Quality in Health Care 1996;5:20–30

Primary hip and knee replacement surgery: Ontario criteria for case selection and surgical priority

C David Naylor, J Ivan Williams, and the Ontario panel on hip and knee arthroplasty

Abstract

***Objectives*—To develop, from simple clinical factors, criteria to identify appropriate patients for referral to a surgeon for consideration for arthroplasty, and to rank them in the queue once surgery is agreed.**

***Design*—Delphi process, with a panel including orthopaedic surgeons, rheumatologists, general practitioners, epidemiologists, and physiotherapists, who rated 120 case scenarios for appropriateness and 42 for waiting list priority. Scenarios incorporated combinations of relevant clinical factors. It was assumed that queues should be organised not simply by chronology but by clinical and social impact of delayed surgery. The panel focused on information obtained from clinical histories, to ensure the utility of the guidelines in practice. Relevant high quality research evidence was limited.**

***Setting*—Ontario, Canada.**

***Main measures*—Appropriateness ratings on a 7-point scale, and urgency rankings on a 4-point scale keyed to specific waiting times.**

***Results*—Despite incomplete evidence panellists agreed on ratings in 92·5% of appropriateness and 73·8% of urgency scenarios versus 15% and 18% agreement expected by chance, respectively. Statistically validated algorithms in decision tree form, which should permit rapid estimation of urgency or appropriateness in practice, were compiled by recursive partitioning. Rating patterns and algorithms were also used to make brief written guidelines on how clinical factors affect appropriateness and urgency of surgery. A summary score was provided for each case scenario; scenarios could then be matched to chart audit results, with scoring for quality management.**

***Conclusions*—These algorithms and criteria can be used by managers or practitioners to assess appropriateness of referral for hip or knee replacement and relative rankings of patients in the queue for surgery.**

(*Quality in Health Care* 1996;5:20–30)

Keywords: hip and knee replacement surgery, case selection, surgical priority.

Departments of Medicine and Surgery, University of Toronto and Institute for Clinical Evaluative Sciences in Ontario
C D Naylor, associate professor and chief executive officer, ICES

Department of Family and Community Medicine, University of Toronto and the Institute for Clinical Evaluative Sciences in Ontario
J I Williams, professor and deputy research director, ICES

Members of the Ontario panel on hip and knee arthroplasty:
Dr J Schatzker (clinical chair), Dr T Axelrod, Dr E Badley, Dr M Bell, Dr A E M Borgiel, Dr R B Bourne, Dr A Fam, Ms E C Finch, Dr A Laupacis, Dr W Palmer, and Dr S Parker.

Correspondence to:
Dr C David Naylor, ICES Sunnybrook HSC, G-106, 2075 Bayview Avenue, North York, Ontario M4N 3M5 Canada

Accepted for publication 1 September 1995

Introduction

Lower limb arthroplasty – replacement of joints with prosthetic devices – can yield dramatic improvements in functional status of patients and their health related quality of life.[1–6] Although arthroplasty is increasing in industrialised countries as the technology diffuses and populations age, surgical rates continue to vary moderately across regions of Canada[7 8] and the United States.[9 10] In the United Kingdom variations in surgical rates for hip replacements have diminished, leading to optimism that need is being met,[11 12] but it is uncertain whether similar shifts have occurred in the United Kingdom for knee arthroplasty – a point of concern as use of knee replacements is more variable and rising faster than hip surgery in North America.[7–10]

Variations in surgery rates raise questions about potential underuse or overuse of procedures. Among the causes of variations are administrative decisions about allocation of resources between and within hospitals, and clinical decisions by referring physicians and surgeons; Canadian[13] and American surveys[14] both showed areas of disagreement when orthopaedic surgeons were asked about indications for knee replacement. Disagreement between doctors was larger still when Canadian general practitioners were surveyed (J G Wright, Knee Patient Outcomes Research Team co-investigator, personal communication), and data in the United Kingdom confirm that the threshold for surgical referral of orthopaedic patients varies among general practitioners.[15] Thus, guidelines for appropriate referral, using simple clinical features, could be valuable to primary care physicians and improve care for patients who might otherwise not receive timely or appropriate referral. Our first objective in this paper is to report on the development and content of such guidelines in Ontario, Canada.

Apart from variations in surgical rates waiting lists constitute another health systems issue in joint replacement surgery. Lengthy queues have been reported in some Canadian provinces,[16] the United Kingdom,[17] and New Zealand.[18] Surveys of Canadian patients suggest that waiting times of three to four months are well tolerated, but dissatisfaction mounts with increasing delays.[19 20] In Canada, and doubtless in other nations, waiting times within regions vary among hospitals and surgeons, and surgeons do not necessarily bring patients with more severe symptoms or disabilities to

the front of the queue.[19–21] Our second objective is to report on the development of a simple scoring system to help surgeons and clinical managers in assigning priority to patients accepted for primary joint replacement. These relative urgency ratings can be used to help to order patients more equitably within queues, independently of typical waiting times, which will tend to vary with local resources.

The guidelines presented have two components: a brief written synopsis of suggestions for case selection or queue management and diagramatic algorithms for quick clinical reference. We have also appended appropriateness scores and urgency scores for individual case scenarios; these can be matched to chart audit data for quality management purposes.

Methods

RATIONALE AND METHOD FOR PANEL PROCESS

The RAND delphi method relies on appropriateness ratings of abbreviated case scenarios, or "indications", by a multidisciplinary panel of experts.[22 23] We previously extended these methods to queuing criteria for cardiac surgery.[24 25] Although such panel methods do not explicitly link the quality of evidence to the ratings for a given indication,[26–28] they are a useful compromise because high quality evidence is lacking on the full range of clinical indications for many procedures,[29] including hip and knee arthroplasty. For example, a computerised and manual literature search carried out before this study (updated for this report) yielded no randomised trials comparing arthroplasty with continued or intensified medical treatment and no detailed studies of patients' functional status in arthroplasty queues. A panel process was the logical recourse to synthesise available observational evidence with clinical experience and judgements.

PANEL MEMBERSHIP

The panel consisted of four consultant orthopaedic surgeons, two consultant rheumatologists, two general practitioners, a general physician with healthcare research interests, an epidemiologist interested in musculoskeletal disease, and a physiotherapist active in practice related to orthopaedics. Panellists were chosen by the research team based on *(a)* credibility as confirmed by discussions with academic orthopaedists and representatives of organised medicine, *(b)* relevant research interests, and *(c)* representation from multiple disciplines, localities, and practice circumstances.

PANEL PROCESS: GENERATION OF CASE SCENARIOS

In order to generate case scenarios the panel first met to discuss available evidence and its limitations. Panellists were given structured abstracts of research reports pertaining to clinical outcomes of hip and knee arthroplasty (list available to readers on request). We included case series if patients were assessed with explicit and consistent methods for before and after comparisons, and if exclusion criteria were documented along with numbers of patients excluded. Benefits for most patients were seen for pain, stiffness, functional capacity, and quality of life.[1–6] These findings were supported by a recent meta-analysis of outcomes of knee arthroplasty showing that over 70% of patients have dramatic functional benefits and over 80% achieve pain relief.[30] However, there were (and remain) no multivariate analyses to differentiate patients after arthroplasty according to short term and longer term benefit risk ratios. Even single factor analyses have been inconclusive – for example, for obesity as a predictor of adverse outcomes.[31–35]

The panel next met to assess factors that affect surgical referral or timing of surgery, or both. The lists of factors had to be clinically credible but brief enough for use in practice. Pain and dysfunction were agreed as the common determinants of surgical referrals. Data from case series (references on request) and a comparative study[36] confirmed that diverse functional scales are responsive to surgical intervention, but data on inter-rater reliability data were lacking except for one little used measure.[37] The popular Charnley modification[38] of the Merle D'Aubigne-Postel scale[39] focuses on ambulation; other joint specific scales assess both walking and other activities of daily living.[40–42] The panel concluded that overall dysfunction might be better captured with a general classification scheme. Again, information was lacking on measurement properties of general scales in routine clinical use.[43] Data supported the validity and reliability of the Western Ontario McMaster Arthritis Index, a disease specific measure for patients with osteoarthritis,[44 45] and the generic SF 36 health status questionnaire.[46–48] These, however, are primarily research tools. For simplicity, the panel returned to the original functional classification of the American College of Rheumatology (see box 1)[49]; the revised criteria[50] were set aside, as they contain less explicit clinical descriptions. For pain on activity and at rest, panellists combined elements from the existing scales to generate descriptions that had face validity for clinical practice. Boxes 1 and 2 show the classification systems used for functional capacity and pain levels, respectively.

For the appropriateness criteria to be useful in surgical audits, the orthopaedic surgeons

Class I	Complete functional capacity with ability to carry on all usual duties without handicaps
Class II	Functional capacity adequate to conduct normal activities despite handicap of discomfort or limited mobility of one or more joints
Class III	Functional capacity adequate to perform only few or none of the duties of usual occupation or of self care
Class IV	Largely or wholly incapacitated with patient bedridden or confined to wheelchair, permitting little or no self care

Box 1 Definitions of functional capacity: original criteria of the American College of Rheumatology[49]

and rheumatologists on the panel suggested considering two other factors. Firstly, reparative procedures such as osteotomy might mitigate need for arthroplasty in younger patients. Indications for tibial osteotomy are still debated,[51-55] and the panel accepted that a decision on whether to carry this out would be made primarily by rheumatologists or orthopaedic surgeons, based on radiological findings and detailed examination. Secondly, the probability of 10 year survival of the prosthesis also bears on the decision to replace a hip or knee joint, particularly in younger people. No validated scoring system for prosthesis survival was located. Among the independent factors identified in the few studies with formal multivariate methods or stratified analyses[56-59] were body weight, rheumatoid arthritis, avascular necrosis of the hip, and operation after traumatic joint disruption. Panellists were again forced to draw on evidence of limited quality, combined with clinical judgement, in grouping risk factors for high, medium, and low prosthesis survival (box 3).

For urgency rating, the panellists sought factors that could be applied by orthopaedic surgeons unaccustomed to ordering queues by severity of disease. The pain and function classifications already mentioned were used again, but panellists highlighted rest pain as a crucial determinant of queue priority. Hence, rest pain was added as a separate factor along with pain on activity, which was defined from activity related elements within general pain (box 2). Problems in work or caregiving were considered separately to capture the potential impact on others from a given level of functional disability. As stiffness without concomitant pain may occasionally impair functional status, the panel was asked to rate some scenarios that were clinically unusual – for example, a patient in functional class III, without notable pain on activities of daily living.

General pain levels:

Mild Pain severity and frequency interferes minimally on an *intermittent* basis with usual daily activities but limits vigorous activities, including *continuous climbing of 15 stairs quickly*. No associated sleep disturbance because of pain. Pain controlled by one or more of the following: non-steroidal anti-inflammatory and non-narcotic analgesic drugs with a tolerable burden of side effects; activity modification; physical modalities such as heat, or cold, or transcutaneous electrical nerve stimulation; strengthening or endurance exercises

Moderate Pain occurs *daily* with movement and interferes significantly with some, but not all, daily activities. More vigorous activities are not done. Occasional sleep disturbance because of pain. Narcotic analgesic drugs are frequently needed for pain control with an ongoing burden of drug side effects. An aid such as a cane is required for prolonged walking

Severe Pain is *constant* and interferes significantly with most daily activities. Pain is not adequately controlled by a combination of a non-steroidal anti-inflammatory and narcotic analgesic drugs; there is a continuing major burden of drug side effects or pain symptoms, or both. Other methods of pain management have not been successful in limiting the burden of pain. Aids are required for mobility (such as a cane, walker or wheelchair, or both)

Specific pain levels at rest:

Mild Constant pain even when not moving, but usually relieved by oral analgesic drugs such as aspirin or acetaminophen, liniments, heat or cold, or both)

Moderate Constant pain even when not moving, usually relieved with narcotic analgesic drugs, such as codeine

Severe Constant, unremitting pain even when not moving, which interferes with sleep and is inconsistently relieved by narcotic analgesic drugs. May require narcotic drugs by injection for relief

Box 2 Definitions of levels of pain

SIMPLIFYING ASSUMPTIONS

It was assumed that the surgeon would consider perioperative risk from medical conditions, threats to limb viability from vascular disease, other potential causes of reduced short term joint survival (active osteomyelitis, acute joint sepsis, severe osteoporosis, and previous trauma with gross distortion of bony architecture), and other conditions that precluded much functional improvement, such as cardiorespiratory or neurological disease that limit activity.

Functional classes II–III:

- Age <45, 45–60, >60
- Pain in general – mild, moderate, severe (class III only)
- Chance of 10 year survival of the prosthesis* – high, medium, low
- For patients <60 years, is patient an osteotomy candidate? – yes, no

Functional class IV:

- Age <45, 45–60, >60
- Pain at rest – mild, moderate, severe
- Chance of 10 year survival of the prosthesis* – high, medium, low
- Expectation of improvement of at least one functional class – yes, no

*Assumed to be affected by obesity, type and level of activity, and nature of the underlying joint problem. Risk factors for medium 10 year prosthesis survival: vigorous activity independent of age, abnormal anatomy as in congenital dislocation of the hip, multijoint involvement as in rheumatoid arthritis, or bone disorders – for example, Paget's disease. Risk factors for low 10 year prosthesis survival: takedown of a previous fusion, a history of previous sepsis in or about the hip or knee, abductor weakness in hip replacement or problem with the extensor mechanisms in a total knee replacement, or dementia. As in the text, these represented panel judgements; evidence from formal multivariate analyses was lacking.

Box 3 Factor lists for appropriateness scenarios

SCENARIO RATINGS

Appropriateness and urgency scenarios were organised into chapters by functional class. Wholly implausible combinations of clinical factors were eliminated. No attempt was made to match the relative numbers of scenarios to their expected frequency in practice, as both common and uncommon clinical presentations had to be covered. Booklets containing the scenarios were posted to the expert panel with instructions and factor definitions. The appropriateness scale ranged from 1 to 9, where 1 was clearly inappropriate, 9 highly appropriate, and 5 a "toss up." As in the RAND method,[22 23] "appropriate" was defined to mean that the expected health benefit (for example, relief of pain, improved functional capacity, better quality of life) exceeded the expected negative consequences (for example, mortality, morbidity, anxiety of anticipating the procedure, pain produced by the procedure, time lost from work) by a sufficient margin that the procedure was worth doing, exclusive of monetary cost. Panellists were advised that ratings would be grouped as follows: 1, 2, and 3 as inappropriate; 4, 5, and 6 as uncertain; and 7, 8, and 9 as appropriate. The ordinal urgency ratings from 1 (highest priority) to 4 (lowest priority) were related to time frames within which surgery should be performed, as follows: ≤1 month; 1–3 months; 3–6 months; and 6–12 months. Panellists understood that ordinal ratings could be used to set relative queue priorities, apart from suggested waiting times. The panellists rated the scenarios and posted back their booklets.

FINAL PANEL MEETING

At the final panel meeting in 1993, each panellist's copy of the scenarios showed his or her personal ratings, along with anonymous ratings of other panellists. Areas of disagreement and ambiguity were discussed and factors finalised (boxes 3 for appropriateness and 4 for urgency). A last set of ratings was then made independently by each panellist.

ANALYSIS OF THE RATINGS

For the appropriateness scale, levels of risk and benefit are not specified for each point, so that the interval from 8 to 9 may not be the same as from 7 to 8. Thus, rather than relying on means and standard deviations of raw ratings,

Functional class II:
- Rest pain – absent, mild
- Pain on activities of daily living – mild, moderate
- Problem in work or care giving – none, minor

Functional class III:
- Rest pain – absent, mild, moderate, severe
- Pain on activities of daily living – absent, mild, moderate, severe
- Major problem in work or care giving – yes, no

Functional class IV:
- Rest pain – absent, mild, moderate, severe
- Improvement in functional status expected – yes, no

Box 4 Factor lists used in urgency scenarios

1 = Inappropriate with strong agreement (eight or more of 11 panellists rated the scenario in the 1–3 range)

2 = Inappropriate with majority agreement (six scores or more are in the inappropriate range and three of the remaining five are in the uncertain range – that is, 4–6)

3 = Uncertain, possibly inappropriate (no more than one score in the appropriate range, and the remaining 10 scores are split between uncertain and inappropriate on a 5:5, 6:4, or 4:6 division)

4 = Uncertain with agreement (eight or more of 11 panellists rated the scenario in the 4–6 range), OR scatter as shown by failure to fit one of the other prespecified patterns of convergence

5 = Uncertain, possibly appropriate (no more than one score in the inappropriate range, and the remaining 10 scores split between uncertain and appropriate on a 5:5, 6:4, or 4:6 division)

6 = Appropriate with majority agreement (mirror image of 2)

7 = Appropriate with strong agreement (mirror image of 1)

Box 5 Method for generating panel summary scores, based on convergence of appropriateness ratings

summary scores were developed for each scenario that reflect the extent of agreement within one of the three prespecified appropriateness categories. Box 5 shows decision rules for these summary scores ; tables 1A–3A in the appendix provide a ratings profile and score for each scenario.

For the urgency scale, the relative priority given to patients with different characteristics is germane for reasons given earlier. Decision rules were again applied to the urgency ratings, generating summary scores for agreement on level of priority (box 6). Table 4A in the appendix gives individual scenario ratings and scores.

Two general measures of agreement among panellists were adapted from the RAND method[22 23] – namely, how often at least 90% of ratings (≥10 of 11 panellists) converged in two adjacent appropriateness categories or two adjacent urgency levels and how often at least 80% of ratings (≥9 of 11 panellists) converged in a single category or level. Agreement

1 = Seven or more ratings in 6–12 months' waiting time category

2 = Five or six ratings in 6–12 months' waiting time category

3 = Seven or more ratings in 3–6 months' waiting time category

4 = Five or six ratings in 3–6 months' waiting time category, and most others in the 1–3 month category

5 = Seven or more ratings in the 1–3 months' category

6 = Five or six ratings in the 1–3 months' category and most others in the <1 month category

7 = Seven or more ratings in the <1 month category

Box 6 Method for generating panel summary scores, based on convergence of urgency ratings

expected by chance was determined from the binomial theorem and confirmed by computer simulations.

Rating patterns were analysed with recursive partitioning software (Knowledge Seeker, Angoss Software, Toronto, Ontario, Canada). Recursive partitioning is a form of stepwise tree based regression as discussed in detail elsewhere[60–62] and we used it to compare the means and distributions of summary scores for sets of scenarios partitioned by different factors. Recursive partitioning sequentially chooses the partitioning factor yielding the largest F value, derived from the variance explained by creation of two or more new branches. For any factor with more than two levels, the program may combine levels with statistically similar scores. Partitioning occurs only if a specified significance level is met, analogous to the α (or P value) threshold before adding another variable to a stepwise regression model. To prevent the creation of too many branches in the tree this threshold was set at 0·01 for the appropriateness analysis. For the urgency analysis, with only 42 scenarios or scores to analyse, a threshold of 0·05 was used.

Recursive partitioning imposes a second significance check when comparing two or more branches from the same limb; differences between branches in mean summary scores have to attain the equivalent of two tailed significance. Adjustment for multiple comparisons is automated in the software. The figures show the exact P values for differences between branches.

Because the numbers of scenarios and factor distributions are arbitrary, clinical logic might not match the statistical algorithm generated by recursive partitioning. Accordingly, we explored the impact of "forcing" particular factors into both the urgency and appropriateness algorithms to create more clinically homogeneous tiers in the tree. The r^2 values for the statistically optimal models were compared with r^2 values for the "forced" models.

Lastly, we used the patterns of the ratings and associated statistical information to generate a set of brief narrative guides about appropriateness and urgency.

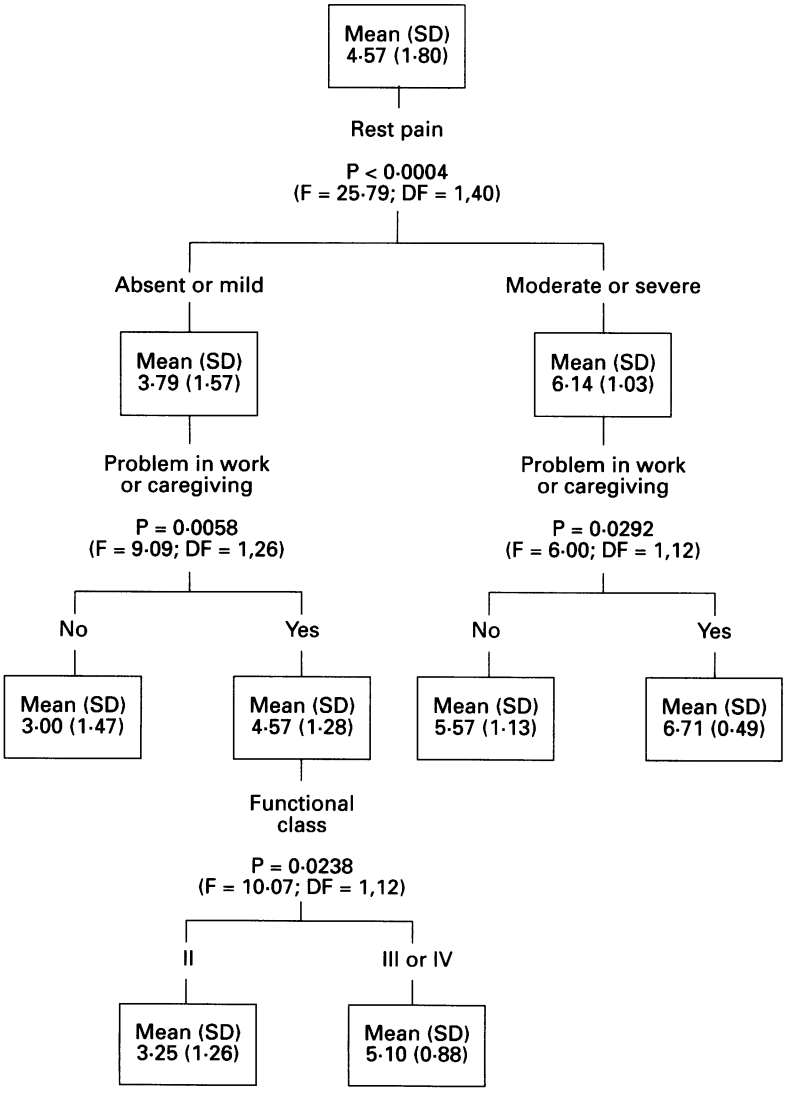

Fig 1 Algorithm of queue priority. Higher scores mean higher priority. For simplicity of use, 2·00 can be subtracted from all average scores shown here, and the score rounded to the nearest decimal place, thereby creating a scoring algorithm that runs from 1·0 to 4·7. P values, corresponding F statistics, and related degrees of freedom all reflect the additional variance explained by partitioning on the factor as shown. This is analogous to a stepwise regression mode.

Results

AGREEMENT

Convergence of ≥9/11 panellists within one of three prespecified appropriateness bands occurred in 60·8% of scenarios versus 4·1% expected by chance alone. Similar convergence within one of four urgency levels occurred in 16·7% of scenarios versus 0·5% expected by chance. Convergence of ≥10/11 panellists within two adjacent appropriateness bands occurred in 92·5% of appropriateness scenarios versus 15% expected by chance alone, and in 73·8% of urgency scenarios versus 1·8% expected by chance alone.

APPROPRIATENESS

Extremely vigorous activity is contraindicated after joint replacement, prosthesis failure rates rise after 10 years, and reoperation is technically difficult. Hence, particularly in people under 60 years of age without considerably impaired function, it is prudent to defer surgery or consider alternative procedures such as high tibial osteotomy or femoral osteotomy whenever feasible.

Functional class II – When pain is mild or osteotomy is an option, joint replacement is deemed inappropriate for patients in functional class II. Even among those who cannot have an osteotomy and who have moderate pain, the appropriateness of joint replacement is uncertain and highly dependent on case specific judgements, unless patients are older and have a good chance of long term prosthesis survival.

Functional class III – Again, osteotomy is deemed preferable to arthroplasty whenever possible in patients younger than 60. For those not able to undergo osteotomy, the need for pain relief must be weighed against chances of long term prosthesis survival. If pain is mild, surgery should be viewed cautiously unless there is a very good chance of long term prosthesis survival. Among older patients (>60), moderate and severe pain are strong

indications for joint replacement when coupled with impaired activities of daily living as is usual in functional class III.

Functional class IV – Patients in this functional class are bedridden or confined to a wheelchair, or both, hence pain on activity is not a factor. Patients with severe rest pain are potentially appropriate regardless of other factors, as joint replacement may be the only option to relieve pain. Surgery is also appropriate if there is some expectation of improvement in function. If, however, the pain level is mild to moderate, and surgery is being undertaken with little expectation of functional improvement, careful weighing of risks and benefits is needed.

URGENCY

Long waiting times for joint replacement are obviously undesirable, and these rankings simply represent an attempt to limit and distribute fairly the burden on patients caused by resource constraints. Although functional class II implies nearly normal functional status (see box 1), patients may fall between functional classes II and III for minor interference with work or caregiving. This has been considered in urgency rankings for people in functional class II.

Functional class II – Patients with mild pain on activity and no rest pain received the lowest priority (summary scores of 1 to 2), and for most candidates surgery was inappropriate. With moderate pain during activity (scores of 2–3) higher priority was assigned. Higher priority still was given if rest pain was present and work or caregiving may be impeded (scores of 4–5).

Functional class III – Other things being equal, urgency was generally moved up one level for patients with functional class III compared with those with functional class II. Patients with severe pain on activity had scores typically of 5 or 6, depending whether accompanying rest pain was absent or mild. With severe pain on activity and at rest, the panel assigned high priority (scores of 7), with surgery to be provided as soon as possible.

Functional class IV – Many patients in functional class IV have longstanding and severe arthritis affecting multiple joints and therefore

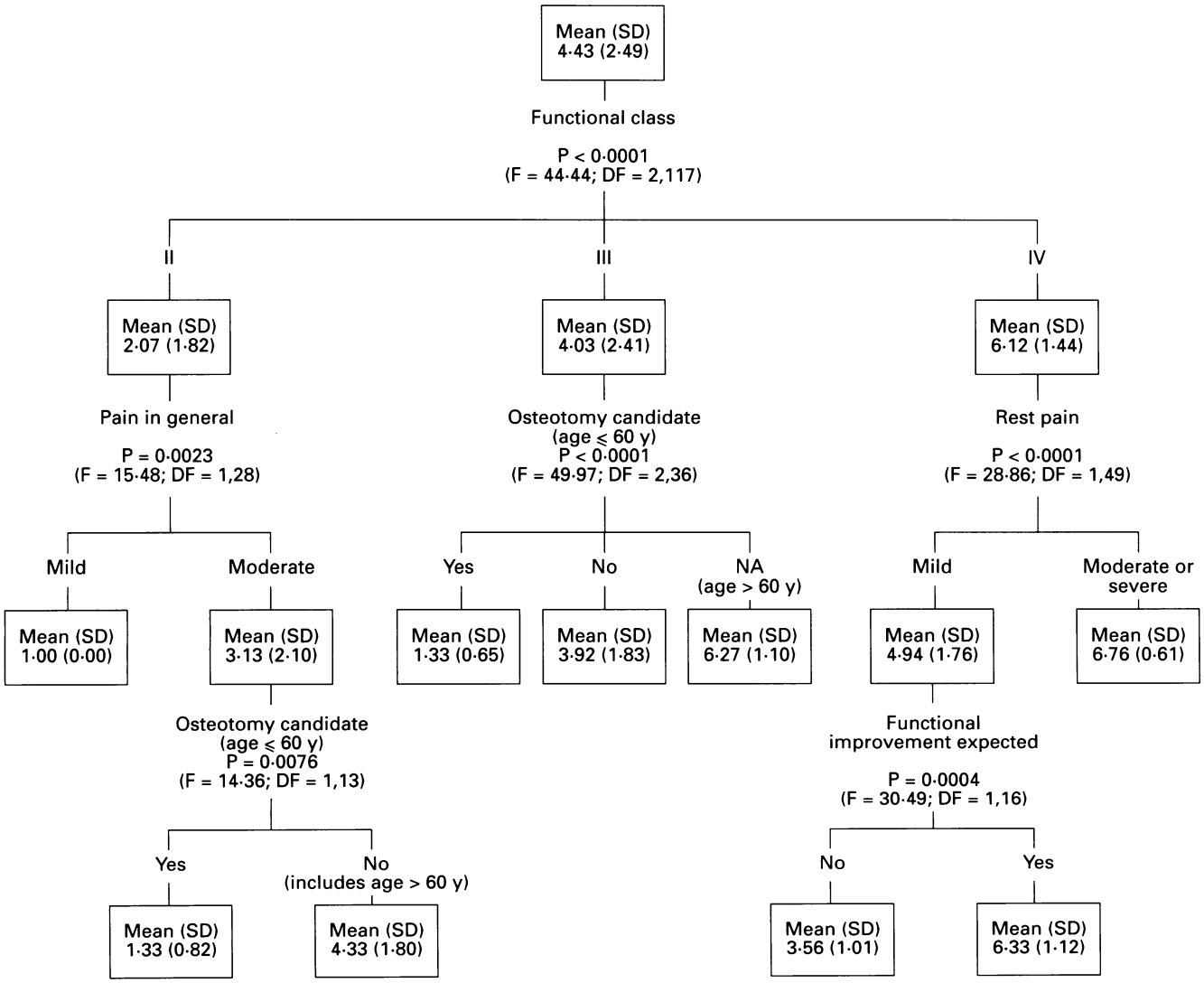

Fig 2 Algorithm of referral and case selection. Higher scores correspond to higher potential appropriateness for surgery. P values, corresponding F statistics, and related degrees of freedom all reflect the additional variance explained by partitioning on the factor as shown. This is analogous to a stepwise regression model.

have limited prospects for improvement of functional status. Surgery should none the less be provided expeditiously in patients with moderate to severe rest pain (scores 5 and 7 respectively). A few patients with moderate rest pain may only recently have become confined to a wheelchair or bed and have good prospects of walking again. Delay may reduce their chances of rehabilitation, and they were given high priority (scores of 7).

ALGORITHMS BASED ON RECURSIVE PARTITIONING

Figures 1 and 2 are the algorithms for appropriateness of referral and relative queue priority. Scanning the summary scores at each step of the analytical tree shows how diverse factors drive categorisations of surgical need and priority.

The appropriateness algorithm explains 84% of the variance in scenario ratings (r^2 0·84). To simplify the algorithm, we tried substituting pain for osteotomy candidacy under functional class III so that pain followed functional class across the entire second tier of the tree; but the r^2 fell to 0·70. For the urgency algorithm, attempts to force in functional status as the first partitioning variable reduced the r^2 from 0·63 to 0·59 and created a more complicated tree.

Discussion

By drawing on a Delphi panel process and keying on simple clinical factors, this report provides materials to help in appropriate and equitable use of hip and knee replacement surgery. The need for referral guidelines was supported by evidence for variations in judgements of general practitioners about appropriateness of hip and knee arthroplasty, suggesting that underuse of hip and knee arthroplasty in regions with low rates of these procedures may arise from resource constraints that narrow the surgical "spout", along with patterns of referral in the primary care "funnel". The need for waiting list guidelines was supported by evidence that patients are not given priority according to severity of symptoms, even though their acceptance of waiting times is affected by the degree of delay and the severity of their pain or disability. The algorithms in the figures may serve as tools to support clinical decisions, or as adjuncts for audits of practice patterns. The lists of scenarios and related appropriateness or urgency scores may also be useful in audits.

The panel exercise was hampered by a lack of relevant research evidence. Orthopaedic surgeons have been "making do without randomised trials"[63] of case selection for lower limb joint replacement; trials instead take surgery as indicated, and randomise patients by type of prosthesis or surgical method. Moreover, a trial comparing outcomes of patients randomised to different waiting periods would be unethical, unless longer waiting periods were accompanied by interventions to forestall surgery or to improve its outcomes. A multidisciplinary panel process, with modified Delphi methods, was accordingly our best option for generating guidelines.

The Delphi method[22 23] draws on evidence, inference, and experience, and permits guidelines or utilisation review criteria to be generated when detailed evidence is not available to allow consideration of all the indications for a procedure. However, this means that "panellists may be pooling ignorance as much as distilling wisdom"[64] in clinical situations where, as was true here, the evidence is very limited. The reproducibility of ratings based on weak evidence may be poor. For example, the specialties and nationalities of panellists have been shown to affect ratings for coronary artery bypass surgery, where a much larger body of evidence is available.[29]

Our methodology contrasts with the informal consensus approach of a 1992 New Zealand panel[65] which aimed to set "boundary" guidelines and eschewed more detailed algorithms or "pathway" guidelines of the type generated here. In laying out some general guidance on appropriate case selection, the convenors of the New Zealand panel cautioned that "there was some difficulty in getting agreed guidelines, particularly because there appears to be no general agreement anywhere, nationally or internationally."[65] However, we achieved far greater agreement between panellists than would be expected by chance alone, and the summary scores take explicit account of panel agreement and disagreement on ratings for particular subgroups of case scenarios.

For the queuing algorithm, the New Zealand panel recommended that surgery in less than six weeks should be reserved for those "suffering constant, unrelieved pain, especially at night and where there is a major threat to independence or occupation."[65] The criteria and associated waiting times are similar to our guidelines and algorithm. Otherwise, the New Zealand panel simply recommended that "patients whose pain is alleviated by rest and analgesics and other conservative measures, but where there is interference with the activities of daily living, should be operated on within six months." The detailed Ontario guidelines should help in structuring more equitable queues.

Clinical limits to the criteria should be acknowledged. For the appropriateness criteria, our target audience in the first instance was the referring general practitioner. We therefore did not deal with severity of disease as shown radiologically, or degree of joint deformity on clinical examination, both of which will enter into the orthopaedic surgeon's decision making. Although candidacy for osteotomy was added as a factor, it would not generally enter into a general practitioner's decisions about referral and was considered secondary to pain and functional status in the algorithms for assessing appropriateness of referral. Technical complexities affecting decisions about repeat arthroplasties are not considered here. Patients differ in their pain thresholds, their tolerance of non-steroidal antiinflammatory drugs or narcotic analgesic drugs, and their thresholds for bringing symptoms to the attention of a physician or surgeon. However, there is no

"gold standard" for objectively determining patients' levels of distress. A further limitation of the exercise is the assumption that patients had been prescreened and found to have an acceptable risk of perioperative complications owing to comorbid conditions. General practitioners, surgeons, or anaesthetists will doubtless exclude patients with severe cardiorespiratory or neurological conditions from consideration for surgery. It is none the less possible that practitioners (and expert panellists) may differ as to what risks are indeed acceptable, and the likelihoods of complications in association with various medical conditions. Thus, whether these materials are used as explicit criteria for the purposes of orthopaedic audit or as practice guidelines for referring physicians, it would be reasonable to gather adjunctive data or develop simple criteria about comorbidity and perioperative risk or both. Lastly, the time frames in the Ontario urgency ratings depended on resource constraints in our jurisdiction. However, the criteria and scoring system were designed to help in ranking patients, independently of specific associated waiting time.

To use these criteria for audit, an individual case can be either mapped on to the nearest matching scenario listed in the appendix or the patient's likely categorisation tracked by following figures 1 or 2. Take, for example, a 70 year old person who presents with impaired daily function and moderate pain on activities of daily living related to osteoarthritis of the hip. This person is in functional class III. The appendix, shows that the category will be "appropriate with strong agreement," unless the chances of prosthesis survival are low, in which case the overall category becomes "appropriate with majority agreement." The same patient can also be tracked down figure 1 within the functional class III arm: the age bracket places the patient under the "not applicable" arm for osteotomy, leading to a mean summary score of 6.27. Similar approaches can be taken for assessing urgency of joint replacement as well, with the starting point at functional class (table 4A, appendix) or degree of rest pain (fig 2).

Results of continuing work in North America by the Patient Outcomes Research Teams on hip and knee arthroplasty should permit refinement of these guidelines. In the interim, research is under way to confirm the feasibility of applying the existing criteria in chart audits, to compare appropriateness as screened with the explicit criteria to the implicit judgements of rheumatologists and surgeons, and to field test the guidelines for day to day use by general practitioners and in surgical practices.

1 Cleary PD, Greenfield S, McNeil BJ. Assessing quality of life after surgery. *Controlled Clin Trials* 1991;**12**: 189–203S.

2 Kantz ME, Harris WJ, Levitsky K, Ware JE Jr, Davies AR. Methods for assessing condition-specific and generic functional status outcomes after total knee replacement. *Med Care* 1992;**30**:MS240–52.

3 Laupacis A, Bourne R, Rorabeck C, Feeny D, Wong C, Tugwell P, *et al.* The effect of elective total hip replacement on health-related quality of life. *J Bone Joint Surg* 1993;**75A**:1619–26.

4 Liang MH, Cullen KE, Larson MG, Thompson MS, Schwartz JA, Fossel AH, *et al.* Cost-effectiveness of total joint arthroplasty in osteoarthritis. *Arthritis Rheum* 1986; **29**:937–43.

5 Jonsson H, Larsson BJ. Functional improvement and cost of hip and knee arthroplasty in destructive rheumatoid arthritis. *Scand J Rheum* 1991;**20**:351–7.

6 Katz J, Larson M, Phillips C, Fossel A, Liang M. Comparative measurement sensitivity of short and longer health status instruments. *Med Care* 1992;**30**:917–25.

7 Cohen MM, de Boer D, Young W. Total hip replacement. In: Naylor CD, Anderson GM, Goel V, eds. *Patterns of health care in Ontario.* Ottawa: Canadian Medical Association, 1994:72–6.

8 Cohen MM, de Boer D, Young W. Total knee replacement. In: Naylor CD, Anderson GM, Goel V, eds. *Patterns of health care in Ontario.* Ottawa: Canadian Medical Association, 1994:76–9.

9 Wennberg JE, Freeman JL, Culp WJ. Are hospital services rationed in New Haven or overutilised in Boston? *Lancet* 1987;i:1185–9.

10 Peterson MG, Hollenberg JP, Szatrowski TP, Johanson NA, Mancuso CA, Charlson ME. Geographic variations in the rates of elective total hip and knee arthroplasties among Medicare beneficiaries in the United States. *J Bone Joint Surg* 1992;**74A**:1530–9.

11 Rajaratnam G, Black NA, Dalziel M. Total hip replacements in the National Health Service: is need being met? *J Public Health Med* 1990;**12**:56–9.

12 Williams MH, Newton JN, Frankel SJ, Braddon F, Barclay E, JAM Gray. Prevalence of total hip replacement: how much demand has been met? *J Epidemiol Community Health* 1994;**48**:188–91.

13 Wright JG, Coyte PC, Hawker G, Bombardier C, Cooke D, Heck D, *et al.* Variation in orthopedic surgeons' perceptions of the indications for and outcomes of knee replacement. *Can Med Assoc J* 1995;**152**:687–97.

14 Tierney WM, Fitzgerald JF, Heck DA, Kennedy JM, Katz BP, Melfi CA, *et al.* Tricompartmental knee replacement: a comparison of orthopedic surgeons' self-reported performance rates with surgical indications, contraindications, and expected outcomes. *Clinical Orthopedics* 1994;**305**:209–17.

15 Lowry RJ, Donaldson LJ, Gregg PJ. Variations in clinical decisions: a study of orthopaedic patients. *Public Health* 1991;**105**:351–5.

16 Ramsay C, Walker M. *Waiting your turn: hospital waiting lists in Canada.* 4th ed. Vancouver: Fraser Institute, 1994.

17 Rigge M. *Guide to hospital waiting lists. 6th ed.* London: College of Health, 1990.

18 Tomlinson M, Cullen J. A clinical audit of patients on an orthopaedic waiting list for greater than two years. *NZ Med J* 1992;**105**:266–8.

19 Coyte PC, Wright JG, Hawker GA, Bombardier C, Dittus RS, Paul JE, *et al.* Waiting times for knee replacement surgery in the United States and Canada. *N Engl J Med* 1994;**331**:1068–71.

20 Ho E, Coyte PC, Bombardier C, Hawker G, Wright JG. Ontario patients' acceptance of waiting times for knee replacements. *J Rheumatol* 1994;**21**:2101–5.

21 Williams JI, Llewellyn-Thomas HA, Arshinoff R, Young N, Naylor CD. *The burden of waiting for hip and knee replacements in Ontario.* Toronto: ICES, 1995: 1–24. (working paper series No 34.)

22 Brook RH, Chassin MR, Fink A, Solomon DH, Kosecoff JB, Park RE. A method for the detailed assessment of the appropriateness of medical technologies. *Int J Technol Assess Health Care* 1986;**2**:53–64.

23 Park RE, Fink A, Brook RH, Chassin MR, Kahn KL, Merrick NJ, *et al.* Physician ratings of appropriate indications for six medical and surgical procedures. *Am J Public Health* 1986;**76**:766–72.

24 Naylor CD, Baigrie RS, Goldman BS, Basinski A, Revascularisation panel consensus methods group. Assessment of priority for coronary revascularisation procedures. *Lancet* 1990;**335**:1070–3.

25 Naylor CD, Baigrie RS, Goldman BS, Cairns JA, Beanlands DS, Berman N, *et al.* Assigning priority to patients requiring coronary revascularization: consensus principles from a panel of cardiologists and cardiac surgeons. *Can J Cardiol* 1991;7:207–13.

26 Phelps C. The methodologic foundations of stdies of the appropriateness of medical care. *N Engl J Med* 1993; **329**:1241–5.

27 Hicks NR. Some observations on attempts to measure appropriateness of care. *BMJ* 1994;**309**:730–3.

28 Naylor CD, Sibbald WJ, Sprung CL, Pinfold SP, Calvin JE, Cerra FB. Pulmonary artery catheterization: can there be an integrated strategy for guideline development and research promotion? *JAMA* 1993;**269**:2407–11.

29 Naylor CD. Grey zones of clinical practice: some limits to evidence-based medicine. *Lancet* 1995;**345**:840–2.

30 Callahan CM, Drake BG, Heck DA, Dittus R. Patient outcomes following tricompartmental total knee replacement – a meta-analysis. *JAMA* 1994;**271**:1349–57.

31 Scott N, Rubinstein M, Scuderi G. Results after knee replacement with a posterior cruciate-substituting prosthesis. *J Bone Joint Surg* 1988;**70A**:1163–73.

32 Ranawat CS, Boachie-Adjei O. Survivorship analysis and results of total condylar knee arthroplasty. *Clin Orthop* 1988;**226**:6–13.

33 Marmor L. Unicompartmental arthroplasty of the knee with a minimum ten-year follow-up period. *Clin Orthop* 1988; **228**:171–7.

34 Kjersgaard-Andersen P, Hvid I, Wethelund JO, *et al.* Total condylar knee arthroplasty in osteoarthritis. *Clin Orthop* 1989;**238**:167–73.
35 Stern S, Insall J. Total knee arthroplasty in obese patients. *J Bone Joint Surg* 1990;**72A**:1400–1.
36 Lagan P, Weiss CA, Hip rating scales: a clinical analysis. *Int Surg* 1981;**66**:331–3.
37 McDowell IW, Martini CJM, Waugh W. A method for self-assessment of disability before and after hip replacement. *BMJ* 1978;ii:857–9.
38 Charnley J. The long-term results of low-friction arthroplasty of the hip performed as a primary intervention. *J Bone Joint Surg* 1972;**54B**:61–76.
39 Merle d'Aubigne R, Postel M. Functional results of hip arthroplasty with acrylic prosthesis. *J Bone Joint Surg* 1954;**36A**:451–75.
40 Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures. An end-result study using a new method of result evaluation. *J Bone Joint Surg* 1969;**51A**: 737–55.
41 Insall JN, Ranawat CS, Aglietti P, Shine J. A comparison of four models of total knee-replacement prostheses. *J Bone Joint Surg* 1976;**58A**:754–65.
42 Aichroth P, Freeman MAR, Smillie IS, Souter WA. A knee function assessment chart. *J Bone Joint Surg* 1978;**60B**: 308–10.
43 Liang MH, Jette AM. Measuring functional ability in chronic arthritis. *Arthritis Rheum* 1981;**24**:80–6.
44 Bellamy N. Pain assessment in osteoarthritis: experience with the WOMAC osteoarthritis index. *Sem Arthr Rheum* 1989;**18**:14–7.
45 Bellamy N, Buchanan WW, Goldsmith C, Campbell J, Stitt L. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988;**15**:1833–40.
46 Ware J, Sherbourne C. The MOS 36-item short-form survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;**30**:473–83.
47 McHorney C, Ware J, Raczek A. The MOS 36-item short-form health survey (SF-36). II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993;**31**:247–63.
48 McHorney C, Ware J, Lu JFR, Sherbourne C. The MOS 36-item short-form health survey (SF-36). III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care* 1994;**32**: 40–66.
49 Steinbrocker O, Traeger CH, Batterman RC. Therapeutic measurement in rheumatoid arthritis. *JAMA* 1949;**140**: 659–62.
50 Hochberg MC, Chang RW, Dwosh I, Lindsey S, Pincus T, Wolfe F. The American College of Rheumatology 1991 revised criteria for the classification of global functional status in rheumatoid arthritis. *Arthritis Rheum* 1992;**35**: 498–502.
51 Broughton NS, Newman JH, Baily RAJ. Unicompartmental replacement and high tibial osteotomy for osteoarthritis of the knee. A comparative study after 5–10 years' follow-up. *J Bone Joint Surg* 1986;**68B**:447–52.
52 Ranawat CS, Padgett DE, Ohashi Y. Total knee arthroplasty for patients younger than 55 years. *Clin Orthop* 1989;**248**:27–33.
53 Coventry MB. Upper tibial osteotomy for gonarthrosis. The evolution of the operation in the last 18 years and long-term results. *Orthop Clin North Am* 1979;**10**:191–210.
54 Coventry MB. Osteotomy about the knee for degenerative and rheumatoid arthritis. Indications, operative technique, and results. *J Bone Joint Surg* 1973;**55A**:23–48.
55 Mont MA, Antonaides S, Krackow KA, Hungerford DS. Total knee arthroplasty after failed high tibial osteotomy. A comparison with a matched group. *Clin Orthop* 1994;**299**:125–30.
56 Surin VV, Sundholm K. Survival of patients and prostheses after total hip arthroplasty. *Clin Orthop* 1983;**177**:148–53.
57 Herberts P, Ahnfelt L, Malchau H, Stromberg C, Andersson GBJ. Multicentre clinical trials and their value in assessing total joint arthroplasty. *Clin Orthop* 1989;**249**: 48–00.
58 Schurman DJ, Bloch DA, Segal MR, Tanner CM. Conventional cemented total hip arthroplasty. Assessment of clinical factors associated with revision for mechanical failure. *Clin Orthop* 1989;**240**:173–80.
59 Sarmiento A, Ebramzadeh E, Gogan W, McKellop H. Total hip arthroplasty with cement. *J Bone Joint Surg* 1990;**72A**: 1470–6.
60 Marshall RJ. Partitioning methods for classification and decision making in medicine. *Stat Med* 1986;5:517–26.
61 Cook EF, Goldman L. Empiric comparison of multivariate analysis techniques: advantages and disadvantages of recursive partitioning analysis. *Journal of Chronic Diseases* 1984;**37**:721–31.
62 Brieman L, Friedman JH, Olshen RA, Stone CJ. *Classification and regression trees.* Belmont, California: Wadsworth, 1984.
63 Dorey F, Grigoris P, Amstutz H. Making do without randomised trials. *J Bone Joint Surg* 1994;**76B**:1–3.
64 Scott E, Black N. When does consensus exist in expert panels [letter]? *J Publ Health Med* 1991;**13**:344.
65 *Hip and Knee Joint Replacements. A consensus development conference report.* Wellington, NZ: National Advisory Committee on Core Health and Disability Support Services, 1992, 10 and 46.

Appendix

Table 1A Appropriateness ratings for scenarios: functional class II

Scenario	*Functional class*	*Pain*	*Prothesis survival*	*Osteotomy candidate*	*Age in years*	*Panelists' scores*			*Total*	*Summary score**
						1–3	*4–6*	*7–9*		
1	II	Mild	High	Yes	<45	11	0	0	11	1
2	II	Mild	High	No	<45	9	2	0	11	1
3	II	Mild	Medium	Yes	<45	11	0	0	11	1
4	II	Mild	Medium	No	<45	10	1	0	11	1
5	II	Mild	Low	Yes	<45	11	0	0	11	1
6	II	Mild	Low	No	<45	10	1	0	11	1
7	II	Moderate	High	Yes	<45	11	0	0	11	1
8	II	Moderate	High	No	<45	3	8	0	11	4
9	II	Moderate	Medium	Yes	<45	11	0	0	11	1
10	II	Moderate	Medium	No	<45	5	6	0	11	3
11	II	Moderate	Low	Yes	<45	11	0	0	11	1
12	II	Moderate	Low	No	<45	9	2	0	11	1
13	II	Mild	High	Yes	45–59	11	0	0	11	1
14	II	Mild	High	No	45–59	9	2	0	11	1
15	II	Mild	Medium	Yes	45–59	11	0	0	11	1
16	II	Mild	Medium	No	45–59	10	1	0	11	1
17	II	Mild	Low	Yes	45–59	11	0	0	11	1
18	II	Mild	Low	No	45–59	11	0	0	11	1
19	II	Moderate	High	Yes	45–59	5	6	0	11	3
20	II	Moderate	High	No	45–59	1	4	6	11	6
21	II	Moderate	Medium	Yes	45–59	8	3	0	11	1
22	II	Moderate	Medium	No	45–59	1	9	1	11	4
23	II	Moderate	Low	Yes	45–59	11	0	0	11	1
24	II	Moderate	Low	No	45–59	3	8	0	11	4
25	II	Mild	High	NA	>60	10	0	1	11	1
26	II	Mild	Medium	NA	>60	10	1	0	11	1
27	II	Mild	Low	NA	>60	10	1	0	11	1
28	II	Moderate	High	NA	>60	0	1	10	11	7
29	II	Moderate	Medium	NA	>60	0	4	7	11	6
30	II	Moderate	Low	NA	>60	1	9	1	11	4

*See box 5 for decision rules regarding summary scores.

Table 2A Appropriateness ratings for scenarios: functional class III

Scenario	Functional class	Pain	Prothesis survival	Osteotomy candidate	Age in years	Panelists' scores			Total	Summary score*
						1–3	4–6	7–9		
31	III	Mild	High	Yes	<45	11	0	0	11	1
32	III	Mild	High	No	<45	2	7	2	11	4
33	III	Mild	Medium	Yes	<45	11	0	0	11	1
34	III	Mild	Medium	No	<45	5	6	0	11	3
35	III	Mild	Low	Yes	<45	11	0	0	11	1
36	III	Mild	Low	No	<45	8	3	0	11	1
37	III	Moderate	High	Yes	<45	10	1	0	11	1
38	III	Moderate	High	No	<45	1	4	6	11	6
39	III	Moderate	Medium	Yes	<45	10	1	0	11	1
40	III	Moderate	Medium	No	<45	2	5	4	11	4
41	III	Moderate	Low	Yes	<45	11	0	0	11	1
42	III	Moderate	Low	No	<45	5	5	1	11	3
43	III	Severe	High	Yes	<45	4	6	1	11	3
44	III	Severe	High	No	<45	1	1	9	11	7
45	III	Severe	Medium	Yes	<45	7	4	0	11	2
46	III	Severe	Medium	No	<45	1	4	6	11	6
47	III	Severe	Low	Yes	<45	8	3	0	11	1
48	III	Severe	Low	No	<45	4	4	3	11	4
49	III	Mild	High	Yes	45–59	6	5	0	11	2
50	III	Mild	High	No	45–59	1	7	3	11	4
51	III	Mild	Medium	Yes	45–59	9	2	0	11	1
52	III	Mild	Medium	No	45–59	2	8	1	11	4
53	III	Mild	Low	Yes	45–59	10	1	0	11	1
54	III	Mild	Low	No	45–59	8	3	0	11	1
55	III	Moderate	High	NA	45–59	0	1	10	11	7
56	III	Moderate	Medium	NA	45–59	0	4	7	11	6
57	III	Moderate	Low	NA	45–59	0	10	1	11	4
58	III	Severe	High	NA	45–59	0	0	11	11	7
59	III	Severe	Medium	NA	45–59	0	0	11	11	7
60	III	Severe	Low	NA	45–59	0	3	8	11	7
61	III	Mild	High	NA	>60	2	2	7	11	6
62	III	Mild	Medium	NA	>60	2	4	5	11	5
63	III	Mild	Low	NA	>60	3	7	1	11	4
64	III	Moderate	High	NA	>60	0	1	10	11	7
65	III	Moderate	Medium	NA	>60	0	1	10	11	7
66	III	Moderate	Low	NA	>60	0	4	7	11	6
67	III	Severe	High	NA	>60	0	0	11	11	7
68	III	Severe	Medium	NA	>60	0	1	10	11	7
69	III	Severe	Low	NA	>60	0	1	10	11	7

Table 3A Appropriateness ratings for scenarios: functional class IV

Scenario	Functional class	Pain	Prothesis survival	Osteotomy candidate	Age in years	Panelists' scores			Summary score*
						1–3	4–6	7–9	
70	IV	Mild	High	Yes	<45	0	3	8	7
71	IV	Mild	High	No	<45	1	7	3	4
72	IV	Mild	Medium	Yes	<45	0	5	6	6
73	IV	Mild	Medium	No	<45	6	2	3	3
74	IV	Mild	Low	Yes	<45	3	4	4	4
75	IV	Mild	Low	No	<45	6	5	0	2
76	IV	Moderate	High	Yes	<45	0	0	11	7
77	IV	Moderate	High	No	<45	0	2	9	7
78	IV	Moderate	Medium	Yes	<45	0	1	10	7
79	IV	Moderate	Medium	No	<45	0	4	7	6
80	IV	Moderate	Low	Yes	<45	0	4	7	6
81	IV	Moderate	Low	No	<45	1	7	3	4
82	IV	Severe	High	Yes	<45	0	0	11	7
83	IV	Severe	High	No	<45	0	0	11	7
84	IV	Severe	Medium	Yes	<45	0	0	11	7
85	IV	Severe	Medium	No	<45	0	2	9	7
86	IV	Severe	Low	Yes	<45	0	1	10	7
87	IV	Severe	Low	No	<45	0	4	7	6
88	IV	Mild	High	Yes	45–59	0	1	10	7
89	IV	Mild	High	No	45–59	0	10	1	4
90	IV	Mild	Medium	Yes	45–59	0	2	9	7
91	IV	Mild	Medium	No	45–59	2	9	0	4
92	IV	Mild	Low	Yes	45–59	1	5	5	5
93	IV	Mild	Low	No	45–59	6	5	0	2
94	IV	Moderate	High	Yes	45–59	0	0	11	7
95	IV	Moderate	High	No	45–59	0	0	11	7
96	IV	Moderate	Medium	Yes	45–59	0	0	11	7
97	IV	Moderate	Medium	No	45–59	0	2	9	7
98	IV	Moderate	Low	Yes	45–59	0	1	10	7
99	IV	Moderate	Low	No	45–59	0	5	6	6
100	IV	Severe	High	Yes	45–59	0	0	11	7
101	IV	Severe	High	No	45–59	0	0	11	7
102	IV	Severe	Medium	Yes	45–59	0	0	11	7
103	IV	Severe	Medium	No	45–59	0	0	11	7
104	IV	Severe	Low	Yes	45–59	0	0	11	7
105	IV	Severe	Low	No	45–59	0	3	8	7
106	IV	Mild	High	Yes	>60	1	1	9	7
107	IV	Mild	High	No	>60	0	6	5	5
108	IV	Mild	Medium	Yes	>60	1	1	9	7
109	IV	Mild	Medium	No	>60	3	4	4	4
100	IV	Mild	Low	Yes	>60	1	2	8	7
111	IV	Mild	Low	No	>60	3	6	2	4
112	IV	Moderate	High	Yes	>60	0	0	11	7
113	IV	Moderate	High	No	>60	0	1	10	7
114	IV	Moderate	Medium	Yes	>60	0	0	11	7
115	IV	Moderate	Medium	No	>60	0	1	10	7
116	IV	Moderate	Low	Yes	>60	0	1	10	7
117	IV	Moderate	Low	No	>60	0	4	7	6
118	IV	Severe	High	NA	>60	0	0	11	7
119	IV	Severe	Medium	NA	>60	0	0	11	7
120	IV	Severe	Low	NA	>60	0	0	11	7

*See box 5 for decision rules regarding summary scores.

Table 4A Urgency ratings for scenarios: functional class II to IV

Scenario	*Functional class*	*Rest pain*	*ADL pain*	*Problem work/care*	*Panelists' scores*				*Summary score**
					1	*2*	*3*	*4*	
1	II	Absent	Mild	No	0	0	2	9	1
2	II	Absent	Mild	Yes	0	0	5	6	2
4	II	Absent	Moderate	No	0	0	6	5	2
5	II	Absent	Moderate	Yes	0	2	9	0	3
7	II	Mild	Mild	No	0	0	5	6	2
8	II	Mild	Mild	Yes	0	1	10	0	3
10	II	Mild	Moderate	No	0	4	6	1	4
11	II	Mild	Moderate	Yes	0	10	1	0	5
13	III	Absent	Absent	Yes	1	4	4	2	4
14	III	Absent	Absent	No	0	0	4	7	1
15	III	Absent	Mild	Yes	1	5	4	1	4
16	III	Absent	Mild	No	0	0	6	5	4
17	III	Absent	Moderate	Yes	2	7	2	0	5
18	III	Absent	Moderate	No	0	1	8	2	3
19	III	Absent	Severe	Yes	5	5	0	1	6
20	III	Absent	Severe	No	1	7	2	1	5
21	III	Mild	Absent	Yes	1	5	4	1	6
22	III	Mild	Absent	No	0	1	5	5	2
23	III	Mild	Mild	Yes	1	5	5	0	4
24	III	Mild	Mild	No	0	1	7	3	4
25	III	Mild	Moderate	Yes	2	9	0	0	5
26	III	Mild	Moderate	No	0	2	8	1	4
27	III	Mild	Severe	Yes	6	5	0	0	6
28	III	Mild	Severe	No	3	7	1	0	5
29	III	Moderate	Mild	Yes	3	6	1	1	6
30	III	Moderate	Mild	No	0	3	6	2	4
31	III	Moderate	Moderate	Yes	5	6	0	0	6
32	III	Moderate	Moderate	No	0	7	4	0	5
33	III	Moderate	Severe	Yes	7	4	0	0	7
34	III	Moderate	Severe	No	5	5	1	0	6
35	III	Severe	Moderate	Yes	7	4	0	0	7
36	III	Severe	Moderate	No	4	7	0	0	5
37	III	Severe	Severe	Yes	9	2	0	0	7
38	III	Severe	Severe	No	7	4	0	0	7
			Improvement of functional status						
39	IV	Absent	NA	Yes	4	4	2	1	5
40	IV	Absent	NA	No	0	2	2	7	1
41	IV	Mild	NA	Yes	4	6	0	1	6
42	IV	Mild	NA	No	0	3	6	2	4
43	IV	Moderate	NA	Yes	8	2	1	0	7
44	IV	Moderate	NA	No	1	7	3	0	5
45	IV	Severe	NA	Yes	11	0	0	0	7
46	IV	Severe	NA	No	9	2	0	0	7

*See box 6 for decision rules regarding summary scores.