Using a point system in the management of waiting lists: the case of cataracts

Eva Rodríguez–Míguez*
Department of Applied Economics, University of Vigo,
Lagoas- Marcosende s/n, 36310 Vigo, Spain.

Carmen Herrero
Departamento de Fundamentos del Análisis Económico,
University of Alicante and IVIE, 03071 Alicante, Spain.

José Luis Pinto–Prades
Department of Economics and Centre for Health Economics,
Pompeu Fabra University, Ramon Trias Fargas 25-27, Barcelona, Spain.

Abstract

In the management of waiting lists point systems could be a useful mechanism to establish priorities amongst patients. In this paper we explore the possibility of using Conjoint Analysis (CA) in order to implement a point system based on social preferences. We conducted an experiment for the specific case of cataract extraction. In spite of the pilot nature of the study the results seems to suggest that CA is a feasible method in order to estimate a point system.

KEY WORDS: point system; waiting lists; cataract; conjoint analysis; rank-ordered logit model.

JEL: C25, I18.

Acknowledgements: We are grateful to two anonymous referees for helpful comments and suggestions. Financial support from CICYT PB97-0120 and SEC2000-1087 is acknowledged.

* Corresponding author. E-mail: emiguez@uvigo.es; Tel: +34 986 813529; Fax: +34 986 812 401.

1. Introduction

When demand exceeds supply in a public health care unit, patients are placed in a queue, i.e. on a waiting list. The queue may be organised on a simple first come first serve basis, or the patients may be prioritised such that patients with certain characteristics get shorter waiting times than others. In the latter case the task of prioritising is normally left to the discretion of clinicians at the health care unit. From the patients' point of view this may be unsatisfactory in terms of lacking transparency, lacking consistency with views of fairness in the wider community and perhaps lacking impartiality in some individual cases. Ideally one could therefore envisage a system in which clinicians assign priority points to patients according to stringent criteria decided by the wider community. This hypothetical system lies behind the suggestion of the British Medical Association when they suggested that waiting time should not be the key element in the management of waiting lists. They suggested that *patients should be given severity scores when they were put on a waiting list, which would reflect their clinical priority and how quickly they should receive surgery* (Fricker, 1999).

If we decide that several characteristics of patients have to be used in order to manage waiting lists we have to decide how to use those characteristics in order to establish priorities. One method is to rank those characteristics by importance and then rank patients according to their position in the most important characteristic. The second most important characteristic is used to rank patients when they are identical in the most important characteristic and so on. For example, let us assume that gender and age are relevant characteristics. Let us also assume that gender is the most important one, e.g., women should go first. Then all women independently of their age would be prioritized over any man. If we want to establish priorities among women, then age would be used. However, there is a second possibility. Even if a characteristic is considered more important than other the position in the most relevant characteristic does not guarantees priority, it also depends on the second characteristic. Following the above example, even if gender is more important, society may want to prioritize a patient like (man, 20 year old) over a patient like (woman, 80 years old). We say that the first method is based on a lexicographic principle and the second method is based on a compensatory principle.

Reforms implemented in the management of waiting lists in New Zealand (Seddon, Almos, Ramanathan, McLaughlin & White, 1999; Jackson, Doogue & Elliot, 1999) have been based in a compensatory point system. This system gives priority to patients on the basis of their clinical and social characteristics. For example, in the case of fertility services, the characteristics that were used to prioritize among patients were probability of success, age, number of children, etc. (Gillet & Peek, 1997). Compensatory point systems have also been used to allocate organs in the US —the UNOS system— (Pierce, Kauffman, Ellison, Edwards, Klein, Wolf et al., 1996; Leffell, Bennett, Ellison & Zachary, 1997). However, the novelty of the New Zealand experience is that they used a point system for elective surgery over a wide variety of health problems.

In the belief that this approach has many positive features, we decided to explore the possibility of developing and applying such a system for a specific condition, namely cataracts, which was chosen simply for convenience. Also, we wanted to based the point system in the preferences of the general population and not in the opinion of a group of experts as was the case in New Zealand. We chose to base our estimates on a sample of the general population because we believe that the choice of the relevant characteristics and their relative importance is not a technical issue. It is an issue related to social values and for this reason it should be based on preferences of the general population (Gold, Siegel, Russell & Weinstein, 1996; Dolan, Olsen, Menzel & Richardson, forthcoming). In this paper we explore if we can incorporate social preferences in the management of waiting lists using conjoint analysis (CA). The objective of the paper is then mainly methodological, that is, if we want to base a point system in social preferences, is CA an appropriate method? This paper reports a study in which a small random sample of the general population were asked to rank a number of hypothetical patients with different characteristics in terms of the priority they thought these patients should have in a queue. By means of CA a model for point assignment is developed.

2. Methods: conjoint analysis

To design a point system that could be used to manage waiting lists for cataracts, and which would take societal preferences into account, the technique of CA was used (Luce & Tukey, 1964). Although CA has been widely used in market research since the mid–1970s, and in health economics since the mid 1990s, its use in the management of waiting lists has been minimal (Ratcliffe, 2000).

CA is used to elicit individuals' preferences for sets of multi–attribute alternatives (products, services, etc.). The technique is based on obtaining weights for the different levels of each attribute that are more consistent with an individuals's overall preferences for the set of alternatives. In this study, the alternatives will be patients who require a given health–care service, i.e., an operation for cataracts, and the attributes will be the characteristics of the patient that are considered to be relevant in deciding their positions on a waiting–list, including severity, time on the waiting–list, etc. Beginning with overall preferences, our objective was to obtain a value for every level of each attribute, to be able to calculate a total priority score for each one.

In compliance with the CA methodology, the following steps were followed to obtain a point system that could be used to prioritize patients for cataract–extraction: a) selection of attributes and levels; b) model specification and selection of combination of attributes (patient types) to be rated in the survey; c) selection of a method for data collection; d) questionnaire design; e) selection of participants; f) selection of the estimation method; and g) selection of tests to assess validity of the results.

A) Selecting attributes and levels.

This selection was based on the review of the existing literature (Mordue, Parkin, Baxter & Steward, 1994; Hadorn & Holmes, 1997; Halliwell, 1998) and on a two 2–hour in–depth interviews with 4 opthalmologists. They worked in two hospitals with a very high volume of cataract operations. They were chosen given their wide experience with the management of waiting lists for cataracts. The process that we followed with clinicians was:

1. We explained to them the point system for cataracts followed in New Zealand (Hadorn & Holmes, 1997).
2. They were asked if they thought that the system was meaningful for them. Given their positive response we asked them about the characteristics of patients that they considered relevant for priority setting according to their clinical experience.
3. They assigned points to those characteristics more relevant for them on a scale from 1 to 9.
4. We chose the characteristics with a score higher than five on average.
5. We asked the ophthalmologists their opinion about the most logical levels in

these attributes.

6. After this meeting, using their opinion and some literature review, we elaborated the levels.

The following attributes and levels were selected for inclusion in the CA.

Visual incapacity (I). This attribute indicates the severity of visual impairment, and is measured by specialists using objective techniques to produce a rating of what is known as *visual acuity*. The concept of visual acuity, however, is very much a medical one, and its inclusion in a questionnaire designed for use in the general population would be impractical. To overcome this problem, and after consulting specialists involved in the study, the concept of visual acuity was converted into an attribute based on the number of activities that the patient could or could not perform. The greater the number of activities which the patient cannot perform due to their impaired vision, the greater their degree of impairment.

In order to develop this attribute, an index of functional impairment called the VF–14 (Steinberg, Tielsch, Schein, Javitt, Sharkey, Cassard et al., 1994) was used. The VF–14 measures patients' level of incapacity when performing 14 vision–dependent activities of daily living, such as driving, reading small print, or watching television. Using the VF–14 it was possible to rank each of the 14 activities included in the instrument according to the degree of visual capacity that was required to perform each activity (Alonso, Espallargues, Andersen, Cassard, Dunn, Bernth-Petersen et al., 1997). This ranking permitted the design of an attribute —visual incapacity— which could be easily understood by members of the general population (see Appendix A). Four levels of visual incapacity were established: a) *mild*, b) *moderate*, c) *severe*, and d) *very severe*.

Limitations in daily activities (L). This attribute refers to the patient's capacity to perform the everyday activities that he or she performed before becoming visually impaired. The information provided by this attribute is different to what is provided by the visual incapacity attribute, as it is possible that the same level of visual incapacity might affect different individuals in very different ways, in terms of their everyday activities. For example, the effect of mild visual incapacity (unable to read small print or to drive) on a pensioner who hardly ever drove, and on a lorry–driver, would clearly be very different, with the latter being much more severely limited, in terms of everyday activities.

This attribute has 2 levels: a) *the patient has some problems in performing everyday activities (for example: working, doing housework, participating in family or leisure activities)* and b) *the patient is unable to perform the majority of his or her everyday activities*. This two levels were chosen following the corresponding dimension of the EQ-5D (Dolan, 1997).

Likelihood of improvement (K). This attribute indicates the likelihood of the patient recovering full sight after the operation. For the sake of simplicity, only total or minimal recovery of sight are considered as possible outcomes. It should be remembered that cataracts do occur along with other eye diseases (macular degeneration, glaucoma, etc.) which reduces the likelihood of a successful operation. The probability of success is also reduced by the presence of comorbid conditions, such as diabetes. This attribute therefore records the effectiveness of the operation. According to the existing literature (Schein, Steinberg, Cassard, Tielsch, Javitt & Sommer, 1995; Mangione, Orav, Lawrence, Phillips, Seddon & Goldman, 1995) and the expert opinion of the four ophthalmologists interviewed, the likelihood of improvement was then divided into three levels: a) moderate (50%), b) high (75%), c) very high (99%).

Patient's age (A). Given that this disease is prevalent among elderly patients, this attribute is categorized as follows: a) *50 years* b) *65 years* c) *75 years*, and d) *85 years of age*. The two extremes (50 and 85) where chosen according to the clinical experience of ophthalmologists. We were told that there were some exceptional cases of patients younger than 50 or older than 85, however, the ophthalmologists considered that most of their patients were within this interval.

Time on waiting list (W). This attribute has been classified using 4 levels: a) *3 months*, b) *6 months*, c) *12 months*, and d) *18 months*. ophthalmologists considered that it was very rare a patient waiting more than 18 months.

The attributes and levels described above define 384 ($4\times2\times3\times4\times4$) patient types. By classifying the patients into types, we avoid any other considerations that are not relevant to prioritize patients.

B) Model specification.

A CA specifies the relationship between the total priority score assigned to a patient j, denoted as V_j, and the weight assigned for each of the levels of the attributes that characterize this patient. In CA the additive model is usually formulated. This model can be specified as follows:

$$V_j = \sum_{t=1}^{5} x_{jt}\alpha_t , \qquad (1)$$

Where x_{jt} indicates the level of attribute t in the patient j and α_t is the vector of parameters to be estimated.

Under an additive model only main effects are estimated (the value of each attribute level assuming everything else held constant) given that no interactions are assumed. For this reason we do not need to use all possible combinations to obtain the parameters of interest. We only need to analyze a fraction of all possible combinations. A nice -although not essential- property of the combinations of attributes chosen is orthogonality, whereby the interattribute correlation is zero (see Appendix B for an intuitive explanation of orthogonality). Such property allows for a minimum number of combinations to estimate main effects only. The SPSS ORTHOPLAN procedure produces cards with this characteristics. Using this statistical package 16 cards were selected (see Appendix B).

C) Choosing a data collection method.

A variety of different preference elicitation methods exist. They include asking respondents to compare a series of two alternatives; to choose one alternative from a set of alternatives; to rate the full set of alternatives using one of a variety of scaling methods, or; ranking the set of alternatives. We consider the latter approach to be more suitable for the objectives of this study, because: first, it is important to know how all of the patients on the waiting list will be ordered; second, the use of a ranking method allows respondents to order the full set of alternatives, without an undue cognitive burden (the use of pair–wise comparisons to rank states would make the exercise unduly arduous); finally, given that the aim is to order patients on the waiting list, an ordinal measure will be sufficient.

D) Designing the questionnaire.

The questionnaire consists of 4 different sections. The first part explains the study's objectives and the different attributes and levels included, and allows the respondent to ask questions about the study. In the second part, the respondent lists the 16 types of patient selected according to the order in which he or she thinks the patients should be operated on. To make the task easier, a description of each of the 16 patients is printed on a separate card, so that the respondent simply has to put them in order. In the third part, the respondent is asked to put the 5 attributes —I, L, K, A and W— in order according to the importance they assign to each attribute in this prioritization exercise. The respondent is also asked whether there are any other patient characteristics they would consider relevant —as this information may well be valuable in the design of future surveys. Finally, the respondent is asked to review his order of the cards and to modify it if they consider it appropriate.

E) Selecting the respondents.

The questionnaire was presented to 100 individuals drawn from the general population of Barcelona (47% men and 53% women) and representing different age groups (23% were between 18 and 29 years of age, 28% between 30 and 44, 31% between 45 and 64, and 18% were over 65). Subjects were contacted by telephone using random dialling. Those that belonged to the appropriate gender and age quota were offered to be interviewed at his/her home. Those who accepted to be interviewed were visited by an interviewer who had been specially trained in the administration of the questionnaire.

F) Estimation method: rank–ordered logit model.

Using the ranking provided by the participants as a starting point, we arrived at an evaluation (weight) for each level of each attribute which is a necessary step towards obtaining a point system —expression (1). Given the type of the information obtained, that is, the ranking of 16 hypothetical patients, we chose a statistical model that estimates the parameters taking into account the ordinal nature of the data. We think that the method that better adapts to our data is the rank-ordered logit model (ROL) (Beggs, Cardell & Hausman, 1981). See Appendix C for a more detailed explanation of ROL.

Two models were estimated for V_j. In the first model (the *general model)* we obtained one parameter for each of the levels of the attributes. In this model, α_t is a vector that has as many elements as the t attribute has levels. As is usual in the literature, to avoid exact collinearity in the estimation, we parametrized the model so that one level of each attribute was excluded. Specifically, we excluded the levels labeled as "mild visual incapacity", "has some problems in performing everyday activities", "moderate likelihood of improvement", "85 years of age" and "3 months on waiting list". Therefore, the parameters associated with the remaining levels of an attribute must be interpreted in relation to the excluded level: a positive (negative) coefficient for a given level indicates a higher (lower) score for this level with respect to the excluded level.

From estimated parameters in the general model, the relative importance (RI) of each attribute is obtained by dividing its range, i.e. the difference between the highest and lowest coefficients, between the sum of the ranges of all attributes (Rosko, DeVita, McKenna & Walker, 1985).

In the second model (the *adjusted model)* continuous variables are approximated by continuous functions. The advantage of this second model is its capacity to provide a valuation for those intermediate levels that were not considered in the survey. In this model the sign of the coefficient in continuous variables indicate the relation between priority and the variables. For example, a positive coefficient in waiting time implies that the longer the patient has been waiting the more priority should receive. The consistent Akaike information criterion (CAIC) has been used to select the adjusted model (see last paragraph in Appendix C).

The computer program used to perform the estimations was LIMDEP 7.0.

G) Validity of the results.

The validity of the results was tested in three ways:

Face validity. The parameters associated to the different levels of each attribute should follow a logical order. We hypothesized that the levels excluded in order to apply the statistical model corresponded to the level of each attribute that gives less priority to patients in a waiting list. For this reason we expected a positive coefficient for each parameter. Furthermore, the value of the parameter should be higher the more we depart from the excluded level. For example, if we think that society prefers to prioritize those patients with more severe visual incapacity (being equal in the rest of characteristics) then the parameter associated to the level "severe" should be higher than the parameter associated to "mild" or "moderate". In the case of continuous variables (adjusted model) we expect a positive sign for likelihood of improvement and waiting time and a negative coefficient for age. This is then a test that should hold at the aggregate level.

Predictive validity. The correlation between the respondents' rank ordering and the order generated by the general model was tested at individual and aggregated levels. The Spearman rank correlation coefficient was used to calculate the correlation between each individual's rank ordering and the order generated by the model, and the average correlation coefficient was then calculated for all participants. To calculate the correlation at an aggregate level, individual orderings were aggregated using the Borda rule, and the Spearman rank correlation coefficient between this ordering and the ROL generated ordering was obtained. The Borda rule assigns points to cards such that the lowest–ranked card of each participant receives zero points, the next–lowest one point, and so on up to the highest–ranked card, which receives 16–1 points. Points for each card are summed across all participants and the place of each card on the social ordering is determined by the total points.

Construct validity. The correlation between the rank ordering assigned directly to the 5 attributes by the respondents and that obtained from their relative importance was calculated. To do so, individual orderings of the attributes were aggregated using the Borda rule and the Spearman rank correlation coefficient between direct ordering and estimated ordering of the attributes was calculated. This comparison, although intuitively attractive, is in fact relatively unorthodox and the results obtained should be treated with caution.

In addition internal consistency and robustness is tested.

Internal consistency. This is a test that is conducted at the individual level. Given the assumption of rationality in the prioritization process, respondents should give greater priority to greater incapacity, greater likelihood of improvement, greater limitations in performing usual activities and longer time on the waiting list. With regard to age, any preference is considered to be rational here. This is compatible with the hypothesis that we did for age above (face validity). At the aggregate level we expect a negative sign for age because it has been found in the empirical literature that it is more frequent to find people tending to choose younger patients over older patients (Busschbach, Hessing & Charro, 1993; Rodríguez & Pinto, 2000). However, at the individual level we cannot consider that a subject is inconsistent if he/she has opposite preferences.

When comparing all pairs of alternatives that could be extracted from the 16 alternatives, 5 pairs were produced where one alternative was clearly superior to the other, i.e. age was the same in the two alternatives, and of the remaining attributes at least one was at a higher level and none at a lower level. In such a situation, the superior alternative should rationally receive a higher ranking. Internal consistency was tested by determining the percentage of respondents who were allocated rational rankings in each of these 5 cases.

Robustness to sample size. To determine the degree of robustness of the rankings to sample size, the Spearman rank correlation coefficient was obtained between the ranking generated by the model using the full sample and the ranking generated using half of the sample —randomly selected.

3. Results

Rankings of the 16 patients and the 5 attributes were obtained from all respondents. The interview lasted 45 minutes on average.

Estimated models.

The estimates generated by the general model are shown in Table 1. All the coefficients estimated were significant, at the 1% level, and were in the expected direction. For example, as visual incapacity increases so does the value of the associated parameter, and hence the likelihood that the patient will be operated on earlier.

As can be seen, the two attributes with the greatest relative importance were visual incapacity (37%) and the patients age (28.7%), whilst the least important was the likelihood of improvement (6.6%).

On the basis of the CAIC value the adjusted model shown in Table 2 was chosen. The adjusted model introduces a lineal adjustment for age, $v_a = \alpha_a A$, and a logarithmic adjustment for time on the waiting list (measured in months), and the likelihood of improvement, $v_w = \alpha_w Ln(W)$ and $v_k = \alpha_k Ln(K)$, respectively. The coefficients of the qualitative variables are barely altered and the continuous variables are in the expected direction. *Age* has a negative sign, and both *likelihood of improvement* and *time on the waiting list* have positive signs.

Coefficients of Table 2 can be used to establish priorities amongst all patients. In order to rank-order patients in the waiting list we will add up the coefficients that represent the situation of each patient. For example, a patient with severe visual incapacity and nine months on waiting list will receive 1.534 points from the first problem and $0.309 \times Ln(9) = 0.679$ points from the second problem so he/she will receive 2.213 points out of these two characteristics. A patient with mild visual capacity will receive 0 points out of this characteristic and if he/she remained fourteen months on waiting list will receive 0.815 points out of this characteristic so he/she will receive a total of 0.815 points of these two characteristics. The higher the score the higher the priority of the patient. Since we are only interested in the ranking of patients, we will be only interested in the ranking properties of the final score and not on the size (distance) between the final score. For example, let two patients A and B, have the following attributes: χ_A= (*very severe, has problems, 60%, 60 years, 4 months)* and χ_B= (*moderate, is unable,* 80%, *70 years, 14 months*). Given that A's total priority score,

$$\hat{V}\left(\chi_A\right) = 1,766 + 0 + 0,457 \times Ln(60) - 0,038 \times 60 + 0,309 \times Ln(4) = 1,785,$$

is greater than B's,

$$\hat{V}\left(\chi_B\right) = 0,716 + 0,766 + 0,457 \times Ln(80) - 0,038 \times 70 + 0,309 \times Ln(14) = 1,640,$$

A will be given greater priority than B. If we estimate the score of all patients in the waiting list we can rank-order the patients according to their characteristics. Those patients with higher scores will be treated first.

Validity.

Face validity. As mentioned, the parameters were in the expected direction.

Predictive validity. With regard to predictive validity, the Spearman rank correlation coefficient between ranking estimated using the ROL and the direct rankings of respondents was 0.67 at the individual level (average) —only 25% had a correlation below 0.64— and 0.98 at the aggregate level.

Construct validity. The ranking of the attributes obtained by aggregating the rankings provided by the respondents coincides with the ranking obtained by estimating relative importance using the ROL (correlation coefficient of 1). Although the comparison of both methods should be seen as merely a simple approximation, the agreement in the ranking reinforces the results obtained.

Internal consistency. Of the 5 pairs of alternatives where the consistency can be tested, one of the pairs was rationally ordered by 100% of the respondents, 3 of the pairs were rationally ordered by 99% of the respondents, and the remaining pair was rationally ordered by 94% of the respondents.

Robustness to sample size. The correlation coefficient between the ordering estimated by the model using the full sample and the other one using half of the sample was 0.96.

4. Discussion

There is currently considerable debate in the health economics literature regarding the characteristics that should be taken into account when deciding on a patient's place on a waiting list, with increasing agreement on the idea that more than merely medical characteristics should be taken into account. However, increasing the number of characteristics to be considered also makes the management of waiting lists more complicated. This paper suggests that a point system may provide a useful tool for health care authorities when establishing priorities. The study also suggests that CA could be a methodology that can be used to incorporate social preferences in the management of waiting lists. It was encouraging to find that all participants were able of ranking the 16 cards with a high level of internal consistency. It seems that this task can be performed by subjects of quite different social conditions. However, more evidence is needed with a larger sample.

A point system may face some problems if it is going to be used in practice. One hypothetical problem could be the attitude of clinicians towards the system. Clinicians may consider such a system as an interference in their clinical freedom. However, when we presented these results to a group of ophthalmologists they considered that such a system could be quite useful to them. They said that the system could take away some of the pressures that patients usually were putting into them to receive higher priority. Doctors thought they could more easily justify their decision appealing to a set of "objective" principles (the point system) that are used consistently in the health system. The second reason was that doctors acknowledged that in some cases they were already taking into account other characteristics apart from waiting time. They acknowledged that the process was quite subjective and sometimes inconsistent. Even more, doctors said that the point system was closer to their preferences than the system based only on waiting time. They said that in everyday practice it was problematic to prioritize patients with greater need (mainly worse visual incapacity) because managers were very reluctant to use other principles apart from waiting time.

It is true that in real decision making a point system may have some other problems that cannot be perceived in a pilot study like this one. For example, the model explains preferences across the 16 hypothetical situations quite well. That is not to say that the correlation between the samples preferences across 16 actual persons would be equally high, since a number of other characteristics of the persons then also would be visible to the judges. However, this is a problem of all models, that is, they have to simplify. On the other hand, if the point system takes into account the most important characteristics that affect priorities, it would be very useful for managers, even if it is not perfect. Of course, in some special cases the point system may no be appropriate because there may be some circumstances that are not be taken into account by the point system.

Another problem that can be raised in the real world is that there may be random error in clinicians' scoring of patients on relevant characteristics. This may reduce the predictive accuracy of the model.

A question related with the practical use of this system is the status that should be assigned to the point system. Should the system be decisive or should it be envisaged as merely as a guide? Talking to doctors we perceived that they preferred the system to be decisive. This is not to say that they think that the system should not admit exceptions.

Another issue that can be raised is the relation of this point system with QALYs.

In principle, there is a relation between both. If we prioritize patients according to QALY we are using health gain as the only characteristic that is taken into account to manage waiting lists. However, the point system includes other variables and then incorporates other kind of preferences. Of course, these preferences could also be incorporated into QALYs, using some kind of adjustment. For example, in the case of age we could construct age-weighted QALYs. With other characteristics like waiting time it would be more problematic. However, we think that the main advantage of a point system over QALYs is that it may be easier to understand by doctors and patients. Consequently it is easier to justify politically and to apply in practical decision making. The priorities obtained with a point system may be closer to social preferences (have higher predictive validity) than priorities obtained using QALYs.

Although the results of the experiment are encouraging the study did have some limitations. On the one hand, patients of a similar type will have the same level of priority. Nevertheless, in some cases patients belonging to the same category may be perceived as being different, particularly if important attributes have not been included in the analysis. For example, the qualitative analysis revealed that whether the patient was responsible for others or not was important for some respondents. The same might be true if the number of levels per attribute were reduced too much. For example, given equality at the other levels, patients who cannot recognize people at close range might be considered to have a much higher priority than those who cannot read large print in a book. With respect to continuous variables, the points assigned to intermediate levels were obtained using an adjustment function. However, these values should be confirmed in future studies. Cards with intermediate levels, for example, could be incorporated into the survey to test the ability of the model to predict the relative position of those levels.

Another limitation of the study was sample size. It should be remembered, however, that this study was a pilot study to determine the validity of the methodology. Although the test of robustness suggest that results were consistent within this particular sample, sample size should be considerably increased to ensure that the results are representative of society's preferences. It would be useful, in future research, to specifically elicit preferences from all the groups directly implicated. This could be done in two ways. On the one hand, focus groups could be carried out with patients, patients' relatives, members of the general population and doctors. This would help not only in the choice of attributes but also in the levels used with each attribute, which is a key feature in obtaining a point system. On the other hand, the survey should be administered not only to members of the general population, but also to health care professionals, patients and members of the patients' families. Although it might be assumed that preferences of the different groups will be similar, this needs to be investigated in future studies. Also, the study does not provide any evidence of test-retest reliability and this should be investigated further.

References

Addelman, S. (1962). Symmetrical and asymmetrical fractional factorial plans. *Technometrics,* 4, 47-58.

Alonso, M., Espallargues, M., Andersen, T.F., Cassard, S.D., Dunn E, Bernth-Petersen P et al. (1997). International applicability of the VF–14. An index of visual function in patients with cataracts. *Ophthalmology,* 104, 799-807.

Beggs, S., Cardell, S., & Hausman, J. (1981). Assessing the potential demand for electric cards. *Journal of Econometrics,* 16, 1-19.

Browning, C.J., & Thomas, S.A. (2001). Community values and preferences in transplantation organ allocation decisions. *Social Science & Medicine*, 52, 853-861.

Busschbach, J.J.V., Hessing, D. J. & Charro, F. T. (1993) The utility of health at different stages in life: A quantitative approach. *Social Science and Medicine,* 37(2), 153-158.

Chapman, R., & Staelin, R. (1982). Exploiting rank ordered choice set data within the stochastic utility model. *Journal of Marketing Research*, XIX, 288-301.

Dolan, P., Olsen, J.A., Menzel, P., & Richardson, J. An inquiry into the different perspectives that can be used when eliciting preferences in health. *Health Economics* (forthcoming).

Dolan, P. (1997). Modeling valuations for EuroQol health states. *Medical Care*, 35(11), 1095-1108.

Fricker J. (1999). BMA proposes strategy to reformulate waiting lists. *British Medical Journal*, 318(7199), 1698–1699.

Gillett, W., & Peek, J. (1997). *Access to infertility services: Development of priority criteria*. Wellington: National Advisory Committee on Health and Disability Support Services.

Gold, M.R., Siegel, J.E., Russell, L.B., & Weinstein, M.C. (1996). *Cost-effectiveness analysis in health and medicine*. New York: Oxford University Press.

Hadorn, D.C., & Holmes, A.C. (1997). The New Zealand priority criteria project. Part 1: overview. *British Medical Journal,* 314(7074), 131–134.

Halliwell, T. (1998). How fair is cataract prioritisation?. *New Zealand Medical Journal*, 1111(1076), 405-407.

Jackson, N.W., Doogue, M.P., & Elliot, J.M. (1999). Priority points and cardiac events while waiting for coronary bypass surgery. *Heart,* 81, 367-373.

Leffell, M.S., Bennett, L.E., Ellison, M.D., & Zachary, A.A. (1997). Impact of changes in the UNOS renal allocation system. *Transplantation Proceedings*, 29(8), 3435-3436.

Luce, D.R., & Tukey, P.E. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1, 1-27.

Mangione, C.M., Orav, E.J., Lawrence, M.G., Phillips, R.S, Seddon, J.M., & Goldman, L. (1995). Prediction of visual function after cataract surgery. *Arch Ophthalmol*, 113, 1305-1311.

Mordue, A., Parkin, D.W., Baxter, C., & Steward, M. (1994). Thresholds for treatment in cataract surgery. *Journal of Public Health Medicine*, 16(4), 393-398.

Pierce, G.A., Kauffman, H.M., Ellison, M.D., Edwards, E.B., Klein, D.H., Wolf, J.S., et al. (1996). UNOS scientific registry: present and future. *Transplantation Proceedings*, 28(4), 2355-2357.

Ratcliffe J. (2000). Public preferences for the allocation of donor liver grafts for transplantation. *Health Economics*, 9, 137-148.

Rodríguez, E. & Pinto, J.L. (2000). The Social Value of Health Programs: Is Age a Relevant Factor? *Health Economic*s, 9, 611-621.

Rosko, M., DeVita, M., McKenna, W.F., & Walker, L.R. (1985). Strategic marketing applications of conjoint analysis: an HMO perspective. *Journal of Health Care Marketing*, 5(4), 27-38.

Schein, O.D., Steinberg, E.P., Cassard, S.D., Tielsch, J.M., Javitt, J.C., & Sommer, A. (1995). Predictors of outcome in patients who underwent cataract surgery. *Ophthalmology*, 102, 817-823.

Seddon, M.E., French, J.K., Almos, D.J., Ramanathan, K., McLaughlin, S.C., & White, H.D. (1999). Waiting times and prioritisation for coronary artery bypass surgery in New Zealand. *Heart*, 81, 586-592.

Steinberg, E.P., Tielsch, J.M., Schein, O.D., Javitt, J.C., Sharkey, P., Cassard, S.D., et al. (1994). The VF-14. An index of functional impairment in patients with cataract. *Arch Ophthalmol*, 112(5), 630-638.

Appendix A

Visual incapacity levels	Activities the patient is unable to perform
Mild incapacity	Driving Reading small print (telephone directory, medicine bottles, etc)
Moderate incapacity	“The above plus the following”: Doing fine handwork (sewing, putting in nails, etc) Reading a newspaper or book / filling out forms
Severe incapacity	“The above plus the following”: Reading traffic, street or store signs Watching televison / seeing stairs Leisure activities such as bowling, gardening or window–shopping Playing cards, dominoes or bingo
Very severe incapacity	“The above plus the following”: Cooking Reading large print in a book or newspaper Recognizing people at close range

Appendix B

Very severe incapacity Has some problems ... Moderate likelihood ... 50 years 3 months on waiting list	Severe incapacity Is unable ... Very high likelihood ... 50 years 6 months on waiting list	Moderate incapacity Is unable .. High likelihood ... 50 years 1 year on waiting list	Mild incapacity Has some problems ... Very high likelihood... 50 years 18 months on waiting list
Very severe incapacity Is unable ... Very high likelihood ... 75 years 1 year on waiting list	Severe incapacity Has some problems ... High likelihood ... 75 years 18 months on waiting list	Moderate incapacity Has some problems ... Very high likelihood ... 75 years 3 months on waiting list	Mild incapacity Is unable ... Moderate likelihood ... 75 years 6 months on waiting list
Very severe incapacity Has some problems ... High likelihood ... 65 years 6 months on waiting list	Severe incapacity Is unable ... Very high likelihood ... 65 years 3 months on waiting list	Moderate incapacity Is unable .. Moderate likelihood ... 65 years 18 months on waiting list	Mild incapacity Has some problems ... Very high likelihood ... 65 years 1 year on waiting list
Very severe incapacity Is unable ... Very high likelihood... 85 years 18 months on waiting list	Severe incapacity Has some problems ... Moderate likelihood ... 85 years 1 year on waiting list	Moderate incapacity Has some problems ... Very high likelihood ... 85 years 6 months on waiting list	Mild incapacity Is unable ... High likelihood ... 85 years 3 months on waiting list

Orthogonality is satisfied when the joint occurrence of any two levels of different attributes appear in cards with frequencies equal to the product of their marginal frequencies (Addelman, 1962). For example, “Severe incapacity” has a marginal frequency equal to 4/16 (it appears 4 times out of 16) and “Is unable to perform the majority of his or her everyday activities” has a marginal frequency equal to 8/16. Therefore, the pair (Severe incapacity, Is unable to perform the majority of his or her everyday activities) should have a frequency equal to 2/16.

Appendix C

To choose an estimation model, we must take the type of information obtained into account. In our study each interviewee ranked 16 patients as those who should be operated on first . The rank–ordered logit model (ROL) not only takes the respondents most preferred alternative into account, as the ordered logit model does, but it also allows for the use of all the information contained in the ranking of alternatives. This additional information can provide more precise estimates of the unknown parameters.

Unlike other methods of analysis used in the management of waiting lists that also explore the information contained in a ranking (Browning & Thomas, 2001), the basic specification used to derive the ROL is the random utility model. This model considers that the preference relation of respondent i over the choice set J can be represented by a utility function U_{ij} that measures the non–observable utility assigned by respondent i to alternative j —defined as a vector of attributes χ_j. U_{ij} is assumed to be composed of two terms, one deterministic, $V\left(\chi_j\right)$, which represents the systematic component of the model, and the other stochastic, e_{ij}, which captures the measurement error. If one assumes that these two components are independent and additive, the random utility model may be written in the form

$$U_{ij} = V\left(\chi_j\right) + e_{ij} \equiv V_j + e_{ij}.$$

We specify a particular linear form for V_j,

$$V_j = \chi_j \alpha = \sum_{t=1}^{5} x_{jt} \alpha_t,$$

where α is the parameters vector (or matrix) to be estimated.

The maximum likelihood method used to estimate the ROL allows us to obtain consistent and (asymptotically) efficient parameters. See Beggs, Cardell & Hausman (1981) or Chapman & Staelin (1982) for a more detailed description of the model.

In order to select the adjusted model for V_j the consistent Akaike information criterion (CAIC) has been used. The CAIC allows us to compare non–nested models based on the log–likelihood function, penalising the decrease in degrees of freedom produced as the parameters increase. Its value is obtained from the expression $-2LnL(\alpha)+(1+Ln\ \ n)k$, where n is the number of observations, k is the number of parameters and $L(\alpha)$ is the log-likelihood function. The model with the smallest CAIC being the most preferred.