

# BMJ Open Eliciting patient preferences, priorities and trade-offs for outcomes following kidney transplantation: a pilot best-worst scaling survey

Martin Howell,<sup>1,2</sup> Germaine Wong,<sup>1,2,3</sup> John Rose,<sup>4</sup> Allison Tong,<sup>1,2</sup> Jonathan C Craig,<sup>1,2</sup> Kirsten Howard<sup>2</sup>

**To cite:** Howell M, Wong G, Rose J, *et al*. Eliciting patient preferences, priorities and trade-offs for outcomes following kidney transplantation: a pilot best-worst scaling survey. *BMJ Open* 2016;**6**:e008163. doi:10.1136/bmjopen-2015-008163

► Prepublication history for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2015-008163>).

Received 11 March 2015  
Revised 2 November 2015  
Accepted 5 January 2016



CrossMark

For numbered affiliations see end of article.

## Correspondence to

Dr Martin Howell;  
martin.howell@health.nsw.gov.au

## ABSTRACT

**Objectives:** Eliciting preferences and trade-offs that patients may make to achieve important outcomes, can assist in developing patient-centred research and care. The pilot study aimed to test the feasibility of a case 2 best-worst scaling survey (BWS) to elicit recipient with kidney transplantation preferences after transplantation.

**Design:** Preferences for graft survival and dying, cancer, cardiovascular disease, diabetes, infection and side effects (gastrointestinal, weight-gain and appearance) were assessed in recipients with transplantation using a BWS (20 scenarios of nine outcomes). Participants chose 'best' and 'worst' outcomes. Responses were analysed using a multinomial logit model. Selected participants were interviewed.

**Outcomes:** Attribute coefficients and survey completion error rates.

**Results:** 81 recipients with transplantation were approached, and 39 (48%), mean age 50.5 years, completed the BWS. 4 (10%) surveys were invalid with major errors and of 35 remaining, 7 of 1400 (0.5%) choices were missing. -23 (59%) took >20 min to complete the survey. 1 was unable to finish, and 1 did not understand the survey. 2 (5%) found it very hard and 14 (35%) moderately hard. Most attribute coefficients were significant ( $p < 0.05$ ) and showed face validity. Graft survival was most important with normalised coefficients from 1 (95% CI 0.89 to 1.11) to 0.06 (95% CI -0.03 to 0.16) for 30 and 1 year duration, respectively. Attribute level coefficients decreased with increasing risk of adverse outcomes. Error rates of 20% and 2% were estimated for dominant attributes '100% risk of dying' and '30 years graft survival', respectively. 7 participants were interviewed regarding counterintuitive selection of '100% risk of dying' as a 'best' outcome. Misunderstanding, not linking dying to graft survival and aversion to dialysis were reasons given.

**Conclusions:** Recipients with transplant recipients successfully completed a complex case 2 BWS with attribute coefficients having face validity with respect to duration of graft survival and risk of adverse outcomes. Areas for refinement to reduce complexity in design have been identified.

## Strengths and limitations of this study

- Ability to elicit preferences and trade-offs for multiple outcomes of varying severity in a way that minimises cognitive burden and error rates.
- Survey design that reflects the complexity of treatment outcomes after kidney transplantation.
- Interviews of selected participants to explore reasons for counterintuitive responses.
- Bias to English speaking recipients with transplantation and with a higher level of education.
- The importance of graft survival may be overestimated due to use of 30 years graft survival in the best-worst scaling survey.

## INTRODUCTION

Compared with dialysis, kidney transplantation offers improved survival and quality of life in most patients with end-stage kidney disease, but lifelong immunosuppression is required to maintain graft function. Immunosuppression is not without harms and may lead to bacterial/fungal or viral infections, post-transplant diabetes mellitus and malignancy.<sup>1-3</sup> Prior research has quantified the frequency and severity of adverse effects associated with long-term immunosuppression,<sup>4-7</sup> and there is a body of evidence showing that recipients with kidney transplantation have a strong focus on graft survival, an aversion to returning to dialysis, and a willingness to accept side effects and adverse outcomes as being a necessary part of the treatment.<sup>8-12</sup> However, no studies have quantified the trade-offs that patients may be willing to make to reduce the impact of debilitating side effects or adverse events such as cancer through minimisation or withdrawal of immunosuppression and the risk of graft dysfunction. To date, as with many chronic diseases, treatment decisions have been predominantly driven by clinicians with

little patient involvement,<sup>13–17</sup> thus, understanding patient preferences and trade-offs is key to facilitating communication and shared decisions.

Understanding patient preferences and values are important to effective communication and facilitating informed and shared decisions that recognise the variation in tolerance of risk.<sup>18–20</sup> While preferences reflect an individual's tastes (likes or dislikes) and ideals (personal values and commitments), they will also reflect their reference point (such as having experienced dialysis or not), psychological traits (imaginative capacity, optimism, inertia), social influences (norms and laws, influence of family and friends) and beliefs (understanding of the likelihood and consequences of an outcome).<sup>21–22</sup> In the context of a complex and long-term treatment that involves trade-off between potentially debilitating side effects and adverse outcomes of immunosuppression and graft dysfunction, clinicians can have a substantial influence on patient preferences through the content and style of communication.<sup>18</sup> Understanding the extent to which patient preferences are underpinned by personal beliefs and how they vary with factors such as age, gender, and previous and current health states, provides a basis for identifying potentially erroneous beliefs<sup>11–23</sup> and the development of communication strategies best suited to facilitate construction of preferences that align with the transplant recipient's own values.<sup>24–25</sup>

The aim of this pilot study was to evaluate the feasibility of best–worst scaling (BWS), a type of discrete choice experiment,<sup>26</sup> to elicit preferences, priorities and trade-offs between outcomes following transplantation of patients with kidney transplants. By contrast with conventional discrete choice experiments, some authors have suggested that BWS offers a number of potential advantages relevant to the assessment of complex treatment outcomes<sup>27</sup> including statistical efficiency,<sup>28–29</sup> the ability to estimate attribute-level coefficients on a common scale allowing for direct comparison both between and within attributes, and may also be less cognitively demanding than conventional discrete choice or ranking experiments.<sup>30</sup> For these reasons, there has been growing interest in the use of BWS surveys in health-related research.<sup>27</sup> However, the application of BWS surveys to complex treatment regimens, such as transplant immunosuppression, has not been undertaken previously.

## METHODS

### Participant selection

Adult patients with kidney transplants (aged  $\geq 18$  years) attending a single transplant centre in Sydney, Australia, were eligible to participate. Respondents could complete the survey while attending the outpatient clinic, by phone or by email. The survey was in English, and non-English speaking patients were excluded.

### BWS methodology

Best–worst attribute scaling, or 'Case 2' BWS<sup>26–27</sup> methodology was used. A 'Case 2' BWS requires participants to choose the best and worst attribute from a single multiattribute profile.<sup>26</sup> Preferences for attributes are inferred from the choices made within the multiattribute profiles rather than from choices made between multiattribute profiles as is the case for traditional discrete choice experiments. Unlike discrete choice experiments, a BWS provides estimates of attribute-level coefficients on a common scale allowing direct comparison of attribute levels within and between attributes. Utility functions are constructed and estimated in accordance with random utility theory as for discrete choice experiments, however, with a BWS, the utility functions are constructed for individual attributes rather than for profiles.<sup>27</sup>

On the basis of the findings of an earlier qualitative study,<sup>10</sup> attributes considered most likely to be of importance to patients with kidney transplants were included. To minimise complexity, nine attributes were presented; dying with a functioning graft, malignancy, cardiovascular disease, diabetes, infection, excessive weight gain, gastrointestinal side effects, changes to appearance and graft survival. A clinically realistic range of probabilities was provided for the nine outcomes.<sup>1–3–31</sup> The options for graft survival were expressed in years. The attributes and attribute levels are shown in [table 1](#).

### Survey design and administration

A d-efficient survey design was generated with parameter values chosen to reflect expected direction of patient preference.<sup>32</sup> The survey consisted of 20 multiattribute profiles, and was presented on paper or an online form. Attribute levels that were expressed as risk of occurrence were represented by words and numbers, and through using pictograms to express the probability. Graft survival was presented descriptively and graphically using a horizontal bar graph showing years of graft survival ([figure 1](#)). At the end of the survey respondents were asked whether they ignored any attributes, how long it took to complete the survey and how hard they found the survey to complete. An open text box inviting any comments was included at the end of the survey. The survey also included questions on demographic and relevant medical details including medication, number of transplants and the time since last transplant. A paper version of the BWS was either handed out or mailed, or a link was emailed to the online version for those who agreed to complete the survey. The surveys were self-completed without assistance from researchers.

### Analysis

All completed surveys were included for analysis with the exception of those containing major errors where it was not possible to identify the best and/or the worst selection. Attribute-specific constants and attribute-level coefficients were estimated using a multinomial logit (MNL) model

**Table 1** Best–worst scaling survey attributes and levels

Attribute	Description	Attribute levels
Graft survival, years	The length of time that the transplanted organ could last	1, 5, 10, 20, 30
Serious adverse outcomes, %		
Length of life	Chance of dying before the graft fails	0, 25, 50, 75, 100
Cancer	Chance of cancer other than skin cancer	0, 10, 20, 30, 50
Cardiovascular disease	Chance of serious heart disease	0, 10, 20, 30, 50
Diabetes	Chance of diabetes after transplant	0, 10, 20, 30, 50
Infection	Chance of serious infection	0, 10, 20, 30, 50
Potentially debilitating side effects, %		
Weight gain	Chance of excessive weight gain	0, 25, 50, 75, 100
Cosmetic side effects	Chance of side effects that will change appearance	0, 25, 50, 75, 100
Gastrointestinal side effects	Chance of severe diarrhoea or nausea	0, 25, 50, 75, 100

with NLOGIT V.5.0 software (<http://www.limdep.com>). For the purpose of the choice analysis, it was assumed that the ‘best’ and ‘worst’ choices were selected sequentially, with the ‘best’ selected first, and that attribute then not available to be selected for the ‘worst’ choice. This assumes that all nine attributes were available for the ‘best’ choice, and that the ‘worst’ choice was made from only eight attributes. The attribute-specific constant for ‘risk of change in appearance’ and the attribute-level coefficient for ‘a 100% risk of change in appearance’ were normalised to 0. The attribute-specific constants thus represent the average of the unobserved effects for the attributes relative to ‘appearance’.<sup>29</sup> The variables were coded such that a ‘best’ choice for an attribute level was assigned a value of 1, a worst choice a value of –1 and a value of 0 assigned when the attribute level was not selected either as a ‘best’ or a ‘worst’. Attribute-level coefficients were estimated from the combined data set of the ‘best’ and the ‘worst’ choices, and assumes that a best choice mirrors a worst choice, and that there are no positive or negative framing effects.<sup>29</sup> As the attribute-level coefficients have the same underlying scale, the coefficients have been normalised to a range 0–1 based on the highest and lowest values from the MNL model. A value of 0 means that attribute level was least preferred and a value of 1 most preferred. All coefficients are expressed as an average with a 95% CI.

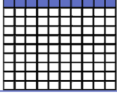
### Respondent interviews

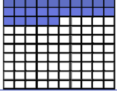
As an additional check of the validity of the BWS attributes and results, a subsample of respondents (n=7) were interviewed by phone to assess the understanding

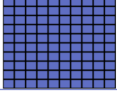
Which is the **best**? Which is the **worst**?

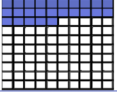
**Health Outcomes**

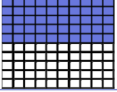
Tick only 1 box Tick only 1 box

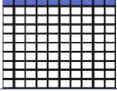
10 out of 100 people will contract a **serious infection** 

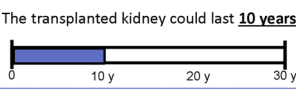
25 out of 100 people will gain a **large amount of weight** 

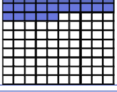
100 out of 100 people will have **severe diarrhoea or nausea** 

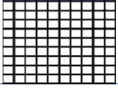
25 out of 100 people will have side effects that **change their appearance** 

50 out of 100 people will develop **serious heart disease** 

10 out of 100 people will develop **diabetes** 

The transplanted kidney could last **10 years** 

25 out of 100 people will **die** before the transplanted kidney fails 

10 out of 100 people will develop a **cancer** other than skin cancer 

**Figure 1** Example of a single scenario from the best–worst scaling survey.

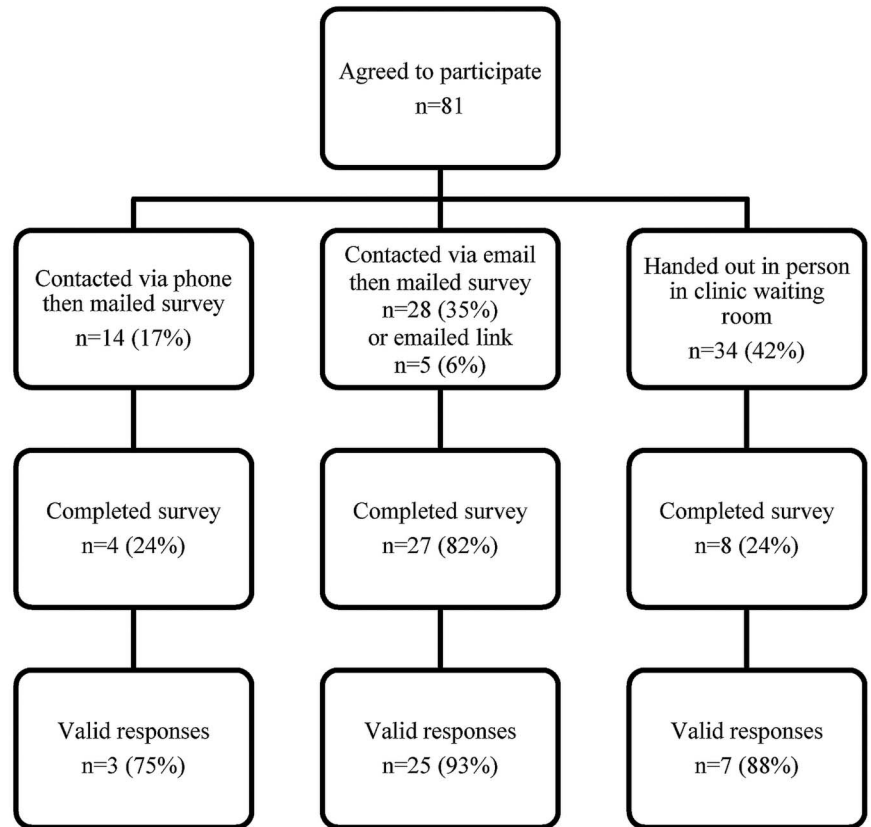
of concepts presented in the pilot survey. The questions focused on the meaning of the attribute of dying with a functioning graft. Individuals were selected on the basis of choices where responses appeared to be counter intuitive, for example, those who choose a very high risk of dying as being the best attribute were interviewed. Interviews were recorded and transcribed.

## RESULTS

### Participant characteristics

A total of 81 patients agreed to complete the BWS and 39 (48%) completed surveys were returned ([figure 2](#)). The respondents, who were aged between 19 and 70 years (mean 50.5, SD 9.7), were predominantly men (66%) and spoke English as a first language (87%), with a majority (69%) having completed education beyond high school. The time since the last transplant ranged from 0.2 to 21.7 years (median 4.7) with 44% having received grafts from deceased donors. Twenty-eight (72%) of the respondents were currently taking tacrolimus and 35 (90%) prednisone. A summary of respondent characteristics is presented in [table 2](#).

**Figure 2** Flow of patients in best–worst scaling survey. A valid survey was one in which single ‘best’ and ‘worst’ choices were clearly indicated by the participant. A valid survey may include individual scenarios that were not completed, or with an error making the selection unclear, and thus have fewer than the 20 valid scenarios available for analysis.



### Supplementary questions

A summary of responses to the supplementary questions is provided in [figure 3](#). The majority of the respondents (n=23 (59%)) reported that they took more than 20 min to complete the survey ([figure 3](#)). One respondent was unable to finish the questionnaire, and one respondent reported that they did not understand the survey, while 2 (5%) and 14 (35%) found it very hard, or moderately hard, respectively.

Twenty-three respondents (59%) stated that they ignored one or more outcomes when completing the BWS. The most frequently ignored outcomes were ‘gaining a large amount of weight’ and ‘change to your appearance’ both of which were ignored by 16 (40%) of the respondents, followed by ‘severe diarrhoea and nausea’ which was ignored by 7 (18%). All other outcomes (risk of cancer, cardiovascular disease (CVD), serious infection, death and graft survival) were ignored by 3 (8%) or fewer respondents ([figure 3](#)).

### BWS analysis

Of the 39 surveys returned, 35 (90%) provided valid data while the remaining four had major errors (missing or multiple choices for either best and/or worst) affecting all 20 scenarios and were excluded from further analysis ([figure 2](#)). Of the valid surveys, 7 of the 1400 (0.5%) choices were either invalid (more than one choice entered) or missing, and were entered as missing data giving a total of 1393 choices. Attribute-specific constants and attribute-level coefficients are summarised in

[table 3](#). The attribute-specific constants for ‘death’ and ‘graft survival’ were both significant ( $p < 0.001$ ) and positive which suggests that, all else being equal, respondents were more likely to choose these attributes as either a best or worst outcome. All other attribute-specific constants were not significant ( $p > 0.05$ ).

Attribute coefficients were statistically significant ( $p < 0.05$ ) for 30 (69%) of the 44 attribute levels with the highest (most preferred) value of 4.58 (95% CI 3.79 to 5.38) being for graft survival of 30 years, and the lowest (least preferred) value of  $-2.61$  (95% CI  $-3.32$  to  $-1.89$ ) for a 50% risk of cancer following transplantation. The normalised (0–1) attribute level coefficients are plotted in [figure 4](#). The overall preference for an attribute is indicated by the value of the attribute-level coefficient with 1 being most preferred and 0 being least preferred. The range of the coefficients for a specific attribute provides an indication of the relative contribution that the attribute is likely to make to the overall ‘value’ or utility of a given scenario, with the wider range indicating a greater contribution. For example, graft survival has the greatest range with an upper value of 1 (95% CI 0.89 to 1.11) for 30 years, and a lower value of 0.06 (95% CI  $-0.03$  to 0.16). All other attributes have coefficient values and upper 95% confidence limits  $< 1$ . The attribute-level coefficients for graft survival decreased with decreasing years of survival, and all attribute-level coefficients were significantly different from each other on the basis of the 95% CIs ([figure 4](#)). A 50% chance of cancer, a 50% chance of CVD, and a

**Table 2** Characteristics of participants

Characteristic	Respondents, n (%)
Age (years)	
19–40	5 (13)
41–50	14 (36)
51–60	14 (36)
61–70	6 (15)
Gender	
Male	23 (66)
Ethnicity	
Anglo-Saxon	23 (59)
Asian	6 (15)
European	5 (13)
Middle East	3 (8)
Samoan	1 (3)
Mauritian	1 (3)
First language	
English	34 (87)
Marital status	
Married/de-facto	34 (81)
Separated/divorced	3 (8)
Employment	
Single	2 (5)
Full time	24 (62)
Part time	2 (5)
Retired	6 (15)
Student	1 (3)
Not able to work	6 (15)
Highest education	
University	14 (36)
Technical college	13 (33)
High school	11 (28)
Primary school	1 (3)
Number of transplants	
1 only	32 (82)
Donor organ type	
Deceased	17 (44)
Living related	18 (46)
Living non-related	4 (10)
Type of transplant	
Kidney	35 (89)
Kidney/pancreas	4 (10)
Time since transplant (years)	
Not stated	3 (8)
0–1	4 (10)
1–3	6 (15)
3–5	11 (28)
5–10	5 (13)
10–15	5 (13)
15–20	3 (8)
More than 20	2 (5)
Immunosuppression	
Tacrolimus	28 (72)
Cyclosporine	4 (10)
Prednisone	35 (90)
Mycophenolate mofetil	25 (64)
Sirolimus	7 (18)

100% chance of debilitating gastrointestinal side effects were not significantly  $>0$  with 95% confidence limits all  $<0$ , and they thus have an equivalent preference to

1 year of graft survival (ie, least preferred). The attribute least likely to contribute to overall utility (ie, the one with the narrowest range of attribute-level coefficients) is the risk of side effects that change appearance with an upper value of 0.62 (95% CI 0.45 to 0.79) and lower value of 0.45 (95% CI 0.34 to 0.57) for a 0% and 75% chance of occurrence after transplantation, respectively, none of which were significantly different from each on the basis of the 95% confidence limits.

Within each attribute, the coefficients for the attribute levels generally moved in line with a priori expectations (figure 4). A higher risk of harm, and lower graft survival were less preferred to lower risk of harms, or higher graft survival. The only exception is ‘dying before the kidney transplant fails’ where the point values for the attribute-level coefficients show a ‘U’-shaped trend (figure 4). A 0% chance of dying, had an attribute-level coefficient of 0.65 (95% CI 0.54 to 0.77), and was greater than all other risk levels, however, attribute-level coefficients were not significantly different from each other for a 25%, 50%, 75% or 100% chance of dying with overlapping 95% CIs. This atypical trend in attribute-level coefficients for the risk of dying may reflect an indifference to an increasing risk of dying for probabilities  $<25\%$ , a rational choice made by some respondents, or the result of misinterpretation of the intended meaning of the attribute.

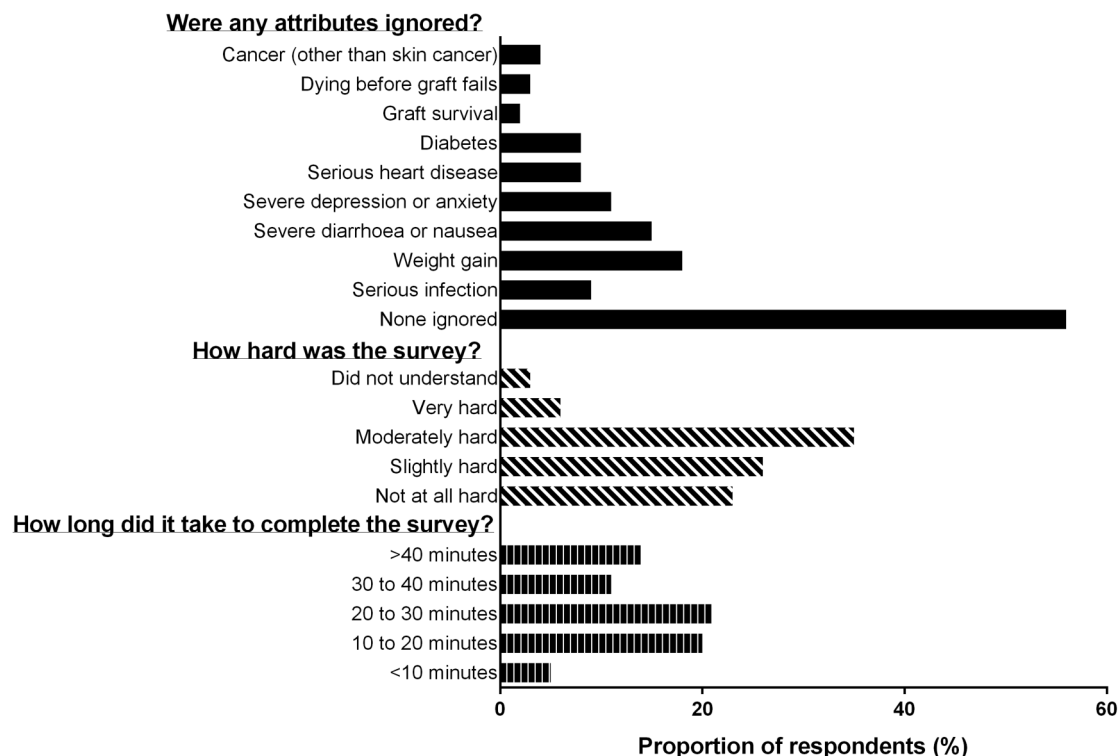
A 100% risk of dying, and 30 year graft survival, are dominant attribute levels, and a choice of ‘best’ for 100% risk of dying would be ‘irrational’, and a choice of ‘worst’ for 30 years graft survival would be ‘irrational’. The frequency of ‘irrational’ choices provides an indication of error associated with these dominant attributes. A 100% risk of dying was available for selection as a ‘best’ or a ‘worst’ choice for 138 cases and was selected as a ‘best’ choice in 28 cases (20%), which is indicative of a high error rate. By contrast, 30 years graft survival was selected as a ‘worst’ choice in 2 (2%) of 138 cases, and suggests a relatively low rate of error particularly compared to a 100% chance of dying. Reasons for the high error rate associated with the risk of dying were explored further through the interviews with selected respondents.

### Participant interviews

Seven participants were purposively selected for interview based on their selection of a high risk of dying as a ‘best’ choice. The three reasons for the selection are:

*Misunderstanding meaning:* The meaning of the attribute was interpreted differently to the intended definition. For example, the response to a question as to what was meant by 100 of 100 people will die was: ‘To me that just means that the kidney will last a hell of a long time or will outlast you’.

*Not linking the risk of dying with the length of graft survival:* The risk of dying was considered in isolation from graft survival rather than meaning a high risk of dying before the graft fails, which in the case of 1 year graft



**Figure 3** Summary of responses to supplementary questions included in the best-worst scaling survey.

survival would mean a short life span. For example: 'Why I said 75 out of 100 people will die before their kidney fails? Well as long as you don't take the questions above [the kidney could last 1 year] into account'.

*Aversion to returning to dialysis:* Dying with a functioning graft was a better option than returning to dialysis even if graft survival was relatively short for example: '...and dying from other reasons rather than having to go on dialysis wait for another one etc'.

## DISCUSSION

This pilot study identified several key findings in relation to the use of a BWS to elicit the preferences and priorities of patients with kidney transplants for outcomes following transplantation. First, even with a relatively small sample size, the majority of the attribute-level coefficients were statistically significant ( $p < 0.05$ ), and demonstrate face validity with respect to relative importance both within and between attributes. Graft survival of 30 years is more important than any other outcome and a low risk of serious adverse outcomes is also highly preferred. A low risk (0%) of serious infection, cancer and CVD has the same level of importance as approximately 20 years of graft survival. A very high risk (100%) of severe gastrointestinal side effects has the same relative importance as a high risk (50%) of cancer and CVD, and short graft survival of 1 year. By contrast, a high risk (100%) of changes to appearance is of low importance relative to all other attributes. Similarly, the range of the

attribute-level coefficients indicate that graft survival is likely to have the greatest contribution to assessment of overall utility of the outcomes after transplantation, with changes to appearance the smallest contribution.

Second, the pilot study had a low rate of major error despite the use of a self-complete paper-based format for a complex survey with 20 'best'/'worst' choice scenarios each containing nine attributes. Of the surveys returned, 10% were unusable, and of the valid surveys only 0.5% of the individual 'best' or 'worst' selections were invalid. However, the estimated error rate for the attribute of 'dying before the transplanted kidney fails' (as indicated by the rate of 'irrational' choice, assuming that a high risk of dying would be 'rationally' identified as an undesirable outcome) was in the order of 20% compared with 2% for graft survival. The interviews with respondents suggest that this error reflected misunderstanding of the meaning of the attribute and a tendency to view the risk of dying in isolation of the years of graft survival. Nonetheless, it may be a 'rational' choice for some individuals reflecting a stated preference for dying over returning to dialysis.<sup>10</sup>

Third, with the exception of the risk of dying, the attribute-level coefficients for adverse outcomes are highest (most preferred) for low risk of occurrence, and lowest (least preferred) for high risk of occurrence of the outcomes. Similarly, attribute-level coefficients for graft survival, are lowest (least preferred) for 1 year graft survival, and increase with increasing years with the maximum value (most preferred) at 30 years. In the case of the risk of dying, the trend for decreasing preferences

**Table 3** Attribute specific constants and attribute level coefficients from a multinomial logit model of the best–worst scaling survey

Attribute	Level*	$\beta$	Lower	Upper	p Value
Attribute specific constants					
Appearance		Ref	Ref	Ref	Ref
Weight		-0.04	-0.49	0.41	0.848
Gastro		-0.31	-0.81	0.20	0.234
Diabetes		-0.39	-0.89	0.11	0.124
Death		1.59	1.23	1.94	0.000
Cardiovascular disease		-0.14	-0.61	0.33	0.548
Cancer		0.06	-0.39	0.51	0.796
Infection		-0.49	-1.16	0.17	0.146
Graft		1.47	1.09	1.84	0.000
Attribute level coefficients					
Appearance	0	1.87	0.65	3.09	0.003
	25	1.43	0.56	2.30	0.001
	50	1.33	0.53	2.12	0.001
	75	0.64	-0.17	1.46	0.122
	100	Ref	Ref	Ref	Ref
Weight	0	1.57	0.51	2.62	0.004
	25	1.49	0.70	2.27	0.000
	50	0.26	-0.51	1.03	0.508
	75	-0.67	-1.47	0.13	0.100
	100	-1.39	-2.09	-0.69	0.000
Gastro	0	2.45	1.36	3.54	0.000
	25	0.44	-0.41	1.29	0.313
	50	-0.88	-1.73	-0.03	0.041
	75	-1.65	-2.49	-0.81	0.000
	100	-2.42	-3.14	-1.70	0.000
Diabetes	0	2.30	1.14	3.46	0.000
	10	1.13	0.18	2.08	0.020
	20	0.66	-0.14	1.46	0.105
	30	0.32	-0.52	1.16	0.453
	50	-1.27	-2.07	-0.46	0.002
Death	0	2.10	1.31	2.90	0.000
	25	0.28	-0.31	0.87	0.356
	50	-0.95	-1.58	-0.32	0.003
	75	-0.52	-1.17	0.12	0.108
	100	0.22	-0.38	0.81	0.472
Cardiovascular disease	0	3.00	2.08	3.96	0.000
	10	0.71	-0.07	1.50	0.075
	20	0.00	-0.80	0.78	0.987
	30	-1.57	-2.42	-0.73	0.000
	50	-2.40	-3.12	-1.70	0.000
Cancer	0	3.12	2.31	3.94	0.000
	10	0.98	0.24	1.72	0.010
	20	0.20	-0.54	0.95	0.595
	30	-0.99	-1.80	-0.17	0.018
	50	-2.61	-3.32	-1.89	0.000
Infection	0	3.36	2.38	4.34	0.000
	10	1.97	1.07	2.88	0.000
	20	2.14	1.26	3.01	0.000
	30	0.71	-0.19	1.61	0.124
	50	-1.56	-2.44	-0.68	0.001
Graft	30	4.58	3.79	5.38	0.000
	20	2.47	1.81	3.13	0.000
	10	1.11	0.52	1.71	0.000
	5	-0.43	-1.05	0.19	0.178
	1	-2.16	-2.83	-1.48	0.000

\*All attribute levels are the risk of occurrence of the outcome expressed as a percentage with the exception of graft which is expressed as years of graft survival.

with increasing risk is less apparent. As noted by de Bekker-Grob *et al*,<sup>33</sup> while most discrete choice evaluations in health assume linearity in both attributes and attribute coefficients, there are substantive reasons for this to not be the case, and that non-linearity should be taken into account when estimating trade-offs between attributes.

## LIMITATIONS

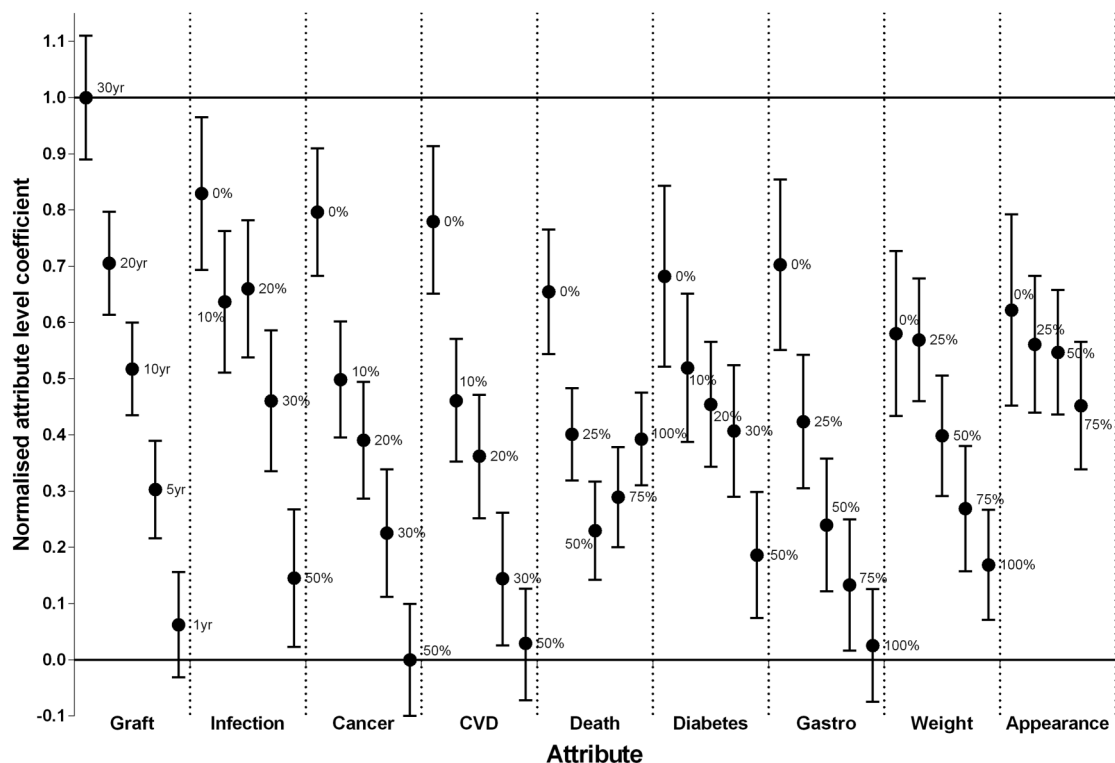
A number of observations from the pilot study warrant consideration in ongoing studies of this type. While the overall response rate for the study was close to 50%, it differed by method of approach: following contact in the clinic and by phone it was <30% compared to email contact which had a 65% response rate. The characteristics of the respondents were, in terms of gender and age, generally representative of the kidney transplant population, however, the sample is biased towards those with higher levels of education, who were white and who had English as a first language. These factors may have influenced the low error rate found in the returned surveys. The patient sample had a restricted range of immunosuppressive agents which may limit representativeness in relation to variation in experience of side effects, for example, 90% of the respondents were taking prednisone and 72% tacrolimus. However, the immunosuppression is characteristic of current clinical practice in Australia and New Zealand.<sup>34</sup>

The upper range of graft survival at 30 years was close to being a dominant attribute with almost all respondents selecting this as the best outcome when it was included in a choice scenario. This may overemphasise the relative importance of graft survival particularly given the low clinical probability of achieving 30 years. By contrast, the relative importance of the 'risk of dying with a functioning graft' may be underestimated because of participant misinterpretation and the subsequent error rate.

The analysis assumed symmetry of the 'best' and 'worst' choices, and that they come from the same underlying utility function. This follows from the approach of using a BWS as a data augmentation technique allowing for smaller sample sizes and less complex experimental designs. This assumption of symmetry may not be the case, and the 'best'/'worst' choice may be influenced by positive and negative framing.<sup>28 29</sup> Furthermore, the analysis assumes that the 'best' and 'worst' choices are made sequentially as 'best' then 'worst' for all scenarios, whereas, it may be the other way around, that is, 'worst' then 'best' for all or for some scenarios, or the choice may be made simultaneously.<sup>28</sup> The assumption of symmetry of choice and sequential selection may result in biased estimates of the attribute-level coefficients.

## IMPLICATIONS FOR FUTURE RESEARCH

The findings of the pilot BWS have implications for future research of patient preferences. The face validity



**Figure 4** Attribute-level coefficients normalised to range 0–1 relative to lowest attribute coefficient for risk of cancer of 50% and highest coefficient for graft survival of 30 years.

and the low major error rate indicate that a BWS is a feasible approach for eliciting preferences for long-term outcomes associated with complex treatment regimens and should also be applicable to conditions other than immunosuppression after transplantation. The types of situations suited to a BWS would be most applicable where treatment regimens cannot easily be described by discrete choices. In the case of recipients of kidney transplants, treatment after transplantation is more a process of adjusting the level and type of immunosuppression to balance benefits (graft survival) against harms (side effects and serious adverse outcomes), rather than a choice between clearly distinct treatment options. Describing treatment outcomes as multiattribute single scenarios with differing attribute levels, as in a BWS, is more realistic than the choice scenarios of a discrete choice experiment. The pilot study has shown that the conduct and analysis of a BWS for complex treatment regimens should carefully consider the meaning and understanding of outcomes, and this may require participant interviews or a ‘think out loud’ approach.<sup>35</sup> Finally, the complexity of the survey may limit participation and result in selection bias. Future studies should explore survey administration techniques aimed at increasing accessibility and response rates.

Specific to the ranking of importance of outcomes after kidney transplantation, the findings of the pilot study can be used to provide for more efficient designs by use of better informed priors, and development of design formats that minimise error resulting from cognitive

burden. For example, the use of a ‘best’ ‘worst’ ‘next best’ ‘next worst’ format to reduce the number of scenarios.

## CONCLUSIONS

The pilot study has indicated that patients with kidney transplants were able to complete a complex BWS of preferences for multiple long-term outcomes after transplantation and are willing to trade-off benefits and adverse outcomes. The pilot survey identified some key findings with respect to preferences and priorities recipients with kidney transplantation that have implications for patient-centred research. In particular, while graft survival is the most important outcome, the pilot study suggests that a low risk of serious adverse outcomes including potentially debilitating side effects, such as severe diarrhoea and nausea, may be of similar importance. This pilot has also helped us identify refinements needed before administration in a larger patient sample, which is now being undertaken.

### Author affiliations

<sup>1</sup>Centre for Kidney Research, The Children’s Hospital at Westmead, Westmead, New South Wales, Australia

<sup>2</sup>School of Public Health, The University of Sydney, Sydney, New South Wales, Australia

<sup>3</sup>Centre for Transplant and Renal Research, Westmead Hospital, Westmead, New South Wales, Australia

<sup>4</sup>Institute for Choice, University of South Australia Business School, North Sydney, New South Wales, Australia

**Acknowledgements** The authors would like to thank the patients with kidney transplants from Westmead Hospital, Sydney, Australia who generously gave their time and shared their opinions.



**Contributors** MH contributed to the study design, design and administration of the BWS, conducted modelling and wrote the manuscript. GW contributed to the study design, data analyses, and manuscript preparation and review. JR contributed to the study design, advised on the design of the BWS, and modelling and reviewed the manuscript. AT and JCC contributed to the study design and manuscript preparation and review. KH contributed to study design, design and administration of the BWS, data analysis and manuscript preparation and review. All authors had full access to all data and analysis.

**Funding** MH is supported by a National Health and Medical Research Council Capacity Building Grant ID 571372. AT is supported by the National Health and Medical Research Council Fellowship (ID 1037162).

**Competing interests** None declared.

**Ethics approval** Ethics approval was obtained from the Human Research Ethics Committee of the Western Sydney Local Health Network, NSW Health (HREC2009/6/4.15 (2956) AU RED09/WMEAD/56).

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** No additional data are available.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

## REFERENCES

- Webster AC, Craig JC, Simpson JM, *et al.* Identifying high risk groups and quantifying absolute risk of cancer after kidney transplantation: a cohort study of 15,183 recipients. *Am J Transplant* 2007;7:2140–51.
- Kasiske BL, Snyder JJ, Gilbertson D, *et al.* Diabetes mellitus after kidney transplantation in the United States. *Am J Transplant* 2003;3:178–85.
- Matas AJ, Gillingham KJ, Humar A, *et al.* 2202 kidney transplant recipients with 10 years of graft function: what happens next? *Am J Transplant* 2008;8:2410–19.
- Ekberg H, Kyllönen L, Madsen S, *et al.* Increased prevalence of gastrointestinal symptoms associated with impaired quality of life in renal transplant recipients. *Transplantation* 2007;83:282–9.
- Kugler C, Geyer S, Gottlieb J, *et al.* Symptom experience after solid organ transplantation. *J Psychosom Res* 2009;66:101–10.
- Moons P, Vanrenterghem Y, Van Hooff JP, *et al.* Health-related quality of life and symptom experience in tacrolimus-based regimens after renal transplantation: a multicentre study. *Transpl Int* 2003;16:653–64.
- Ponticelli C, Colombo D, Novara M, *et al.* Gastrointestinal symptoms impair quality of life in Italian renal transplant recipients but are under-recognized by physicians. *Transpl Int* 2010;23:1126–34.
- Buldukoglu K, Kulakac O, Kececioglu N, *et al.* Recipients' perceptions of their transplanted kidneys. *Transplantation* 2005;80:471–6.
- Goldade K, Sidhwani S, Patel S, *et al.* Kidney transplant patients' perceptions, beliefs, and barriers related to regular nephrology outpatient visits. *Am J Kidney Dis* 2011;57:11–20.
- Howell M, Tong A, Wong G, *et al.* Important outcomes for kidney transplant recipients: a nominal group and qualitative study. *Am J Kidney Dis* 2012;60:186–96.
- Nilsson M, Forsberg A, Bäckman L, *et al.* The perceived threat of the risk for graft rejection and health-related quality of life among organ transplant recipients. *J Clin Nurs* 2011;20:274–82.
- Prasad GVR, Nash MM, McFarlane PA, *et al.* Renal transplant recipient attitudes toward steroid use and steroid withdrawal. *Clin Transplant* 2003;17:135–9.
- Murray MA, Brunier G, Chung JO, *et al.* A systematic review of factors influencing decision-making in adults living with chronic kidney disease. *Patient Educ Couns* 2009;76:149–58.
- Fortnum D, Smolonogov T, Walker R, *et al.* 'My kidneys, my choice, decision aid': supporting shared decision making. *J Ren Care* 2015;41:81–7.
- Murray MA, Bissonnette J, Kryworuchko J, *et al.* Whose choice is it? Shared decision making in nephrology care. *Semin Dial* 2013;26:169–74.
- Ahmad N, Ellins J, Krelle H, *et al.* *Person-centred care: from ideas to action. Bringing together the evidence on shared decision making and self-management support.* London: The Health Foundation, 2014:100.
- Schoen C, Osborn R, Squires D, *et al.* New 2011 survey of patients with complex care needs in eleven countries finds that care is often poorly coordinated. *Health Affairs* 2011;30:2437–48.
- Epstein RM, Peters E. Beyond information: exploring patients' preferences. *JAMA* 2009;302:195–7.
- Kramer MHH, Bauer W, Dicker D, *et al.* The changing face of internal medicine: patient centred care. *Eur J Intern Med* 2014;25:125–7.
- Barry MJ, Edgman-Levitan S. Shared decision making—pinnacle of patient-centered care. *N Engl J Med* 2012;366:780–1.
- Hausman D. *Preferences, value, choice, and welfare.* New York: Cambridge University Press, 2012.
- McCaul KD, Peters E, Nelson W, *et al.* Linking decision-making research and cancer prevention and control: important themes. *Health Psychol* 2005;24:S106–10.
- Montori VM, Brito J, Murad M. The optimal practice of evidence-based medicine: incorporating patient preferences in practice guidelines. *JAMA* 2013;310:2503–4.
- Mühlbacher A, Juhnke C. Patient preferences versus physicians' judgement: does it make a difference in healthcare decision making? *Appl Health Econ Health Policy* 2013; 11:163–80.
- Siegel CA. Shared decision making in inflammatory bowel disease: helping patients understand the tradeoffs between treatment options. *Gut* 2012;61:459–65.
- Flynn TN. Valuing citizen and patient preferences in health: recent developments in three types of best-worst scaling. *Expert Rev Pharmacoecon Outcomes Res* 2010;10:259–67.
- Lancsar E, Louviere J, Donaldson C, *et al.* Best worst discrete choice experiments in health: methods and an application. *Soc Sci Med* 2013;76:74–82.
- Giergiczny M, Hess S, Dekker F, *et al.* Testing the consistency (or lack thereof) between choices in best-worst surveys. *Third International Choice Modelling Conference*; Sydney, Australia, 2014.
- Rose J. *Interpreting discrete choice models based on best-worst data: a matter of framing.* Paper No. 12-3103-1. Transportation Research Board 93rd Annual General Meeting. Washington DC, 2014.
- Whitty JA, Burton P, Kendall E, *et al.* Harnessing the potential to quantify public preferences for healthcare priorities through citizens' juries. *Int J Health Policy Manag* 2014;3:57–62.
- Lamb KE, Lodhi S, Meier-Kriesche HU. Long-Term Renal Allograft Survival in the United States: A Critical Reappraisal. *Am J Transplant* 2011;11:450–62.
- Bliemer MCJ, Rose JM. Construction of experimental designs for mixed logit models allowing for correlation across choice observations. *Transport Res B-Methodol* 2010;44:720–34es
- de Bekker-Grob EW, Rose JM, Bliemer MCJ. A closer look at decision and analyst error by including nonlinearities in discrete choice models: implications on willingness-to-pay estimates derived from discrete choice data in healthcare. *Pharmacoeconomics* 2013;31:1169–83.
- McDonald S, Clayton P, Hurst K. *ANZDATA Registry Thirty Fifth Annual Report.* Adelaide, South Australia, 2013.
- Whitty JA, Walker R, Golenko X, *et al.* A think aloud study comparing the validity and acceptability of discrete choice and best worst scaling methods. *PLoS ONE* 2014;9:e90635.