Wide Social Participation in Prioritizing Patients on Waiting Lists for Joint Replacement: A Conjoint Analysis

L. Sampietro-Colom, MD, PhD, M. Espallargues, MD, PhD, E. Rodríguez, PhD, M. Comas, BSc, J. Alonso, MD, PhD, X. Castells, MD, PhD, J. L. Pinto, PhD

**Objective.** The aim was to develop a priority scoring system for patients on waiting lists for joint replacement based on a wide social participation, and to analyze the differences among participants. **Methods.** Conjoint analysis. Focus groups in combination with a nominal technique were employed to identify the priority criteria (N = 36). A rank-ordered logit model was then applied for scoring estimations. Participants (N = 860) represented: consultants, allied-health professionals, patients and their relatives, and the general population of Catalonia. **Results.** Clinical and social criteria were selected, and their relative importance (over 100 points) was: pain (33), difficulty in doing activities of daily living (21), disease severity (18), limitations on ability to work (10), having someone to look after the patient (9), being a caregiver (6), and recovery probability (4). Estimated criteria coefficients had the expected positive sign and all were statistically significant (P < 0.001). There were differences between groups; pain was rated higher by patients/relatives, and difficulty in doing activities was rated lower by patients/relatives and the general public. Most interaction terms for these criteria and groups were significant (P < 0.001). Consultants and allied-health professionals had the most similar prioritization pattern (τ = 0.97). **Conclusion.** Both clinical and social criteria are considered for prioritization of joint replacement surgery from a wide social perspective. The preference among professional and social groups varies and this might impact the result of patient prioritization. A wide social participation for obtaining adequate prioritizing systems for patients on waiting lists is desirable. **Key words:** priority setting; elective surgery; hip/knee replacement; conjoint analysis; preferences; social participation. (Med Decis Making 2008;28:554–566)

Waiting lists are common in health care systems with universal coverage. This is especially true for elective surgical procedures such as total hip or knee replacement. In Catalonia, a Spanish region with about 7 million inhabitants, the waiting lists for joint replacement (JR) are high both in volume and in waiting time. The management method for the patients on these waiting lists is not explicit, which raises concerns by the patient and public about equity.

Several international initiatives have developed scoring systems that prioritize patients according to their needs. Scoring systems are appropriate when the prioritization depends upon more than 1 relevant attribute. Scoring priority systems were initially applied to organ transplant and elective surgery. Most of these systems are based on criteria derived from a consensus obtained from physicians and other health professionals. The patients’ or other social groups’ views had little or no direct input. If preferences for prioritization among the patients, the general population, and health professionals differ, the systems based on these studies may not reflect societal preferences.

Who should be consulted when deciding such criteria? Arguably, under a tax-based system the participation of all individuals, including consultants, general practitioners, nurses and allied-health professionals, social workers, patients and their relatives, and the general public, would be preferable. Broad participation is especially needed when values, perceptions, and preferences regarding health care differ among patients, physicians, and the public, as they often do.

Address correspondence and reprint requests to L. Sampietro-Colom, MD, PhD, Office for the Assessment of Innovation and New Technologies, Fundació Clinic, Villarroel 170, Barcelona 08036, Spain; telephone: 34-93-2275400, ext. 4090; e-mail: lsampiet@clinic.ub.es.

DOI: 10.1177/0272989X08315235

**COMMUNITY BASED PREFERENCES**
This study aimed to develop a priority scoring system for patients on waiting lists for JR based on a process with wide social participation. We also assessed differences in preferences among participants.

METHODS

We used conjoint analysis (CA) to develop a point-count scoring system for setting priorities. The technique is used to rate the relative importance of different attributes in the provision of a good or service. The technique is based on the premise that any good or service can be described by its characteristics, or attributes, and that the extent to which an individual values a good or service is dependent on the levels of these characteristics. Therefore, CA allows the establishment of a citizen’s preferences among multiple attribute alternatives. In this study, the alternatives were the patients’ order in the waiting list, and the attributes were the criteria considered as relevant for prioritization purposes. CA was developed in 5 steps: 1) identify the relevant criteria for patient prioritization; 2) assign levels to the criteria; 3) select the patient scenarios; 4) establish preferences; and 5) statistical analysis. These 5 steps were grouped into 2 stages.

Stage 1: Criteria Identification, Level Assignation, and Scenario Selection

Criteria were identified and defined by means of focus groups. A total of 4 groups were organized, including: the general population (individuals with neither the clinical condition studied nor close relatives affected by it; group 1); patients and close relatives, group 2; allied-health professionals (GPs, nurses, social workers, and physiotherapists; group 3); and consultants (orthopedic surgeons, rheumatologists, rehabilitators, and general practitioners [GPs]; group 4). Allied-health professionals were separated from consultants to avoid domination by the latter. GPs were included in groups 3 and 4 because their practice provided them with both a clinical and a social perspective of the disease process.

Each group consisted of about 5–10 individuals plus an experienced moderator and an observer. The groups were selected on the basis of convenience to reflect different sociodemographic and professional characteristics. The moderator conducted all sessions using a standard guide to assure consistency. All group discussions were tape recorded and subsequently transcribed to allow external validation. Two independent reviewers analyzed the transcripts to elicit the meaning of the criteria identified by groups.

A nominal group technique was subsequently used with the aim of summarizing the group discussions. Participants in each group scored, on a scale from 1 (least important) to 9 (most important), the criteria previously identified by the focus group. The mean score for each criterion was obtained and used to create a ranking list for each group. Seven criteria were selected from these lists as the relevant prioritization attributes. These took into account the criteria with the highest mean value and in the event of a tie the criterion identified by the greatest number of groups was chosen. This cutoff of 7 criteria was established because it was a manageable number both for the subsequent scenario prioritization task and for the application of the scoring system in everyday clinical practice. A financial incentive (60 euro) was provided for participating in this stage of the project. The criteria identification was carried out at the headquarters of the Catalan Ministry of Health in the year 2000.
Criteria were categorized into different relevant levels by the research team based on a review of previous experiences. All possible combinations of criteria levels were generated with each combination becoming a patient scenario. A fractional factorial technique was performed to select an achievable number of scenarios to be prioritized by the participants (Orthoplan procedure in SPSS 11.0: SPSS Inc., Chicago, IL). This technique guarantees that the selected scenarios maintain orthogonality (i.e., there is no collinearity among the scenarios’ criteria levels), and valid results can be obtained for the unselected scenarios.

**Stage 2: Scoring System Estimation**

We interviewed new representative samples of the participant groups. A simple random sample was drawn from the list of registered professionals in Catalonia for selection of consultants and allied-health professionals (group 1). Patients and relatives (group 2) were recruited from a multistage probability sampling, the primary sampling unit being the public hospitals stratified by geographic size, complexity, and JR volume, and the second the individual. For the latter, quota sampling was used according to age, gender, and operated/awaiting from the Catalan Register of patients on waiting lists. Finally, a multistage probability random sample was obtained from the general population. The first sampling unit was the municipality stratified by geographic area and size. The second sampling unit was the individual, obtained from the phone directories, by means of quota sampling according to age, gender, and geographic size. Based on a previous study, a minimum of 279 participants per group was considered necessary to estimate the priority scores, in a measurement scale from 0 to 100, with a precision of 5 points (type I error of 0.05).

Individuals were presented with the selected patient scenarios and asked to rank them from the highest to the lowest priority for surgery. Written information regarding the meaning of each criterion, and levels, was attached to the scenarios in a language understandable to each group, especially clinical information. Participants were also asked to order from highest to lowest the importance of the 7 criteria used in the scenarios. Face-to-face interviews were used for the general public and patients/relatives. A postal survey was used for consultants and allied-health professionals. After several reminders, nonrespondent professionals were contacted by telephone and asked to return the questionnaire. No financial incentive for participants was provided.

**Statistical Analysis**

**Score Estimation**

A rank-ordered logit model was applied for scoring estimations. This is a parametric multivariate regression technique where the independent variables are the different characteristics that define the scenarios (i.e., patients), and the dependent variable is the ranking of the scenarios obtained from participants. This model took into account the respondent’s most preferred scenario and allowed all the information contained in the ranking of scenarios to be used. Estimated coefficients represent the scores given by the participants for each criteria level. Results were weighted by the number of participants in each group to allow for an equal contribution. The log-likelihood ratio test was used to test whether adjacent parameters within a criterion were different from each other. To facilitate the application of the scoring system in clinical practice, coefficients were transformed to a 0–100 point-count linear scale so that the patient scenario with the highest surgical priority would score 100 points. The score for a specific patient scenario was obtained by the addition of the scores of each criterion level, which will give us the position of the patient in the waiting list. The relative importance of 1 criterion (over 100 points) was obtained by dividing its range, the difference between the highest and the lowest level’s coefficient, by the sum of all the criteria ranges.

**Validity of the Model**

The validity of the model was tested using different approaches. We evaluated face validity as whether the parameter’s coefficients increase with the severity of the level of each criterion. Construct validity was assessed by correlating the ranking of the criteria provided by each participant (i.e., N = 860) with the ranking resulting from the model. We assigned a score from 1 to 7 to each criterion (1 point for the criterion placed in first position, 2 points for the second, and so on) for each respondent. The criteria were then ranked according to their mean importance; and this rank was correlated (Spearman coefficient) with the relative importance obtained from the model. Predictive validity was assessed by the correlation (Spearman coefficient) between the ranking of scenarios given by each
participant (observed) and the ranking of scenarios obtained through the model (predicted). Finally, we estimated the percentage of participants with low percentage of dominance (i.e., given 2 scenarios with all levels of criteria equal but 1, the scenario with the worst level is prioritized before the other). Since we did not find in the literature what is considered a “low percentage” of dominances, we considered a “low” percentage of dominance ≤ 25%.

We also analyzed the stability of ranking information. Typically, the stability of ranking information is the lowest in the middle ranks. To test this hypothesis, we estimated the rank-ordered logit model for the pooled top 3 and the pooled bottom 3 scenario choices (set 1) and compared the results with those from the middle placed scenarios in the ranking. To compare both models we used the likelihood ratio test reported by Ben-Akiva and others.

Analysis of Group Differences in Preferences

To assess difference in preferences between groups, we 1) constructed a separate model for each of the 4 groups of participants (subgroup analysis) and compared the relative importance of each criterion by group; 2) constructed an overall model, then introduced a main effect term for the group and interaction terms (group*criterion). The log-likelihood ratio test was used to assess the statistical significance of the interaction terms. The reference group was changed to achieve all possible comparisons between groups; 3) compared the scores of all possible patient scenarios and their ranks according to the preferences reported by participant groups. For each participant group a rank of the 288 patients’ scenarios was generated applying the estimated scores from each subgroup regression model. These 4 ranks were correlated by pairs, which allowed 6 possible comparisons, by means of the Spearman’s correlation coefficient. Mean differences in scores, employing the 100-point scale, and in ranks between pairs of the same patient scenario were also calculated. The statistical software used was SPSS 11.0 (SPSS Inc.) and LIMDEP 7.0.

RESULTS

Stage 1: Criteria identification, Level Assignation, and Scenario Selection

A total of 36 individuals participated in the 4 focus groups and identified about 15 different criteria. There were no substantial differences between groups in the number and type of criteria. Seven criteria were selected and subjectively grouped into 2 larger meaningful domains. The first domain was composed of the clinical criteria, which included disease severity (according to clinical exploration and complementary tests), pain, and recovery probability. The second domain was composed of the social criteria, which included difficulty in doing activities of daily living (ADL), limitations on the ability to work, having someone to look after the patient, and being a caregiver. Two or 3 levels were established for each criterion. The combination of levels of criteria produced 288 possible scenarios (2 criteria with 3 levels and 5 criteria with 2 levels: \(3^2 \times 2^5 = 288\)). Sixteen

---

**Scenario A**

1. Very severe disease
2. Mild pain
3. High (>75%) probability of recovery
4. He/she has great difficulty in doing activities of daily living
5. His/her ability to work is limited due to the disease
6. He/she has someone to look after him/her (the patient)
7. He/she (the patient) has nobody to look after

---

**Figure 1 Example of a patient clinical scenario for prioritization of joint replacement (including clinical and social criteria) and explanations of the wording presented to each group of participants.**

1. **Consultants, allied health-professionals, patients and relatives, general public:** The diagnostic tests show advanced joint degeneration (>III Lawrence-Kellgren level affecting one compartment. Hip: allowed squat < 90°. Knee: mobility limitation > 30°, frontal alignment in varus > 30° and valgux > 20°, showing instability)

2. **Consultants, allied health-professionals, patients and relatives, general public:** Drugs for pain have intermittently been taken during the previous months. Pain appears intermittently; sometimes of high intensity and with some movements the pain increases. He/she is lame.

3. **Consultants, allied health-professionals, patients and relatives, general public:** After surgery, patient has > 75 possibility of success out of 100.

4. **Consultants, allied health-professionals, patients and relatives, general public:** Allows to walked 1-10 blocks, always needs support (e.g. walking stick) to walk and to go up and down the stairs, he/she needs a large shoehorn to put on socks and shoes, can use public transport with some difficulties, needs help when showering, needs help to go shopping, and needs help to stand up and sit down in a chair or bed.
scenarios were finally selected using the fractional factorial technique and exemplified in Figure 1.

**Stage 2: Scoring System Estimation**

A total of 860 individuals were interviewed to obtain the scoring system: 300 from the general public, 347 patients and relatives, 117 allied-health professionals, and 96 consultants. Table 1 shows the main characteristics of each participant group.

Table 2 presents the groups’ overall estimated coefficients for each criterion level. All coefficients within each attribute were statistically different from the omitted category ($P < 0.001$) and from the adjacent category (results not showed). That is, participants discriminated clearly among all levels within each attribute. Moreover, the greater the level of each criterion (e.g., pain severity), the higher the coefficient’s value obtained. Severe pain had the highest relative importance score, followed by being unable to do most of ADL. Recovery probability had the lowest score. Clinical criteria, such as disease severity, pain, and recovery probability, accounted for 55 (over 100 points) of the total relative importance and social criteria accounted for the rest (Figure 2).

The model showed face validity (e.g., severe pain scores higher than moderate pain and the coefficients signs are all positive and statistically significant). The model also showed a good construct validity (Spearman $r = 0.93$ was observed between the order of criteria estimated from the model and the order given by all participants) and a good
predictive validity (Spearman $r = 0.99$ was observed between the order of the scenarios applying the model values and the order given by participant). Ninety-three percent of participants (range = 84.9%–100%) had a low percentage of dominances. The stability of the results from the baseline model was rejected (likelihood ratio test = 23.29, which is above the 5% significance critical $\chi^2$ value). Lack of stability introduces bias when estimating the full rank model. The last column in Table 2 shows that there is only a small influence of this bias in the ranking of the scenarios (i.e., the correlation between the results from the full model and the set 1 model was 0.97).

Subgroup analysis shows that the largest coefficient was associated with pain in all groups. Difficulty in doing ADL and disease severity came next but in different order depending on the group (Table 3). The other criteria were, in order of importance, limitation on ability to work, having someone to look after the patient, being a caregiver, and recovery probability. However, there were differences in preferences between the groups. The coefficient associated with pain was highest in patients/relatives (41 of 100 points), followed by consultants (35 of 100 points), and then the general public and allied-health professionals (both 29 of 100 points). Difficulty in doing ADL was weighted lower by patients/relatives and the general public than by allied-health professionals and consultants (12 of 100, 17 of 100, 25 of 100, and 27 of 100, respectively; $P < 0.001$). The reverse was found for disease severity (20 of 100, 26 of 100, 16 of 100, and 10 of 100, respectively; $P < 0.001$). All the groups similarly valued the rest of criteria.

Correlation of the ranks for all possible patient scenarios was the highest between consultants and allied-health professionals ($r = 0.97$) and the lowest between consultants and the general public ($r = 0.77$).

### Table 2  Rank-Ordered Logit Models Estimated for All Groups Together (Overall) for Prioritization of Joint Replacement ($N = 860$)

<table>
<thead>
<tr>
<th>Estimated Coefficients$^a$</th>
<th>95% Confidence Interval</th>
<th>Value (0–100 Scale)$^b$</th>
<th>Value (0–100 Scale) Set 1$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>–1.64</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Disease severity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate (ref)</td>
<td>0</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>Severe</td>
<td>0.92</td>
<td>0.89—0.98</td>
<td>17.63</td>
</tr>
<tr>
<td>Pain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mild (ref)</td>
<td>0</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>Moderate</td>
<td>0.88</td>
<td>0.85–0.91</td>
<td>16.96</td>
</tr>
<tr>
<td>Severe</td>
<td>1.73</td>
<td>1.71–1.75</td>
<td>33.28</td>
</tr>
<tr>
<td>Recovery probability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate (ref)</td>
<td>0</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>High</td>
<td>0.20</td>
<td>0.18–0.22</td>
<td>3.80</td>
</tr>
<tr>
<td>Difficulty in doing ADL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has some difficulty (ref)</td>
<td>0</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>Has great difficulty</td>
<td>0.52</td>
<td>0.50–0.55</td>
<td>10.04</td>
</tr>
<tr>
<td>Unable to do most of ADL</td>
<td>1.06</td>
<td>1.03–1.09</td>
<td>20.46</td>
</tr>
<tr>
<td>Limitations on ability to work</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No/does not work (ref)</td>
<td>0</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>Yes</td>
<td>0.53</td>
<td>0.51–0.55</td>
<td>10.29</td>
</tr>
<tr>
<td>Has s/o to look after the patient</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes (ref)</td>
<td>0</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>No</td>
<td>0.45</td>
<td>0.43–0.47</td>
<td>8.61</td>
</tr>
<tr>
<td>Be a caregiver</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (ref)</td>
<td>0</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>Yes</td>
<td>0.31</td>
<td>0.29–0.33</td>
<td>5.93</td>
</tr>
</tbody>
</table>

Note: ref = reference category; ADL = activities of daily living; s/o = someone.

a. $P < 0.001$.

b. The estimated coefficients and standard errors have been transformed to a 0–100 scale, so that the patient in the worst situation will score 100 points and the patient in the best situation will score 0 points.

c. Model for only the top 3 and the bottom 3 choices.
Interaction terms (group*criterion) were statistically significant ($P < 0.001$) for the majority of comparisons (data not shown). There were a few differences between health professionals and consultants. In general, more differences were likely to appear for criteria that received higher weights, such as pain, difficulty in doing ADL, and disease severity (88% of these interaction terms were significant; $P < 0.001$). The score differences between groups were more likely to be in the center of the rank distributions (intermediate patients’ scores). A $2 \times 2$ group comparisons of all possible scenarios would lead to an average difference between 3.7 and 12.1 points. Importantly, these figures would translate to a mean difference between 15.9 and 46.4 positions, respectively, in the rank (Table 4).

When scores from subgroup analysis are applied to the scenarios, differences in their waiting list rank were observed (Table 5).

Figure 2  Relative importance of each prioritization criterion for joint replacement according to participating group and all groups together.
As in many other areas of health care, there has been little public discussion concerning the setting of priorities for waiting lists.37 There is little information on whether the diverse individuals involved agree on a more explicit prioritization for waiting lists and on what criteria the list should be based.38 Our study has provided a prioritization system for patients waiting for JR. This prioritization system was based on a wide societal participation using CA to elicit preferences. This method identified and quantified differences among the participating groups, and it allowed us to estimate the potential impact on the order of patients in the waiting list when applied.

### Table 3

Rank-Ordered Logit Models Estimated for Each Group for Prioritization of Joint Replacement: Estimated Coefficient (EC), Confidence Interval (CI), and Transformed Values (TV)

<table>
<thead>
<tr>
<th>Group 1 General Public (n = 300)</th>
<th>Group 2 Patients and Relatives (n = 347)</th>
<th>Group 3 Allied-Health Professionals (n = 117)</th>
<th>Group 4 Consultants (n = 96)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC&lt;sup&gt;a&lt;/sup&gt;</td>
<td>CI&lt;sup&gt;b&lt;/sup&gt;</td>
<td>TV&lt;sup&gt;c&lt;/sup&gt;</td>
<td>EC&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Constant</td>
<td>–1.37</td>
<td>—</td>
<td>–1.32</td>
</tr>
<tr>
<td>Disease Severity Moderate (ref)</td>
<td>—</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>Severe</td>
<td>1.17</td>
<td>1.12–1.23</td>
<td>34.19</td>
</tr>
<tr>
<td>Pain Mild (ref)</td>
<td>—</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>Moderate</td>
<td>0.69</td>
<td>0.63–0.75</td>
<td>13.72</td>
</tr>
<tr>
<td>Severe</td>
<td>1.29</td>
<td>1.25–1.33</td>
<td>25.65</td>
</tr>
<tr>
<td>Recovery probability Moderate (ref)</td>
<td>—</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>High</td>
<td>0.09</td>
<td>0.05–0.90</td>
<td>1.79</td>
</tr>
<tr>
<td>Difficulty in doing ADL Has some difficulty (ref)</td>
<td>—</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>Has great difficulty</td>
<td>0.30</td>
<td>0.25–0.35</td>
<td>6.16</td>
</tr>
<tr>
<td>Unable to do most of ADL Limitations on ability to work No/does not work (ref)</td>
<td>—</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>Yes</td>
<td>0.48</td>
<td>0.44–0.52</td>
<td>9.54</td>
</tr>
<tr>
<td>Has s/o to look after the patient Yes (ref)</td>
<td>—</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>No</td>
<td>0.40</td>
<td>0.36–0.44</td>
<td>7.95</td>
</tr>
<tr>
<td>Be a caregiver No (ref)</td>
<td>—</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>Yes</td>
<td>0.26</td>
<td>0.22–0.30</td>
<td>5.37</td>
</tr>
</tbody>
</table>

Note: ref = reference category; ADL = activities of daily living; s/o = someone.

a. P < 0.001.
b. 95% confidence interval.
c. The estimated coefficients and standard errors have been transformed to a 0–100 scale, so that the patient in the worst situation will score 100 points and the patient in the best situation will score 0 points.

**DISCUSSION**

As in many other areas of health care, there has been little public discussion concerning the setting of priorities for waiting lists.37 There is little information on whether the diverse individuals involved agree on a more explicit prioritization for waiting lists and on what criteria the list should be based.38 Our study has provided a prioritization system for patients waiting for JR. This prioritization system was based on a wide societal participation using CA to elicit preferences. This method identified and quantified differences among the participating groups, and it allowed us to estimate the potential impact on the order of patients in the waiting list when applied.

The relevant criteria identified in this study do not differ substantially from those in other studies (Canada7 and New Zealand5). Similarly to these previous studies, where only clinicians selected and scored the criteria, age was not a priority criterion.
This criterion was heavily discussed during the focus group. However, there was no agreement on whether the younger or older individuals deserved prioritization, and it was not finally selected. Nevertheless, participants chose the criterion of work limitation, which may act as a surrogate of age as is demonstrated in a subsequent pilot study. It is difficult to make a more quantitative comparison between our coefficients and those in the Canadian and New Zealand studies since attribute and levels were slightly different.

The amount of time waiting for surgery is the unique explicit criterion for prioritization presently used in Catalonia (Spain) and in other countries with a national health system. Time spent on a list is an extrinsic factor not related to the patient’s need for surgery. The criteria identified in our study included both clinical and social domains. All these criteria reflected the patient’s need for surgery and represented what society, including health professionals, preferred for patient prioritization. Making these criteria explicit, and implementing the prioritization system, might lead to greater equity because patients with the highest need would be operated on first. However, nationwide implementation of a scoring system may be difficult. A recent study showed that the use of a scoring system for prioritizing patients on waiting lists, although mandatory, was not followed by all the physicians. Several reasons may exist. A mandatory system for prioritization may...
threaten a physician’s authority. Conversely, the physician may see the system as a helpful tool when explaining to a patient why they have to wait a much longer time or why they do not need surgery.40

Another explanation might be the reluctance of clinicians to use the scoring system on the grounds that it increases bureaucracy; however, in a pilot study, scoring a patient did not take more than 2–3 minutes.43 Moreover, other health care professionals, such as nurses, could score the social criteria and leave the clinical criteria to the physicians to simplify the procedure.

We observed some differences in prioritization preferences among participants. Differences concentrated on the valuations of disease severity, pain, and difficulty in doing ADL. Patients and relatives gave less weight to difficulty in doing ADL and scored pain much higher than the other groups. This may be explained by previous disease experience, such as suffering from pain, adaptation to everyday activities, and ability to cope with impairments that seem worse to the unaffected population.42

Although consultants should be intimately familiar with the patients they treat and their health status, disparities between them have been observed.18,43–46 Physicians often underestimate the impact of health status issues compared with the patients’ perception.43 However, in our study both groups of health professionals gave higher scores to pain and disability than to criteria such as disease severity or probability of recovery. In fact, disease severity was perceived to be more important the less technically understood it was: consultants (10 of 100 points), allied-health professionals (16 of 100 points), patients/relatives (20 of 100 points), and general public (35 of 100 points). Severity might have been underscored by professionals because they are aware of the possible discrepancy between clinical signs, and/or diagnostic tests, and functional capacity.1 Therefore, they may consider it as a poor measure of current patient disability. Also, the criterion wording per se, for instance the use of “severity” instead of “state,” might have influenced the response of those groups with no medical background.

Recovery probability receives the lowest weight both in the aggregate and in the subgroup analysis results. The fact that JR has been shown to be highly effective47,48 may have led participants to consider this criterion to be a poor discriminator for prioritization. However, other reasons may exist. The literature suggests that when an attribute includes uncertainty, such as treatment success, it may lead to significant framing effects depending on whether the information is expressed in losses or gains. Additionally, prospect theory states that in decisions involving uncertainty, individuals have a generalized aversion to loss.49 In our study, recovery probability was presented as a gain, or possibility of success, and the potential benefit offered by surgery would compensate for the risk of failure. Therefore, it was given a low relative importance in the system and showed a small difference within levels. Finally, ethical considerations might have played a role in the low score obtained. There is evidence of similar trade-offs between efficiency and equity in other health-related contexts.42

Would differences observed between groups impact the prioritization of patients when our scoring system was applied? Although the scoring criteria pattern was the same, the estimated weights were not. This would lead to variation in the patient’s position in the list, and to different waiting times, depending on which group scores were considered. The discrepancies would be greater for patients with some “worst” levels for highly weighted criteria. Therefore, looking at differences in preferences for prioritization of patients awaiting for surgery, and the impact on the position of the patient when using the scores from one or another group, how should such disagreement be resolved when designing a priority index? Someone may argue that health professionals must have a higher weight because they are more knowledgeable about health care. Others may be in favor of giving higher weight to patients’ answers since they bear the actual positive or negative consequence of treatment. Finally others may agree on weighting higher the general public answers since they are objective citizens, with the potential to be a patient in the future, and their taxes finance health care. However, developing a priority system for waiting lists implies value judgments and not just technical knowledge. Therefore, in the development of a social priority scoring system we recommend both wide societal participation from the beginning as well as to weight equally the answers from the different citizens groups. Moreover, we think that the higher the participation level by patients in setting priorities for waiting lists, the greater will be the acceptability and trust in the process. However, this may clash with corporative attitudes50–53 and raise costs of health care policy making.

Some limitations of our study deserve comment. First, the low participation of health care professionals, especially consultants, in both the first and the second stage of the study may have introduced a selection bias. However, it was reassuring that the
criteria identified were consistent across the groups and did not differ from those previously described in the literature. In the second stage there were no statistically significant differences in gender and age between those health professionals who answered the questionnaire and those who did not. Moreover, the main symptom of the lack of power is getting nonsignificant results when differences are evident. In our case, the results from the rank-ordered logit show statistically significant parameters. Second, although we are confident that the criteria for identification and selection are valid\(^5,7,38\) some dependency between the clinical and the social criteria (difficulty in doing ADL) might exist. The wording and levels were also simplified to develop patients’ scenarios, and the operational descriptions were adapted to the language of the target population. These modifications might have introduced both a label and a wording bias and delivered information asymmetrically among groups, which may be responsible for some of the observed differences. The prioritization system gave rise to 288 different patient profiles (scenarios). However, participants must be assumed to provide participants with a manageable number of scenarios while still allowing to obtain the scores for all possible combinations. The main-effects design precludes analyzing interactions among attributes. In our case, a possible interaction between pain and limitation in ADL may affect rankings.

It is worth mentioning that the task of ranking scenarios showed good intraobserver reliability in a pilot study (test-retest correlation coefficient \(r = 0.8\); data not shown). Evidence suggests that the more complex a choice becomes in terms of the number of options (criteria) and the higher the variability within these options, the more often there exists the motivation to simplify the decision process and focus on 1 or 2 options.\(^49\) This could explain in part the consistency in the order of the criteria according to their relative importance across the 4 groups. However, interviewer training and not limiting the interview duration might have decreased the likelihood of this phenomenon as well.

Additionally, it is known that it is easier to rank the most preferred and least preferred alternatives than the alternatives in the middle ranks, whose preferences may not be so clear for participants, thus questioning the pooling of all patient scenarios. Using the likelihood ratio tests we have explored this issue (see Methods section). The results show that, although the order of relative importance of the criteria is the same, there is statistical difference among parameters, confirming this already known methodological problem. However, our study differs from the traditional market research, where the interest is focused in knowing the first scenarios chosen, because we want to know the preferred order of the range of patients in a waiting list. Therefore, from a “health policy” point of view we think the results obtained are useful. Finally, we did not postvalidate the results estimated from the analysis with a different sample population. However, the correlation between the order of preference of criteria given by participants during the survey with their relative importance estimated by the model was high (\(r = 0.93\)).

In conclusion, the increase in consumerism and the requests for more accountability advocates the involvement of citizens in health care decisions. This is most important when preferences for health care options by the individuals involved have been seen to differ and value judgments are involved in the patient prioritizing decision. The potential impact on practice could be important. Moreover, having an objective and explicit scoring system derived from all concerned individuals would be expected to provide a need-based patient prioritization, improve equity, and may increase the credibility of citizens in the system.

REFERENCES

COMMUNITY BASED PREFERENCES


