medical care. *New England Journal of Medicine*, 1993, 329, 1241–45.
31. Phelps, C. E. Appropriateness studies. *New England Journal of Medicine*, 1994, 330, 433–34.
32. Quintana, J. M. *Empleo de la metodología de uso apropiado en el estudio de la utilización de un procedimiento quirúrgico: Prótesis de cadera.* Leioa, Spain: The University of the Basque Country, 1998.
33. SAS Institute Inc. *SAS procedures guide*, version 6. Cary, NC: SAS Institute, 1994.
34. Schneider, A. J. L. Assessment of risk factors and surgical outcome. *Surgical Clinics of North America*, 1983, 63, 1113–26.
35. Schoenbaum, S. C. Toward fewer procedures and better outcomes. *JAMA*, 1993, 269, 794–96.
36. Shekelle, P., Kahan, J. P., Bernstein, S. J., et al. The reproducibility of a method to identify the overuse and underuse of medical procedures. *New England Journal of Medicine*, 1998, 338, 1888–95.
37. Singh, M., Nagrath, A. R., & Maini, P. S. Changes in trabecular pattern of the upper end of the femur as an index of osteoporosis. *Journal of Bone and Joint Surgery*, 1970, 52-A, 457–67.
38. Tierney, W. M., Fitzgerald, J. F., Heck, D. A., et al. Tricompartmental knee replacement: A comparison of orthopedic surgeons' self reported performance rates with surgical indications, contraindications and expected outcomes. *Clinical Orthopaedics and Related Research*, 1994, 305, 209–17.
39. Total hip replacement: NIH consensus conference. *JAMA*, 1995, 273, 1950–56.
40. Walraven, C. V., Peterson, J. M., Kapral, M., et al. Appropriateness of primary total hip and knee replacements in regions of Ontario with high and low utilization rates. *Canadian Medical Association Journal*, 1996, 155, 697–706.
41. Weisberg, S. *Applied linear regression*, 2nd ed. New York: Wiley, 1985.
42. Wennberg, J. The paradox of appropriate care. *JAMA*, 1987, 258, 2568–69.
43. Winslow, C. M., Solomon, D. H., Chassin, M. R., et al. The appropriateness of carotid endarterectomy. *New England Journal of Medicine*, 1988, 318, 721–27.