Same patient, different advice: a study into why doctors vary

T Rakow, C Bull

Aim: To understand why doctors differ in their recommendations in situations where there is little certainty about the long term outcomes of the possible treatment options.

Methods: A correlational design was used to examine the relation between preference for different treatment options and beliefs about likely outcomes for these options. Eighty doctors, with a mean of nine years in paediatric cardiology/surgery, attending a conference on serious congenital heart disease were studied. Main outcome measures were: ratings of the extent to which each of four treatment options were favoured; and subjective probabilities for three outcomes—death, survival with “good heart function” (New York Heart Association functional class [NYHA] I or II), and survival with “poor heart function” (NYHA III or IV)—for different treatment options over a 20 year time frame.

Results: Preference for one treatment option over another was most closely associated with the subjective estimate of the additional years with “good heart function” that it offered 10–20 years after surgery (Pearson’s $r = 0.66$, $p < 0.001$). In influencing a preference, the possibility of early death was subordinate to optimising the late outcome.

Conclusions: Doctors’ treatment preferences are consistent with selecting the option that maximises the chance of the best outcome (long term survival with good heart function). Doctors’ recommendations imply that they place more value on years of life in the child’s far future than on life-years in the immediate future.

The clinical situation examined in this paper is common in paediatrics: several treatment options are available, each with different profiles of immediate risk and scant availability of long term outcome information. The research question is also of significance across paediatric specialties as it examines the variability in doctors’ recommendations in the face of such uncertainty and the reasons behind the variability. The methodology borrows from the “paper patient” techniques used in research in clinical judgement and decision making.

Our study sought to examine the dominant influences on decision making in a surgical context:

1. Differences in judgement. Differences in opinion about the likely outcomes of surgery may account for differences in doctors’ preferences for different surgical options. Specifically, we might anticipate that the preferred option corresponds to the option with the highest (subjective) probability of putting or keeping the patient in a state with high utility (for example, alive and well).

2. Differences in utility. Doctors may believe that the numbers of life-years associated with two options are the same, but still prefer one option over the other. For instance, individuals might value survival with good heart function and poor heart function differently or value “years soon” differently from “years to come” and this balance may itself vary with the age of the patient. Studies of adult patients’ treatment preferences generally reflect a preference that good health in the near future carries greater weight than good health in the far future. This domain is less well studied in the decisions parents or doctors make on behalf of children.

3. Differences in “local practice”. Preferred management strategies and protocols differ between institutions. It is possible that practice influences preference, either with or without influencing the subjective probability of outcomes. Once a pattern of practice is established, it is possible to follow precedent for an individual case without recourse to reasoning about risk.

The particular situation we use as an example of the clinical situation of interest relates to a surgical decision in the context of complex congenital heart disease. Surgery for univentricular heart is a good example of decision making under uncertainty because surgical risks are high and even risk adjusted models of short term outcome have poor predictive value. Also, the variety and rapid evolution of strategies for palliation means that medium term follow up is frequently unavailable and there is inevitably no evidence at all about patterns of wellbeing or attrition in adult life. These are the uncertain circumstances in which the paediatric physician or surgeon must often function and offer a good environment for us to examine what doctors are trying to optimise when they make a recommendation.

“Univentricular hearts” are a “family” of rare congenital cyanotic conditions in which the systemic and pulmonary circulations are supported predominantly by a single functioning ventricle. Most patients would die without palliation in infancy. The dismal subsequent natural history of univentricular heart prompted the introduction of more aggressive palliations beyond infancy, including the bi-directional Glenn shunt (superior caval vein to pulmonary artery anastomosis) and a variety of total right ventricular bypass operations that divert the whole caval return to the lungs, offering the patients full systemic arterial saturation in return for a high systemic venous pressure. Current techniques generally consist of a bi-directional Glenn and construction of a connection between the inferior caval vein and the pulmonary artery, and are usually called “total cavopulmonary connections” (TCPC). Recent practice has swung towards “staging” the TCPC operation, performing the bi-directional Glenn first and “completing” the inferior pathway some years later. As long

Abbreviations: NYHA, New York Heart Association; TCPC, total cavopulmonary connections

Original Article

See end of article for authors’ affiliations

Correspondence to:
Dr T Rakow, Dept of Psychology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK, timrakow@essex.ac.uk
14 January 2003
term complications of TCPC began to emerge, the issue of whether the bi-directional Glenn alone might be the preferred “definitive palliation” for univentricular hearts began to be debated. Thus for an individual patient with univentricular heart, both “no treatment” and different surgical options are available.

Given their differing profiles of risks and benefits over time, choices between these options are far from straightforward. Is a management strategy that seems to offer a greater chance of 20 year survival justified if it carries a high risk of early mortality? Is it acceptable to try to maximise overall survival if this means selecting an option with poor prospects for good heart function?

METHODS

The participants were 80 doctors attending a paediatric cardiology conference held in the UK, with sizeable representation from European and Commonwealth countries. Of those who indicated their specialty, there were 64 cardiologists, six surgeons, and four from other specialties. The mean time spent in the specialty of paediatric cardiology/surgery was nine years, with 39 doctors describing their current position as “consultant (or equivalent)”.

Knowledge elicitation

In a structured case-conference format familiar to cardiac units, the participants listened to a presentation consisting of the case history and the cardiac haemodynamics obtained at catheterisation of an anonymised real patient, and viewed the corresponding echocardiogram and angiogram. The case was chosen because its difficult underlying anatomy and preceding history would force clinicians to rely largely on subjective judgements—the objects of interest to this study (box 1).

Participants were then asked to rate four possible courses of action for this patient on a visual analogue scale from −5 (“strongly dislike”) to +5 (“strongly favour”):

- Option 1: no surgical intervention.
- Option 2: a “one stage” total cavo-pulmonary connection operation now.
- Option 3: a bi-directional Glenn shunt now as definitive palliation.
- Option 4: a “two stage total cavo-pulmonary connection” with bi-directional Glenn now and completion in five years.

To limit conference time spent on the task, participants were assigned to one of three groups: (A) comparing options 1 and 2 (n = 25); (B) comparing options 2 and 3 (n = 30); and (C) comparing options 2 and 4 (n = 25). The one stage total cavo-pulmonary connection was chosen as a “reference option” to be considered by all doctors. After a short demonstration of multistate survival curves, and a demonstration of how to approach the task, the participants then completed a multistate survival graph for their two options. By adding two lines to the grid provided (fig 1), doctors could indicate their estimates over a 20 year time frame of the chance of:

(a) Mortality.
(b) Survival with poor heart function (New York Heart Association functional class (NYHA) III/IV).
(c) Survival with good heart function (NYHA I/II).

The two options for which graphs were drawn were then rated for a second time (from −5 to +5 as before).

Finally, participants supplied estimates of the numbers of bi-directional Glenn, one stage total cavo-pulmonary connection, and total cavo-pulmonary connection completion operations performed at their institution in the past 12 months; and gave details of their specialty and experience.

Analysis

Each participant had provided a pair of actuarial estimates for two options: one stage TCPC and an “alternative option” (which varied between participants). We explored the relation between their preference among their two options (reflected in the corresponding option ratings) and their beliefs about the survival and wellbeing relating to the two options (reflected in the corresponding survival curves).

Option ratings were recorded to the nearest 0.2, measuring the distance to the centre of the cross placed on the rating scale. From the option ratings we calculated the difference between the rating for the alternative option over the one stage TCPC. We denote

### Table 1: Rating of the four options

<table>
<thead>
<tr>
<th>Option Rating</th>
<th>Mean</th>
<th>SD</th>
<th>Number of first preferences*</th>
</tr>
</thead>
<tbody>
<tr>
<td>No intervention</td>
<td>−3.03</td>
<td>2.08</td>
<td>2</td>
</tr>
<tr>
<td>One stage TCPC</td>
<td>−1.32</td>
<td>3.08</td>
<td>9</td>
</tr>
<tr>
<td>Bi-directional Glenn</td>
<td>1.84</td>
<td>2.44</td>
<td>22.5</td>
</tr>
<tr>
<td>Two stage TCPC</td>
<td>2.55</td>
<td>2.28</td>
<td>44.5</td>
</tr>
</tbody>
</table>

*Two equal ratings are scored as 0.5.
this as “preference”, a single measurement that could take a positive, zero, or negative value, and used it as a dependent variable in subsequent analyses.

For the purposes of our analyses, the multistate survival curves were “read” as providing subjective estimates of the varying probability over time that a particular individual that is, the patient considered) would be in one of the outcome states (dead, alive with good heart function, or alive with poor heart function). This requires the vertical (percentage) scale to be converted to probability (values 0 to 1). The total area under the upper curve then represents the expected number of years survival (in this time frame). Similarly the area under the lower curve represents the expected number of years with good heart function (NYHA I or II), and the area between the upper and lower curves indicates the expected time with poor heart function (NYHA III or IV). For each clinician we calculated these areas for both of the options they considered.* For some analyses the time scale was divided in two: “medium term” outcome (0 to 10 years) and “long term” outcome (10 to 20 years). Thus for the 0–20 year time frame, each participant provided subjective estimates of the years that the patient would be expected to be alive, alive and in NYHA I/II, and alive but in NYHA III/IV for one stage TCPC and an “alternative option”.

To use these data as predictors in subsequent analyses, from the pair of multistate survival graphs provided by each clinician, we calculated the corresponding difference between the expected number of years alive (the difference in areas under the upper survival curve). We denote this difference as the clinician’s perception of the “advantage” of the alternative option over a one stage TCPC; again this measurement could take a positive, zero, or negative value. We then examined the relation between each clinician’s “preference” and the corresponding “advantage” for the pair of treatment options they considered. Moreover, from the areas in different sections of the grid bounded by the two survival curves (alive in NYHA III/IV, alive in NYHA I/II), and the 0–10 and 10–20 year time ranges, eight separate “advantage” measures were calculated, and their relation to “preference” was explored. Specifically, we used correlation and multiple regression to explore the extent to which option “preference” (positive or negative) was related to the corresponding “advantage” (positive or negative) in the expected number of years in each disease state associated with choosing the alternative option over the one stage TCPC.

RESULTS
Fifty of a total of 80 doctors provided a complete set of option ratings and two subjective multistate survival graphs with both lines added. The remaining participants provided a partial set of option ratings or drew one rather than two lines on the grid provided for one or both of the options they considered. Therefore the data from these 30 participants were included in some, but not all, analyses.

*Subjective estimates of the percentages in the three states were read from the graphs to the nearest 1% with the aid of a grid on a transparent film at 30 days (the left vertical axis) and then at 2.5 year intervals from 2.5 to 20 years, giving measures at nine time points. To take account of perceived operative mortality or change in functional status associated with TCPC completion, separate readings were taken before and after TCPC completion (as inferred by the presence of a vertical or near vertical line at 5 years).
Favoured treatments
Seventy eight participants completed ratings for every option. The two stage total cavo-pulmonary connection (TCPC) was the highest rated option, being the most favourably rated of the four options for 57% of the participants; however, each of the four options featured as the first choice of some clinicians (table 1). The standard deviation of the ratings indicates a considerable spread in how favourably the one stage TCPC option was viewed.

Judgements of the probability of outcome states
Figure 2 shows graphs representing the mean values of outcome likelihoods for the four options obtained from all doctors. There was considerable variation in how doctors assessed the likely outcome of all four options, as illustrated in fig 3. The greatest variation was in the assessment of the chance of medium term survival without intervention and the long term prospects following a bi-directional Glenn.

Relation between option preference and judgements concerning likely outcome
We examined the correlation between the expected actuarial survival in various states given by doctors and their preferences for the alternative option over a one stage TCPC (table 2, final column). Positive preference for a treatment was significantly related to survival advantage (additional years alive), more so in the long term (10–20 years) than in the medium term (0–10 years). Multiple linear regression confirmed that long term survival advantage is more closely related to preference than medium term survival advantage (table 2, analysis 1). The (unstandardised) regression coefficients \( b \) values can be used to infer estimates of the relative “value” placed on each state of health for the purpose of making (or justifying) the ratings provided in this exercise. These coefficients imply that each year of life in the latter half of our 20 year time frame is “worth” about two years in the earlier half.\(^\dagger\) Also, preference was more closely associated with survival advantage in NYHA I or II than with either survival advantage in NYHA III or IV or with survival advantage per se (for which \( r = 0.57 \)). Regression coefficients obtained from a further multiple linear regression imply that a year in NYHA I or II carries the same value as approximately two years in NYHA III or IV in predicting preference (table 2, analysis 2). Combining these factors (time and heart function), it can be seen that the advantage measure for later survival in NYHA I or II is the one most closely associated with preference (\( r = 0.66 \)). Two multiple regressions indicated that once the association between this advantage measure and preference are taken into account, other advantage measures are not significantly associated with preference (table 2, analyses 3 and 4). Thus, once clinicians’ expectations about survival in NYHA I or II in the long term are known, knowing their expectations about other outcomes (for example, early death) does not significantly improve our ability to account for their preference.

In a subsidiary analysis, we noted that those doctors assigned to condition B could be separated into those preferring the one stage TCPC to the definitive bi-directional Glenn (eight doctors), and those with the reverse preference (15 doctors). Both groups viewed the likely outcome for TCPC similarly, but differed quite notably in their perception of the likely outcome for a bi-directional Glenn. Those preferring TCPC felt it offered a better chance of long term survival than a bi-directional Glenn (57% versus 49% at 20 years) as well as a greater chance of good heart function (36% versus 23% at 20 years). The reverse was true for those preferring a Glenn, who felt it offered better long term prospects than a TCPC, both for survival (67% versus 56% at 20 years) and good heart function (46% versus 30% at 20 years). Thus it would seem that preferences are matched to beliefs about likely outcomes, particularly the perceived long term prospects.

\(^\dagger\)Some caution should be exercised here, as estimates of the chance of survival beyond 20 years were not elicited. Prospects beyond 20 years are likely to correlate more strongly with long term outcomes than with medium term outcomes. If doctors take outcomes beyond 20 years into account this would then inflate the apparent value of survival between 10 and 20 years. This does not invalidate the assertion that longer term outcomes are highly weighted, but does influence the estimate of relative value.
Table 2  Multiple linear regression and Pearson correlation for the relation between each measure of “advantage” and the “preference” for the alternative option over one stage TCPC (the dependent variable, DV)

<table>
<thead>
<tr>
<th>Analysis: “main effect” of ...</th>
<th>R² for analysis</th>
<th>Predictor variables (IVs); difference in expected years</th>
<th>b (regression weight)</th>
<th>β (standardised weight)</th>
<th>Significance (p value for b/β)</th>
<th>Pearson’s r between IV and DV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Time</td>
<td>0.323</td>
<td>Medium term survival (0–10 years)</td>
<td>0.66</td>
<td>0.13</td>
<td>0.361</td>
<td>0.45**</td>
</tr>
<tr>
<td>2. Functional status</td>
<td>0.426</td>
<td>Long term survival (10–20 years)</td>
<td>1.27</td>
<td>0.47</td>
<td>0.001</td>
<td>0.56**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In NYHA class I or II</td>
<td>1.15</td>
<td>0.81</td>
<td>&lt;0.0005</td>
<td>0.63**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In NYHA class III or IV</td>
<td>0.58</td>
<td>0.24</td>
<td>0.178</td>
<td>-0.36*</td>
</tr>
<tr>
<td>3. Time and functional status</td>
<td>0.460</td>
<td>In NYHA I/II medium term</td>
<td>0.87</td>
<td>0.32</td>
<td>0.342</td>
<td>0.50**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In NYHA I/II long term</td>
<td>1.59</td>
<td>0.63</td>
<td>0.002</td>
<td>0.66**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In NYHA II/IV medium term</td>
<td>0.93</td>
<td>0.27</td>
<td>0.348</td>
<td>-0.37*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In NYHA III/IV long term</td>
<td>0.35</td>
<td>0.08</td>
<td>0.595</td>
<td>-0.15</td>
</tr>
<tr>
<td>4. Time and outcome (“best” and “worst” outcomes)</td>
<td>0.460</td>
<td>In NYHA I/II medium term</td>
<td>-0.06</td>
<td>-0.02</td>
<td>0.907</td>
<td>0.50**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In NYHA I/II long term</td>
<td>1.25</td>
<td>-0.50</td>
<td>0.027</td>
<td>0.66**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dead in the medium term</td>
<td>-0.93</td>
<td>-0.16</td>
<td>0.348</td>
<td>-0.45**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dead in the long term</td>
<td>-0.35</td>
<td>-0.13</td>
<td>0.595</td>
<td>-0.56**</td>
</tr>
</tbody>
</table>

For Pearson’s r, *p<0.05, **p<0.01.

Other influences on preference

The relation between an individual doctor’s treatment preferences and his or her own institution’s pattern of practice was examined using multiple regression (to allow adjustment for any differences attributable to specialty or experience). Doctors from institutions performing greater numbers of one stage TCPCs rated the one stage TCPC more favourably (p = 0.018) and the definitive bi-directional Glenn less favourably (p = 0.013). In contrast, those from institutions performing greater numbers of two stage TCPC completions made lower ratings for the one stage TCPC (p = 0.022) and higher ratings for the bi-directional Glenn (p = 0.031). These are not surprising findings but do reflect concordance between the participants’ reactions to this exercise and their clinical practice.

DISCUSSION

Many areas of paediatric practice are characterised by a lack of information about the long term outcomes of current treatment. This is particularly true when a condition is so uncommon that no centre will follow enough cases to provide a useful description of the issues, or when the field is so fast moving that the outcomes of management 10 or 20 years ago—even if available—are not thought very helpful in formulating beliefs about the late outcomes of current treatment. This study documents enormous variation in beliefs about the late outcomes of current practice.

The relation between an individual doctor’s treatment preferences and his or her own institution’s pattern of practice was examined using multiple regression (to allow adjustment for any differences attributable to specialty or experience). Doctors from institutions performing greater numbers of one stage TCPCs rated the one stage TCPC more favourably (p = 0.018) and the definitive bi-directional Glenn less favourably (p = 0.013). In contrast, those from institutions performing greater numbers of two stage TCPC completions made lower ratings for the one stage TCPC (p = 0.022) and higher ratings for the bi-directional Glenn (p = 0.031). These are not surprising findings but do reflect concordance between the participants’ reactions to this exercise and their clinical practice.

This study follows a long line of research which has used “paper patients” to investigate clinical judgement and decision making. Prior research in this tradition has shown that doctors’ clinical judgements based on written case summaries can be well matched to the equivalent judgements based on physical examination and patient histories obtained face-to-face. Subjective multistate survival graphs proved to be a practical and effective method of making explicit the implicit beliefs that a large number of senior doctors held about the likely outcomes of different surgical options for a paediatric cardiac patient. This adjunct to critical clinical thinking allowed us to identify the areas of greatest professional debate in the application of clinical knowledge to this case. The rating of options was found to be consistent with local practice, which can be viewed as indicative of the face validity of our approach. However, by focusing on a single case we were limited in the extent to which we could examine the impact of differences between doctors in the perceived utility of different outcomes. Also, time constraints only allowed us to explore in detail two management options with each participant.

The extent of the variation in beliefs about the likelihood of outcomes may be surprising, but can be viewed as indicative of considerable variance in the published literature. Doctors’ estimates of the likelihood of early mortality following one stage TCPC for this patient ranged from 2% to 35%. Published outcomes for patients with “high risk” TCPCs exist and include a series with one early death in 68 such operations (1.5%), but series of patients (like the one presented in our study) with a Damus-Kaye-Stansel anastomosis are small. However, from larger series using risk adjustment indicators such as younger age and previous PA banding, operative risk exceeding 30% is not implausible. Similarly, the wide spread of estimates of early mortality following a bi-directional Glenn does not exceed what can plausibly be sustained by reference to published series, even without consideration of variation between centres not reflected in the literature.

Publications detailing the outcomes of right heart bypass operations and the bi-directional Glenn beyond 10 years are limited; and the 15 and 20 year outcomes for contemporary modifications of these operations, which were first performed in the late 1980s, are of course unknown. Nevertheless, the demands of clinical decision making necessitate the formation of an opinion concerning the long term outlook for these palliations, regardless of whether such opinions are implicit or made explicit as we have done in this exercise. This study indicates that beliefs about the likelihood of good heart function in the long term appear particularly important in determining preference for different procedures. Yet this is the outcome about which there is least evidence and greatest uncertainty.

The finding that long term outcomes appear more significant than early outcomes in determining preference is consistent with the common practice in health economics and decision analysis of discounting future life years (or quality adjusted life years). The reversed pattern observed in this study might reflect a complex societal attitude relating to the relative “values” of life in childhood and adulthood. It might possibly be a feature of “proxy decision making”, in which agents other

www.archdischild.com
than the patient himself are crucial. Alternatively, if this reversal represents a difference in the values of parents and doctors, it could be a source of conflict in decision making. Though it would be difficult to undertake an exactly parallel study with a group of parents, some further examination of the issues behind this finding are probably necessary.

For the decision-theoretical element of our analysis we have made the, perhaps unwarranted, assumption that there is a rationale behind the doctors’ choices that we can attempt to infer from the profiles of preference and risk assessments they offer. However, it is important to consider how management habits might influence preference and beliefs about outcomes, because, in principle, management recommendations might so strongly reflect local precedent that it reduces to “rule based” behaviour that essentially precludes reasoning about risk. Ratings for the different surgical options for the patient we presented were positively correlated with the frequency of these options in participants’ institutions; option ratings were in turn related to beliefs about the likely outcomes of these operations. Because a potentially circular relation exists between practice and belief, we cannot establish which is dominant from this exercise.

Some psychological accounts of decision making point to habitual preferences or dispositions that determine how an individual evaluates competing options where outcomes are uncertain. Thus, rather than comparing all possible risks and benefits, people consider a subset of potential outcomes and their attendant probabilities. People motivated by “safety” or “security” primarily consider the likelihood of the worst and close-to-worst outcomes, whereas the likelihood of the best and close-to-best outcomes are mainly considered by those who are motivated by “opportunity” or “potential”. Choice is then determined by selecting the option that minimises the chance of a bad outcome, or maximises the chance of a good outcome, depending on one’s typical perspective. The clinical details of the case chosen for this study meant that the option everyone would agree on (a safe operation with great late benefit) was not available. Instead, we were able to show that for the clinicians, preference was more consistent with selecting those that minimise the likelihood of a poor outcome. From this perspective, the doctors in this study appear to be oriented towards “opportunity” or “potential” rather than “safety” or “security”.

ACKNOWLEDGEMENTS

We thank Charles Vincent and Nigel Harvey for their comments on an earlier draft of this paper. At the time this work was undertaken, the first author was supported by a grant from the UK Medical Research Council and was affiliated to the Clinical Risk Unit, Department of Psychology, University College London.

REFERENCES