The stated and tacit impact of demographic and lifestyle factors on prioritization decisions for cardiac surgery

F. KEE, P. McDONALD, J. R. KIRWAN, C. C. PATTERSON and G. LOVE

From the Departments of Epidemiology and Public Health and Medicine, Queen’s University of Belfast, Belfast, Northern Health and Social Services Board, Ballymena, and Department of Medicine, University of Bristol, Bristol, UK

Received 11 November 1996

Summary

In a clinical judgement analysis, we used linear regression models to reflect the impact of clinical and non-clinical cues on priority decisions, by comparing the stated prioritization policies of 30 clinicians with their actual policies as revealed by an appraisal of 50 ‘paper patients’. Correspondence was modest for some cues, e.g. 25 doctors said they accounted for age, but age only had a significant bearing in the derived decision models of two doctors. Correspondence between the derived and expressed weights was greatest for clinical angina grade and the presence of left main stem stenosis. Correlation between the rank order of importance between the two models was poor for most of the cues, and statistically significant only for smoking. However, stated policies made it appear that lifestyle factors such as smoking habit would influence prioritization decisions for most clinicians but policies derived from actual prioritization decisions seldom related to lifestyle or demographic variables. There were significant differences in the degree of correlation between the two models according to the experience of the clinician. However, correspondence was not significantly better for doctors with cardiological training than those without. The overall contribution of demographic and lifestyle factors to decision making appears to be small, suggesting that they should be omitted from prioritization guidelines.

Introduction

Despite much rhetoric on the pros and cons of according priority for coronary artery bypass surgery to particular types of patients, there is little data in the UK which actually demonstrates the impact of clinical and non-clinical factors on waiting times for surgical revascularization. A prominently-debated issue in this country has been the influence of demographic or lifestyle factors, such as age or smoking habit, on waiting-list priority. Naylor et al. have used a ‘clinical judgement analysis’ (CJA) approach to devise an urgency score based on judgements made on a matrix of ‘paper’ patients, (defined in terms of prognostic predictors, such as angina class or the number of stenosed vessels), but no account was taken of patients’ demographic and lifestyle characteristics. However, if it was deemed that a smoker or an obese subject might gain less from bypass surgery than a non-smoker, it might seem reasonable to accord that patient less priority. We have previously surveyed the views of over 500 local general practitioners on what factors they would deem it justifiable to take account of when assigning priority for coronary bypass surgery. Generally, when such factors were taken into account, it was because they were deemed to affect the efficacy of the operation.

Apart from the fact that doctors’ assessments of cardiac risk may be quite inaccurate, a problem with the direct survey approach is that doctors are often poor at identifying the key influences on their decisions. Clinical judgement analysis can generally better describe the basis of decision-making than can doctors acting alone by introspection, and making these processes explicit can improve consensus on judgements about disease severity and response to treatment.
A more accurate reflection of current practice may thus be obtained using the CJA approach. We used this methodology to compare the stated prioritization policies of a group of doctors attending a regional symposium on the subject with their actual policies as revealed by an appraisal of 50 paper cases.

Methods

Our methods have been described in detail elsewhere but, essentially, the key task which the clinicians were asked to undertake was an appraisal of ‘paper’ patients. In fact these cases were based on a random sample of real patients who had undergone bypass surgery for coronary artery disease in Northern Ireland in 1991.

Paper cases were devised following a pilot study with two consultant cardiologists. Each patient was described on a single A4 page and was characterized by 10 clinical ‘cues’ including age, sex, smoking habit (smoker vs. non-smoker), percentage of ideal body weight, presence or absence of significant co-morbidity (such as chronic airways disease), results of an exercise stress test (high risk vs. not high risk), angina severity (modified Canadian Cardiovascular Class), left ventricular function (normal or abnormal), presence or absence of left main stem stenosis, and the total number of significantly diseased vessels. Each participating doctor was visited by one of us (FK) to explain the background for the proposed exercise and was given a folder of 50 ‘paper patients’ to assess. A sample patient is shown in Appendix I.

Importantly, they were asked to consider the cases as identical in all other respects apart from the given clinical cues. The doctors were first asked to indicate their perceptions of the likely efficacy of bypass surgery in its ability to improve the patient’s symptoms, reduce the risk of infarction and to extend the patient’s life. They were then asked to re-arrange the cases, in any way they saw fit, until reaching a final priority order for surgery (1 → 50).

Statistical methods

Multiple regression analysis develops an equation to express the relation between one variable, the dependent variable, and several others called the predictors or independent variables. In the case of CJA, the dependent variable is the judgement, in this case the rating of priority for bypass surgery, and the independent variables are the clinical cues. In practice, linear regression models have provided fairly robust representations of clinical judgements. While we have previously used all-possible-subsets regression for model selection, we found that broadly similar judgement policy models were selected using a stepwise (backwards) procedure. Consequently, for computational efficiency, the latter was our method of choice. There is no entirely satisfactory method of overcoming the problem of cue intercorrelation, about which much as been written in the CJA literature and this is why we set a relatively conservative p value for inclusion in the model (0.10) so that we minimized the risk of inappropriately rejecting cues.

The contribution of each clinical variable or cue to the judgement policy model is represented by its contribution to R$^2$ which was assessed by dropping each variable in turn from the model (the change in the type II Sum of Squares, $cR^2$). We explored an alternative approach by which comparison between equations from different judges is more convenient, when $cR^2$ is expressed relative to that of all the other cues in the equation ($rcR^2$), a method which standardizes for variation in the models’ explanatory power.

Although neither method entirely overcomes the problem of collinearity, the rank order of importance of the cues in the decision models was not changed.

Prior to undertaking this exercise, each participating clinician completed a short questionnaire to identify their background training (e.g. whether a cardiologist or not) and to state how much impact they thought each clinical cue had on their assessments of priority for surgery. This was to be done by assigning a number (or ‘weight’) to each cue such that the sum of all the weights was 100.

These ‘expressed’ weights for each clinical cue were compared, using Spearman’s rank order correlation, to the contribution which each made to R$^2$ (the explained variance) for that doctor’s derived decision model. In this latter case, we felt it more logical to use a fully-saturated model (and the respective contributions to the total sum of squares), as the questionnaire to obtain expressed weights had asked the doctors to ensure summation to 100. The presence or absence of co-morbid conditions was not specified explicitly in the questionnaire but was specified under an ‘other factors’ category by just four doctors. For the purposes of the rank order correlation, it has been excluded from both the stated and derived policies prior to analysis. To determine whether the correlation between expressed and derived or tacit decision models differed between groups of doctors we transformed the coefficients to a z score by a method provided by Fisher:

$$z = \frac{1}{2} \log_e (1 + r) - \log_e (1 - r)$$

which is distributed normally with a standard error

$$\sigma = 1/\sqrt{(n-3)}.$$
the Regional Workshop. All participants were consultant physicians, cardiologists \( n = 15 \) or cardiac surgeons \( n = 4 \).

Following a stepwise backwards elimination procedure, a model to explain the prioritization decisions was derived for each of the 30 judges. The extent to which each clinical cue added explanatory power to the final decision model is given by its contribution to the overall variance \( (R^2) \). The models generally had high explanatory power, the mean \( R^2 \) being 82% but increasing to 86% when the doctors’ perceptions of the efficacy were initially forced into the model. The relative importance or ‘relative weight’ of each of the nine major clinical cues is given for the derived and the expressed prioritisation policies in Figure 1. Figure 2 compares the frequency with which each cue was either in or out of the decision-making model (i.e. had either a non-zero or zero weight) and to this extent, at least, allows the degree of correspondence between the derived and expressed weights to be assessed. The correspondence was poor for some of these cues. For example, twenty-five doctors said they accounted for age when they prioritized patients, but this was the case in the derived decision models in only two instances. Twenty-seven participants said they were influenced by smoking, but smoking only had a significant bearing in the derived decision models of five of these doctors. The apparent correspondence was greatest for clinical angina grade and the presence of left main stem stenosis.

Table 1 gives the coefficients of correlation between the rank order of importance of each cue in the derived and the expressed decision models.

![Graph](image)

**Figure 1.** Contribution of each clinical cue to the variance explained by derived prioritization decision model and clinical cue weights in expressed decision model.
Figure 2. Inclusion/exclusion of clinical cues in expressed and derived prioritization decision policies.

Table 1 Correlation of rank order of importance of clinical cues between expressed and derived prioritization decision models

<table>
<thead>
<tr>
<th></th>
<th>Doctor’s experience</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trained in cardiology</td>
</tr>
<tr>
<td>Age</td>
<td>0.10</td>
</tr>
<tr>
<td>Sex</td>
<td>0.07</td>
</tr>
<tr>
<td>Smoking status</td>
<td>0.37*</td>
</tr>
<tr>
<td>Body mass index</td>
<td>−0.03</td>
</tr>
<tr>
<td>Stress test</td>
<td>0.05</td>
</tr>
<tr>
<td>Angina grade</td>
<td>0.26</td>
</tr>
<tr>
<td>Left ventricular function</td>
<td>0.15</td>
</tr>
<tr>
<td>Left main stem stenosis</td>
<td>0.15</td>
</tr>
<tr>
<td>Number of affected vessels</td>
<td>−0.003</td>
</tr>
</tbody>
</table>

*Overall significant difference, p < 0.05. **Significant difference according to experience of clinician.

Discussion

Both the National Audit Office and the Clinical Standards Advisory Group have advocated the establishment of a broader consensus, if not national guidelines, on criteria for prioritizing cardiac surgery waiting lists. Despite the apparent polarization of views on what impact it would be proper for demographic and lifestyle characteristics to have on any such guideline (the issues were widely debated a few years ago in the pages of the BMJ1–4), we have shown that their overall contribution to the decision-making process is generally small. The most clearcut example of this is for smoking habit. Although 90%
of our participating clinicians expressed the view that they would think it reasonable to take this factor into account when prioritizing patients, analysis of their actual decisions showed that the patient’s smoking habit was accounted for by only 16% of doctors. Before we can achieve any consensus, therefore, the challenge is to clarify the extent of real rather than expressed divergences of policy. Clinical judgement analysis can at least better describe these differences and can reduce the extent of superficial or false agreement which results from inadequate introspection.

Although we found the contribution of ‘non-medical’ factors was small in doctors’ tacit and actual policies for prioritization, this might nevertheless have important implications. High or low medical risks might determine priority for many cases but those with intermediate risks may be near a decision threshold, and therefore be disproportionately susceptible to the influences of demographic or lifestyle ‘cues’. It was recently concluded, for example, that the use of angiography in the Myocardial Infarction and Intervention project, did not seem to be driven by mortality risk stratification, and it may be that doctors take an individual view of the likely efficacy of revascularization for certain patients.

It may be expected that a measure of consensus already exists regarding the major clinical factors such as left main stem disease, angina grade and ejection fraction which are known to predict outcome, and a number of groups around the country are devising priority scoring systems based on these variables. In our study, the greatest weight was given to ‘prognostic’ anatomical factors and angina severity in both the expressed and derived decision models. However, while almost all our clinicians reported taking account of these factors in their decision-making, this was reflected in their actual decisions mainly for angina grade and left main stem stenosis. About half of the derived decision models did not include one or more of stress test, left ventricular function or the number of affected vessels. Although various meta-analyses have already demonstrated how these cues affect revascularisation outcome, previous studies have shown that doctors tend to over-estimate risk and are inclined to believe that most patients who die while waiting for revascularization would have been ‘saved’ by surgery, which is patently not the case. More salutary in this regard is that the correspondence in our study between the expressed and derived decision models of cardiologists was not significantly better for doctors with cardiological training than those without. Indeed others have found that though confidence increases with experience, prognostic accuracy does not, and teaching physicians to make better prognostic estimates may not alter their treatment decisions.

Overall, these doctors’ descriptions of the importance they give to various clinical variables when prioritizing patients for cardiac surgery correlate only modestly with the weights reflected by their actual decisions on paper cases. In general, many of the doctors believed they used more cues than they actually did. In this regard our findings mirror those from a study of GPs’ prescribing habits for patients with hyperlipidaemia. Indeed, it is known from psychological research that even experts can only consider a limited number of cues in making judgements, and that people have great difficulty in forming accurate assessments of risks and probabilities.

A reduction in interpersonal variation in judgement is an essential prerequisite to co-operative decision making and the use of CJA to reveal the systematic element of these variations seems to provide an avenue for reaching agreed policies. While some might criticize the ostensibly ‘artificial’ method of appraising paper patients, several studies have demonstrated that judgements made in response to paper cases resemble those made with actual patients and that ‘process’ or ‘cognitive’ feedback (i.e. revealing cue weights) can both improve the teaching of complex diagnostic strategies, and surpass the agreement between clinicians that mere discussion and exchange of ideas might achieve. A consensus conference might be deemed successful if it produces guidelines that can both reflect the resolved views of local practitioners and are grounded in the best evidence available. Consensus has been the more difficult to achieve with respect to prioritization for cardiac surgery because the original clinical trials were not designed to determine the risks of delay, a fact acknowledged by the Ontario group. Judgements about the appropriateness of care may nevertheless be affected by how the outcomes and the risks are perceived. Within the context of our own workshop, we had to acknowledge that occasionally ‘we may enter a debate with an established main stem stenosis. About half of the derived decision models did not include one or more of stress test, left ventricular function or the number of affected vessels. Although various meta-analyses have already demonstrated how these cues affect revascularisation outcome, previous studies have shown that doctors tend to over-estimate risk and are inclined to believe that most patients who die while waiting for revascularization would have been ‘saved’ by surgery, which is patently not the case. More salutary in this regard is that the correspondence in our study between the expressed and derived decision models of cardiologists was not significantly better for doctors with cardiological training than those without. Indeed others have found that though confidence increases with experience, prognostic accuracy does not, and teaching physicians to make better prognostic estimates may not alter their treatment decisions.

By our participating doctors sharing their decision models in discussion groups, the impact of prior beliefs was made the more apparent, as was their divergence from the actual prioritization decisions made. Our debates may actually have been much more protracted if not sterile, had it not been made clear that while many doctors claimed to be influenced by demographic or lifestyle cues in prioritizing for surgery, far fewer of these decisions were significantly affected in practice.

Following presentation and discussion of the CJA findings, there was no support for inclusion of demographic or lifestyle cues in any local prioritization.
References


Appendix 1.

**Patient No. 10**

Please examine the patient's details given below and respond to the questions at the bottom of the page.

<table>
<thead>
<tr>
<th>AGE</th>
<th>39</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEX</td>
<td>MALE</td>
</tr>
<tr>
<td>SMOKING STATUS</td>
<td>NON-SMOKER</td>
</tr>
<tr>
<td>PERCENTAGE OF IDEAL WEIGHT</td>
<td>114%</td>
</tr>
<tr>
<td>EXPECTED PROCEDURE RELATED MORBIDITY/MORTALITY</td>
<td>AVERAGE</td>
</tr>
<tr>
<td>EXERCISE STRESS TEST</td>
<td>NOT HIGH RISK</td>
</tr>
<tr>
<td>CCS ANGINA GRADE</td>
<td>CCS I</td>
</tr>
<tr>
<td>LEFT VENTRICULAR FUNCTION</td>
<td>NORMAL</td>
</tr>
<tr>
<td>LEFT MAIN STEM STENOSIS %</td>
<td>0 - 49%</td>
</tr>
<tr>
<td>NUMBER OF DISEASED VESSELS - other than left main stem</td>
<td>0</td>
</tr>
</tbody>
</table>

Indicate your views on the following questions by drawing a short vertical line somewhere on the horizontal lines below.

- By having surgery, how much improvement in this patient's anginal symptoms would you expect
  - [ ] None
  - [ ] Symptoms eliminated

- What difference will surgery make to the chances of non fatal infarction?
  - [ ] No difference
  - [ ] Eliminate risk

- By having surgery, what GAIN in life expectancy is this patient likely to acquire?
  - [ ] Yrs. Mths.

- Indicate your views on the urgency of the case for surgery, by drawing a short vertical line at some point on the horizontal line below
  - [ ] Routine waiting list entry
  - [ ] For next available theatre list

- What do you consider to be the maximum safe waiting time for this patient.

  **Maximum safe waiting time**

  Insert code number →

  "1" = 1wk; "2" = 2 - 4 wks; "3" = 5 - 12 wks; "4" = 3 - 6 months; "5" = 6 - 9 months; "6" = 9 - 12 months