European Journal of Human Genetics (2013) 21, 1202–1208

www.nature.com/ejhg

ARTICLE

Eliciting preferences for priority setting in genetic testing: a pilot study comparing best-worst scaling and discrete-choice experiments

Franziska Severin*[,1], Jörg Schmidtke[2], Axel Mühlbacher[3] and Wolf H Rogowski[1,4]

Given the increasing number of genetic tests available, decisions have to be made on how to allocate limited health-care resources to them. Different criteria have been proposed to guide priority setting. However, their relative importance is unclear. Discrete-choice experiments (DCEs) and best-worst scaling experiments (BWSs) are methods used to identify and weight various criteria that influence orders of priority. This study tests whether these preference eliciting techniques can be used for prioritising genetic tests and compares the empirical findings resulting from these two approaches. Pilot DCE and BWS questionnaires were developed for the same criteria: prevalence, severity, clinical utility, alternatives to genetic testing available, infrastructure for testing and care established, and urgency of care. Interview-style experiments were carried out among different genetics professionals (mainly clinical geneticists, researchers and biologists). A total of 31 respondents completed the DCE and 26 completed the BWS experiment. Weights for the levels of the six attributes were estimated by conditional logit models. Although the results derived from the DCE and BWS experiments differed in detail, we found similar valuation patterns in the DCE and BWS experiments. The respondents attached greatest value to tests with high clinical utility (defined by the availability of treatments that reduce mortality and morbidity) and to testing for highly prevalent conditions. The findings from this study exemplify how decision makers can use quantitative preference eliciting methods to measure aggregated preferences in order to prioritise alternative clinical interventions. Further research is necessary to confirm the survey results.

European Journal of Human Genetics (2013) **21,** 1202–1208; doi:10.1038/ejhg.2013.36; published online 13 March 2013

Keywords: resource allocation; priority setting; discrete-choice experiment; best-worst scaling; genetic testing

INTRODUCTION

The use of genetic tests in Europe is expanding at a rapid pace.[1,2] Even though technical improvements are leading to decreasing laboratory costs per tested mutation, overall expenditure on genetic tests is likely to increase, also as a result of the challenges of interpreting DNA sequences. On the basis of theoretical considerations, a recent study demonstrated that, even today, funding is insufficient to provide all potentially beneficial tests to the public.[3] Faced with a large number of novel genetic tests, service providers are therefore likely to face limited budgets that do not allow them to offer all valuable tests to their patients.

Therefore, decisions may have to be made on how to allocate resources between competing tests. This is usually referred to as 'prioritisation', ie placing genetic tests into a rank order to choose the tests that are the most important to provide.[4] Up to now, there has been no practical guidance on how to prioritise genetic tests. As a result, decisions are likely to be left to the providers of genetic tests, resulting in heterogeneity within and across health-care systems.[5] In a recent study, Canadian decision makers reported that, in the absence of transparent and principle-based methodologies, decisions are often made on an ad hoc basis.[6] Ad hoc decision making imposes the risk of an unfair distribution of benefits and disadvantages.[7]

To prepare guidance for fair and more harmonised prioritisation, it is desirable to obtain evidence of the actual value judgements of key stakeholder groups (such as policy makers, funders, scientists, clinicians, industry and patients). Indeed, the inclusion of multiple perspectives may be considered the most important element in fair priority setting.[8] The aim of this study is to contribute to a more structured approach in eliciting and comparing preferences across key stakeholders in order to support the development of an explicit approach for prioritising genetic tests.

Establishing the 'value' of health-care interventions is a complex task. In the case of genetic testing, multiple benefits need to be accounted for.[9] Clinical studies may conclude that genetic tests offer no value if the screening does not change the course of treatment or improve patient outcomes. However, patients, doctors and other stakeholders may place great value on genetic testing for a variety of non-medical reasons (especially the value of knowledge for making decisions about life planning rather than to guide clinical interventions[10,11]). Indeed it has been argued that only by considering

[1]Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Health Economics and Health Care Management (IGM), Neuherberg, Germany; [2]Hannover Medical School, Institute for Human Genetics, Hanover, Germany; [3]Hochschule Neubrandenburg, IGM - Institut Gesundheitsökonomie und Medizinmanagement Neubrandenburg, Germany; [4]Ludwig-Maximilians-University, Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Clinical Center Munich, Germany

*Correspondence: F Severin, MPH, Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Health Economics and Health Care Management (IGM), Ingolstädter Landstrasse 1, Neuherberg 85764 Germany. Tel: +49 89 3187 4052; Fax: +49 89 3187 3375; E-mail: franziska.severin@helmholtz-muenchen.de

Received 30 January 2012; revised 1 February 2013; accepted 6 February 2013; published online 13 March 2013

health and non-health criteria simultaneously can the full value of genetic tests be captured.[12]

Different approaches are available when analysing multiple criteria to guide health-care decision making. In the last decade, discrete-choice experiments (DCE) have become widespread practice when analysing respondents' perspectives regarding resource allocation decisions, for example to determine the public's distribution preferences within the UK NHS[13] or to consider the implementation of lung health programmes in Nepal[14] (see De Bekker-Grob *et al*[15] for an overview). However, till now, no DCEs have been conducted to support resource allocation decisions for genetic testing. Best-worst scaling (BWS) is an alternative preference electing technique that is assumed to have a number of advantages over DCE. In particular it may impose less cognitive burden on respondents as it presents profiles one by one rather than two (or more) at the same time. However, BWS is a relatively new method, and there is a lack of evidence demonstrating the superiority of this method.

In this paper, we are interested in testing the feasibility of both the DCE and the BWS approach in the context of decision making for priority setting in genetic testing. The study is undertaken with the objective of informing the design of a larger survey on priority setting criteria. It is the first stage in a project designed to establish priority scores for different genetic testing options. Ultimately, the project seeks to integrate priority scores into the Clinical Utility Gene Cards[16] in order to assist in appraisal regarding the tests' relative importance in the case of resource scarcity.

MATERIALS AND METHODS

Priority setting criteria and levels

DCE and BWS experiments are both preference eliciting methods that analyse the relative importance of several attributes and their corresponding levels. We developed attributes for inclusion in the experiments from a qualitative review of the literature on the criteria proposed for priority setting relevant to genetic testing. We hypothesised that both the characteristics of the individuals undergoing the test and the attributes of the test itself influence the order of priority.

As there is growing awareness that incorporating patients' (or patient representatives') opinions in priority setting decisions can help to ensure the fairness and legitimacy of the decision outcomes,[17] we subsequently conducted a series of qualitative interviews with patient representatives from different countries in Europe. The overall aim of this was to: (1) assess whether characteristics identified in the literature review were confirmed as being important from the patients' point of view; and (2) establish whether patients think that there are other essential factors not mentioned in the literature. On the basis of the literature research and the results of qualitative interviews, a preliminary list of attributes was established. Next, the criteria were checked for conceptual overlap as this would hinder accurate analysis.[18]

Table 1 presents the levels that were assigned to each attribute. These levels were selected to represent the available testing options. To be confident that the levels reflected clinical scenarios and are therefore meaningful for the respondents, members of the research team identified a suitable clinical situation for each level from the published literature.

Design of the discrete-choice experiment

In the DCE, the respondents were asked to state their choice from two hypothetical alternatives. On the basis of five criteria with two levels and one criterion with three levels, 96 ($2^5 \times 3^1$) different scenarios could be defined (full factorial design) resulting in over 4500 combinations of pairwise choices. Clearly, this would be an impossible task for any person conducting the DCE. Therefore, we reduced the number of choice sets using a D-efficient design generated with SAS software (SAS Institute Inc., Cary, NC, USA, Version 9.2).[15] Following this approach, we created 12 choice tasks similar to that shown in Figure 1 (resulting in a D-efficiency of 90.57).

Each choice task involved a forced choice between the two different genetic tests presented. The forced choice design was chosen as neither opting out nor keeping the status quo appeared to be realistic decision options in clinical practice. We decided in favour of a generic presentation of the scenarios using the label 'testing option A/B' to ensure that participants base their decision on the attribute levels rather than on the disease name.[19] Findings from earlier DCEs support a generic presentation of genetic testing options.[20,21]

Additionally, a dominant choice task, in which one testing option was assumed to be superior to the other in all attributes was included in the questionnaire to test participants' understanding of the DCE format. We assumed that participants prefer diagnostic testing for highly prevalent, treatable and severe conditions where no alternative to genetic testing exists and the infrastructure for delivering the testing is widely established.

Design of the BWS experiment

The BWS experiment used the same attributes as the DCE, but the nature of the choice tasks was different. Rather than deciding which of two profiles they prefer, respondents are presented with one profile, in which they have to choose the 'most important' item (from which they derive the highest utility) as well as the 'least important' item (from which they derive the lowest utility) based on the levels displayed in a given situation. For the BWS experiment, the full factorial design resulting from all possible combinations of attribute levels was reduced using an orthogonal array resulting in 24 options. As we assumed that 24 choice tasks are still too many to evaluate for one respondent we blocked the design into two versions (using the SAS Block procedure). Figure 2 shows an example of a BWS exercise.

BWS has a potential for bias, when so called 'easy choices' are presented, ie profiles in which one attribute is at the top (bottom) whereas all other attributes are at the bottom (top) level. When such states appear this is likely to result in greater choice consistency (variance heterogeneity).[22] Therefore, we tried all possible coding schemes and chose the one that minimised this problem arising from easy choices.

Administering the questionnaire

Both experiments were conducted alongside the European Human Genetics Conference (Amsterdam, the Netherlands, 28–31 May 2011/ Nürnberg, Germany, 23–26 June 2012). Attendees were approached during breaks and asked to participate in a face-to-face interview. Only one interviewer (FS) administered the questionnaire to standardise and maintain the quality of interviews and data collection.

Respondents received either the DCE or one of the BWS questionnaires. The questionnaires were presented in three sections. The first section consisted of socio-demographic information, including gender, profession and nationality. The second section consisted of the DCE or the BWS experiment. The attributes and levels relating to the choices were described to each participant. While answering the choice tasks, we asked participants to imagine that they were clinical geneticists faced with the difficult choice of how to allocate their budget. The third section contained a series of follow-up questions regarding participants' understanding of the choice format and potential improvements in the survey instrument. During the interviews, participants were encouraged to discuss any problems arising.

Data analysis

Discrete-choice experiment. The choices observed in a DCE are assumed to reveal the true latent but unobservable utility (ie measure of individual value or benefit). As shown in Equation 1, the utility U^i_{jt} that individual i derives for each of j choice alternatives in a choice questions t can be decomposed in a determinant part V_{jt} and random error term denoted by:

$$U^i_{jt} = V^i_{jt} + \varepsilon^i_{jt} \equiv \sum_{k=1}^{K} (\beta^i_k \times X_{kjt}) + \varepsilon^i_{jt} \quad (1)$$

where β^i_k represents the marginal utility of an additional unit of attribute k, and X_{kjt} is the value of attribute k shown in alternative j, question t. ε denotes the random error term. Assuming that ε is independently and identically

Table 1 Explanation of prioritisation criteria and levels in the DCE

Criteria	*Explanation*	*Level*	*Level definition*
Prevalence of the condition within the target group	The phenotypical prevalence (genotype prevalence [a] penetrance) may differ between different conditions and within different target groups. Society may favour genetic tests for common conditions as they are more likely to detect an affected person.	Less than 0.05%	In line with the definition of the European Commission of Public Health[a], we define a rare condition as affecting less than 0.05% of a target group.
		Higher than/equal to 0.05% but less than 25%	We define a medium frequent condition as affecting less than 25% but more than or equal to 0.05% of a target group.
		Higher than/equal to 25%	Other
Severity of condition	The severity of the conditions tested for may be unequal. There might be justification in giving priority to severely ill patients as they are in greater need of health care.	Highly severe	As Huntington disease is a frequently cited very severe genetic condition, we define 'highly severe' as a patient suffering from Huntington disease or a condition of comparable severity.
		Moderately severe	Other
Urgency of care	In the considered setting, the tests' aim may be to define a diagnosis or to predict the future risk of developing a disease. There might be a preference for conducting diagnostic testing as people with established symptoms face greater urgency for health resources.	Diagnostic	'Diagnostic' testing is defined as any genetic test that aims to establish a diagnosis in a symptomatic person.
		Predictive	Other
Clinical utility	Although some tests may have clinical utility, which could lead to a reduction in mortality/mobility (eg due to the availability of treatment or preventive options), other tests may be mainly for personal knowledge. Society may wish tests leading to mortality/morbidity reduction to have higher priority than those carried out just for personal knowledge.	Mortality/morbidity reduction	'Mortality/morbidity reduction' is the case if the patient gains health benefit through available treatment or preventive options.
		Personal knowledge	Other
Alternatives available	There might be alternative means of diagnosing or predicting genetic conditions (eg clinical, biochemical). Society may have a preference for those tests where no alternative exists for diagnosis/prediction.	Not available	Alternatives are 'not available' if there are no alternative options for establishing a diagnosis/predicting the risk of contracting a disease.
		Available	Other
Infrastructure for deliverability of testing and subsequent care	There might be a difference between competing tests in terms of availability of the infrastructure for the delivery of the test and routes for follow-up medical care (eg psychological care, medical treatment, inpatient health care). Society may have a preference for conducting those tests primarily where the infrastructure is fully established	Established	We define the infrastructure to be fully 'established' if the test is routinely offered within genetic clinics.
		Not established	Other

[a]http://ec.europa.eu/health/rare_diseases/policy/index_en.htm (downloaded 18 April 2011).

	Testing Option A	Testing Option B
Prevalence of the condition within the target group	higher than/equal to 25%	less than 0.05%
Severity of condition	highly severe	moderately severe
Urgency of care	diagnostic	predictive
Clinical utility	mortality/morbidity reduction	personal knowledge
Alternatives available	not available	available
Deliverability of testing and subsequent care	not established	established
Which one would you choose? Please tick a box	☐	☐

Figure 1 Example DCE question.

MOST IMPORTANT		LEAST IMPORTANT
	The **prevalence of the condition** within the target group is higher than **25%**	
	The condition tested for is **moderately severe**	
	The test is a **predictive test**	
	The test leads to **mortality/morbidity reduction**	
	There are **no alternatives** to genetic testing available	
	The infrastructure for delivering testing and subsequent care is **fully established**	

Figure 2 Example BWS question.

distributed (IIA) following a type 1 extreme value distribution the conditional logit regression model can be applied. The probability of observing a particular sequence of choices C in choice questions t with j alternatives is then given by:

$$\mathrm{Prob}[C_{j1}^{i},\ldots,C_{jT}^{i}]=\prod_{t=1}^{T}\left[\frac{\exp\left[\sum_{k=1}^{K}(\beta_{k}^{i}\times X_{kjt})\right]}{\sum_{j=1}^{J}\exp\left[\sum_{k=1}^{K}(\beta_{k}^{i}\times X_{kjt})\right]}\right] \quad (2)$$

Specifically, the econometric model estimate of our study has the following form:

$$\begin{aligned}U_{jt}^{i}&=\beta_{\mathrm{Prevalence\_high/medium}}*\mathrm{Prevalence}\\&+\beta_{\mathrm{Severity\_high}}*\mathrm{Severity}+\beta_{\mathrm{Urgency\_diagnostic}}*\mathrm{Urgency}\\&+\beta_{\mathrm{Utility\_mortatlity/morbidity\ reduction}}*\mathrm{Utility}+\beta_{\mathrm{Alternatives\_available}}*\mathrm{Alternatives}\\&+\beta_{\mathrm{Infrastructure\_established}}*\mathrm{Infrastructure}+\varepsilon_{jt}^{i}\end{aligned} \quad (3)$$

A main effect model was used, which assumes that all attributes have an independent effect on the choices (ie all interactions between attributes were zero) as including attribute interactions increases the requirements of the experimental design. Exclusion of interaction terms may be justified as the main effects typically account for 70–90% of explained variance in DCEs.[23]

Effect coding was used to transform the attributes' levels into L-1 dummy variables. Given the coding, we expected all β coefficients to have positive signs, indicating an increase in utility.

BWS

Although the choice task was different in the BWS experiment—respondent i always chooses the most and the least important item from a given set of items t—this choice task can still be modelled in the discrete-choice framework outlined above:[22] a utility function can be specified for each possible most/least important pair measuring the difference in utility between each pair chosen as most important and least important. Thus in each choice profile, respondents could choose between 30 most/least important pairs. According to Potoglou *et al*.[24], the utility of choosing for example 'high severity' as most important and 'personal knowledge' as least important within a given choice question t can be specified by:

$$\begin{aligned} &U(\text{severity\_high, utility\_personal knowledge})^i_{jt} \\ &= \left[\beta_{\text{severity\_high}} * (1, \text{ if severity is high; } 0 \text{ otherwise})\right] \\ &- \left[\beta_{\text{utility\_personal knowledge}} * (1, \text{ if utility is personal knowledge; } 0 \text{ otherwise})\right] \end{aligned} \quad (4)$$

Whereas it is necessary to fix one level of each attribute to avoid over specification of the model in the DCE, in BWS experiments, this need only to be done for one attribute level. In our model the preferences for each attribute level were estimated relative to infrastructure 'not established'. All estimates of the model can then be interpreted relative to the omitted level.

RESULTS

Respondents

A sample of 31 respondents completed the DCE questionnaire and 26 (13 per version) completed the BWS questionnaires. Table 2 provides an overview of the sample's socio-demographic characteristics. On average, it took participants about 25 min to complete the interview for both DCE and BWS questions.

Model estimation

Table 3 presents the estimated coefficients of the conditional logit model using data from the DCE and BWS experiments, respectively.

In the DCE, one respondent failed to answer the 'dominant choice task' (the option that was dominant to the alternative in all provided attributes) correctly. We tested the 'irrational responder', demonstrating that the exclusion did not have major effects on the results. Therefore, data from this respondent were included in the analysis as suggested by Lancsar and Louviere.[25] All coefficients resulting from the DCE analysis have a positive sign and thus have the expected direction. Staging of the attribute 'prevalence of the condition' is consistent across attribute levels (ie $\beta_{\text{Prevalence\_high}} > \beta_{\text{Prevalence\_medium}}$). Except for the level coefficient referring to Infrastructure 'established', all coefficients were significant ($P<0.05$). Coefficients in the DCE present the part worth utilities associated with changes in each attribute level compared with the reference case of this attribute. For example, a genetic test leading to mortality/morbidity reduction resulting from available treatment or preventative options has a higher level of utility than testing options that are conducted only for personal knowledge (everything else being equal).

In the BWS experiments, the preferences for each attribute level was estimated relative to Infrastructure 'not established'. As such, each coefficient can be interpreted as the utility on a common underlying preference scale.[26] Coefficients with a positive sign indicate a higher level of utility compared with the reference level, and vice versa for negative coefficients. In contrast to the data derived from DCEs the staging of the two lower attribute levels of 'prevalence of the condition' is not consistent across levels (ie $\beta_{\text{Prevalence\_medium}} < \beta_{\text{Prevalence\_low}}$). However, these parameter estimates were not statistically significant. The staging of the other attributes levels was in line with the a priori expectations. The level estimate for the availability of alternatives was also not statistically significant.

Table 2 Sample characteristics

Characteristics	*Number of people in DCE* (n = 31)	*Number of people in BWS* (n = 26)
Gender		
Male	11	17
Female	20	9
Profession		
Clinical geneticist	14	6
Molecular geneticist	2	1
Laboratory scientist	2	3
Biologist	6	4
Researcher	3	11
Regulatory affairs	2	—
Industry	1	—
Patient supporter	1	1
Nationality		
Germany	5	2
Sweden	2	1
Denmark	2	3
United Kingdom	4	2
Netherlands	5	5
Switzerland	2	—
Croatia	1	—
Belgium	4	3
Russia	2	—
Portugal	1	1
Finland	1	—
Spain	1	1
Canada	1	1
Italy	—	1
Australia	—	3
USA	—	1
Turkey	—	1
Slovenia	—	1

Comparison of the values derived from DCE and BWS

In order to compare estimates derived from the DCE and the BWS experiments, the difference in preferences within an attribute has to be examined. Therefore, we calculated the marginal value of moving from the lowest attribute level to a higher level as outlined by Potoglou *et al*.[24] As BWS and DCE have different underlying preference scales the coefficients cannot be compared directly.[27] However, the relative size can be assessed by dividing all attribute levels by a fixed attribute level and thus by scaling all levels relative to this.[24] For this division, we have chosen the 'severity' level of 'high'.

Figure 3 presents an overview of rescaled values derived from the DCE and BWS experiments. As can be seen, the attributes 'availability of treatment options' to reduce mortality and morbidity' and 'highly prevalent conditions' have been valued most highly in both experiments even though utility weights differ. Although the middle and rear parts of the rank ordering of attribute levels differ in detail, the results suggests that, in both experiments, the participants attached a higher utility to testing for severe conditions than to the non-availability of alternatives or to diagnostic testing.

Follow-up: participants' cognitive ability

Table 4 presents data on respondent self-reported difficulty in understanding and answering the choice formats. The majority of

Table 3 Results from the conditional logit models

		DCE		*BWS*	
Criteria	*Levels of criteria*	*Coefficient*	P-*value*	*Coefficient*	P-*value*
Prevalence of the condition within the target group	Higher than/equal to 25%	0.3415	0.0012	1.9237	<0.0001
	Higher than/equal to 0.05% but less than 25%	0.2730	0.0083	0.1993	0.3910
	Less than 0.05%	Reference		0.2065	0.3666
Severity of condition	Highly severe	0.3160	<0.0001	1.6419	<0.0001
	Moderately severe	Reference		0.7556	0.0002
Urgency of care	Diagnostic	0.1376	0.0450	1.4433	<0.0001
	Predictive	Reference		0.7949	0.0001
Clinical utility	Mortality/morbidity reduction	0.4863	<0.0001	3.1582	<0.0001
	Personal knowledge	Reference		-0.4986	0.0086
Alternatives available	Not available	0.3044	<0.0001	0.9687	<0.0001
	Available	Reference		0.1269	0.5172
Infrastructure for deliverability of testing and subsequent care	Established	0.1066	0.1311	1.2803	<0.0001
	Not established	Reference		Reference	

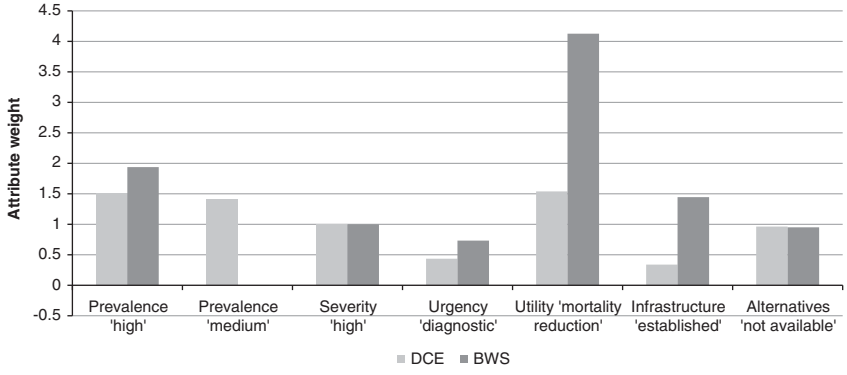

Figure 3 Comparison of rescaled DCE and BWS coefficients.

respondents to the DCE reported that they found the choice tasks not very difficult to understand. However, the respondents reported that the DCE questions were difficult to answer. In contrast, more than half the participants in the BWS exercise found the BWS tasks difficult or very difficult to understand and nearly half of the participants found the questions very difficult to answer.

On conclusion of the DCE/BWS exercise, we discussed the choice tasks with the respondents in order to learn about potential improvements in survey design. Issues covered in the follow-up questions included the attributes and levels and the choice exercise as a whole. Further information on the follow-up questions and the attribute selection are available in a Supplementary Appendix available from the corresponding author on demand.

Table 4 Difficulty in understanding and answering DCE/BWS questions

	DCE		*BWS*	
	Difficulties in understanding the choice task (%)	*Difficulties in answering the choice task (%)*	*Difficulties in understanding the choice task (%)*	*Difficulties in answering the choice task (%)*
Very difficult	9.7	19.4	15.4	46.2
Difficult	16.1	64.5	38.5	46.2
Not very difficult	74.2	16.1	46.2	7.7

DISCUSSION

To our knowledge, this is the first study testing the feasibility of the DCE and BWS formats in weighting different key considerations for priority setting in genetic testing. Given the exploratory nature of the study, the findings are not generalisable. The study is used to exemplify how quantitative preference eliciting methods can be used to informing decision makers about aggregated preferences.

The results derived from the DCE and BWS experiments indicate that the genetic experts surveyed in this study consider clinical utility (in terms of the availability of treatment and prevention options) as the most decisive criterion for prioritising genetic tests. Thus, genetic tests that lack a treatment strategy and therefore merely provide information had an increased probability of being rejected. This finding corresponds with the observation that, generally, the effectiveness of a health technology and thus improved health outcomes are seen as key criteria in priority setting decision making.[28] Most regulatory and reimbursement agencies assess the scientific evidence of a technology's effectiveness if explicit coverage decisions are made.[5]

Independent of whether or not treatment is available, the participants exhibited a very strong preference for testing for highly prevalent conditions. This finding may indicate that participants accounted for the criterion of related health need when prioritising genetic tests or that they aimed at maximising health benefits from scarce resources.

Comparing DCE and BWS

Although we found similar preference patterns, as outlined above, values derived from the DCE and BWS differed in detail. While this could reflect bias due to true differences in preferences between DCE and BWS participants, it might also result from methodological effects. Flynn[22] discussed further how the different nature of the choice task might lead to disagreement between the parameter estimates. Although Potoglou *et al.*[24] found similar preference patterns for quality of life domains in BWS and DCE experiments, the similarity of the findings derived from DCE and BWS data still remains unclear within the context of priority setting. Further research is needed to assess under which circumstances and due to what reasons the results derived from the BWS approach differ from the values derived from the DCE.

Comparing the DCE and the BWS approaches, BWS is cited as having a number of advantages because it presents respondents with profiles one by one rather than two at a time.[29] Thus, BWS has the advantage that all coefficients can be estimated on a common scale. This allows comparison of utilities across all attribute levels (and not just differences within each attribute).[30] Moreover, although all preference eliciting methods are cognitively burdensome for participants, the one profile approach is often assumed to be less cognitively demanding for participants. However, within this study, we could not find any evidence supporting this assumption in the context of priority setting. Indeed, answering BWS questions was found to be even more burdensome by respondents than answering traditional DCE questions.

In on-going research on priority setting in genetic testing, the advantages of the BWS approach must be traded off against the potential for bias that may result from an additional cognitive burden placed on respondents as well as the lack of published experiences with this novel elicitation technique. Concerning the project on prioritising genetic tests it appears more suitable to rely on results from the DCE approach as it has been applied and discussed more widely in the scientific literature and it seems to be less cognitively burdensome for respondents to answer.

Feasibility of the choice formats

The findings from this pilot study also provided insight into the feasibility and practicality of the BWS and DCE approaches. In general, the individuals we approached seemed to be involved in the experiments, which might be viewed as an important indicator of survey acceptability. During this pilot study, we found broad support regarding the necessity of the research project as such and did not encounter principally negative reactions, eg, participants opposing such research, because they believed unlimited genetic testing should be made available.

Respondents participating in the DCE experiments in general revealed fewer problems in answering and completing the questions than participants in the BWS. This may be alleviated by modifying the introduction to the BWS choice tasks. In the DCE, the staging and direction of the coefficients were consistent with a priori expectations and thus support the theoretical validity of the choice experiment.[31] The BWS experiment, in contrast, revealed irrational answerers for the two lower levels of prevalence of the condition, which might be due to bias resulting from cognitive overload.

Methodological limitations

The methodological designs used in this pilot study are not complex, as only small factorial designs are used in both experiments, and analysis was carried out using a main effect conditional logit model. This might be a methodological limitation, as the model assumes no correlation in unobserved factors over the alternatives (independence of irrelevant alternatives (IIA)). This assumption might not be realistic for genetic testing in any given situation. However, the relatively simple approach helps to keep the design straightforward and make the findings communicable to interested people who have little experience with preference eliciting methodology. Further research should use more sophisticated methods for data analysis and compare the results with the approach used here.

The findings presented here are the combined results for all participants regardless of their nationality or profession. The small sample size limits our ability to conduct separate analyses for different subgroups. Further research based on a revised version of this survey among a large number of geneticists and patient representatives is necessary to document the similarities and differences between value judgements across countries and professions. The questions of which stakeholders should be involved in the final decision and for which opinions should be given more weight if priorities differ across groups then present a new challenge.

Further ethical considerations

Empirical data about a population's value judgements can be considered a relevant input into decisions about prioritising health-care resources. However, given the lack of ethical reflection, it can hardly be claimed that they have sufficient normative power to determine decisions as important as those about providing or withholding health care. Therefore, such empirical evidence should be complemented by theoretical considerations about which criteria can be considered a reasonable basis for fair decision making.[32]

Despite the use of the best available evidence and thorough ethical reflection, reasonable people may still disagree about which criteria should have a role in decision making and about what their relative importance should be. It has therefore been claimed that decisions about limits in health care can only be legitimate if they also meet the criteria of procedural fairness. To achieve this, Daniels and Sabin[33] have proposed conditions of accountability for reasonableness, which include the recommendation that decisions and their rationales should be made transparent, that they should rest on reasons that fair-minded parties can agree and are relevant to the decision and that there is a mechanism of challenge and appeal available to the relevant stakeholders.

Such theoretical considerations as well as a criterion of procedural fairness should therefore complement the use of DCE or BWS experiments when developing guidance for prioritising genetic tests.

CONCLUSION

Scientific evidence about value judgements regarding different prioritisation criteria can provide important insights for a rational approach to priority setting. This exploratory study presents an example of how DCE and BWS experiments can be used to collect such evidence for priority setting in genetic testing.

The findings presented here show that the methods are feasible, but that they exhibit particular strengths and limitations. This underlines the importance of pilot testing the DCE and BWS format before its wider application using both qualitative and quantitative methodology. Further research needs to trade off the methodological advantages of the BWS approach against potentially biased results when using this new elicitation method.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The authors appreciate the assistance of the participants in the DCE and BWS experiments who provided their insight and time in completing this pilot study. We wish to thank Axel Rene Cyranek, Helmut Farbmacher, Matthias Hunger and Jürgen John for their methodological advice regarding the experimental design and data analysis and Christine Fischer for providing clinical expertise. This study was partly supported by EuroGentest (EuroGentest 2, Unit 2: Genetic testing as part of health care, Work package 6 (WH Rogowski)).

1 Javaher P, Kaariainen H, Kristoffersson U *et al*: EuroGentest: DNA-based testing for heritable disorders in Europe. *Community Genet* 2008; **11**: 75–120.
2 Schmidtke J, Pabst B, Nippert I: DNA-based genetic testing is rising steeply in a national health care system with open access to services: a survey of genetic test use in Germany, 1996-2002. *Genet Test* 2005; **9**: 80–84.
3 Krawczak M, Caliebe A, Croucher PJ, Schmidtke J: On the testing load incurred by cascade genetic carrier screening for Mendelian disorders: a brief report. *Genet Test* 2007; **11**: 417–419.
4 National center for priority setting in health care. Resolving health care's difficult choices — Survey of priority setting in Sweden and an analysis of principles and guidelines on priorities in health care. Linköping 2008.
5 Rogowski WH, Hartz SC, John JH: Clearing up the hazy road from bench to bedside: a framework for integrating the fourth hurdle into translational medicine. *BMC Health Serv Res* 2008; **8**: 194.
6 Adair A, Hyde-Lay R, Einsiedel E, Caulfield T: Technology assessment and resource allocation for predictive genetic testing: a study of the perspectives of Canadian genetic health care providers. *BMC Med Ethics* 2009; **10**: 6.
7 Caulfield T, Burgess MM, Williams-Jones B: Providing genetic testing through the private sector: a view from Canada. *Can Policy Res* **2**: 72–81.
8 Martin DK, Giacomini M, Singer PA: Fairness, accountability for reasonableness, and the views of priority setting decision-makers. *Health Policy* 2002; **61**: 279–290.
9 Rogowski WH, Grosse SD, John J *et al*: Points to consider in assessing and appraising predictive genetic tests. *Community Genet* 2010; **1**: 185–194.
10 Rogowski WH, Grosse SD, Khoury MJ: Challenges of translating genetic tests into clinical and public health practice. *Nature Rev Genet* 2009; **10**: 489–495.
11 Grosse SD, Wordsworth S, Payne K: Economic methods for valuing the outcomes of genetic testing: beyond cost-effectiveness analysis. *Genet Med* 2008; **10**: 648–654.
12 Mills A, Bennett B, Bloom G, Angel M, González-Block M, Pathmanathan I: *Strengthening health systems: the role and promise of policy and systems research*. Global Forum for Health Research, Alliance for Health Policy and Systems Research, Geneva: Switzerland. Available at. http://www.who.int/alliance-hpsr/resources/Strengthening_complet.pdf
13 Green C, Gerard K: Exploring the social value of health-care interventions: a stated preference discrete choice experiment. *Health Econ* 2009; **18**: 951–976.
14 Baltussen R, Ten Asbroek AH, Koolman X, Shrestha N, Bhattarai P, Niessen LW: Priority setting using multiple criteria: should a lung health programme be implemented in Nepal? *Health Policy Plan* 2007; **22**: 178–185.
15 De Bekker-Grob EW, Ryan M, Gerard K: Discrete choice experiments in health economics: a review of the literature. *Health Econ* 2010; **21**: 145–172.
16 Schmidtke J, Cassiman JJ: The EuroGentest clinical utility gene cards. *Eur J Hum Genet* 2010; **18**: 1068.
17 Daniels N, Sabin J: *Setting Limits Fairly: Can We Learn to Share Medical Resources?*. Oxford University Press, Inc., 2002.
18 Mangham LJ, Hanson K, McPake B: How to do (or not to do). Designing a discrete choice experiment for application in a low-income country. *Health Policy Plan* 2009; **24**: 151–158.
19 Viney R, Lancsar E, Louviere J: Discrete choice experiments to measure consumer preferences for health and healthcare. *Expert Rev Pharmacoecon Outcomes Res* 2002; **2**: 319–326.
20 Regier DA, Ryan M, Phimister E, Marra CA: Bayesian and classical estimation of mixed logit: an application to genetic testing. *J Health Econ* 2009; **28**: 598–610.
21 Regier DA, Friedman JM, Makela N, Ryan M, Marra CA: Valuing the benefit of diagnostic testing for genetic causes of idiopathic developmental disability: willingness to pay from families of affected children. *Clin Genet* 2009; **75**: 514–521.
22 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**: 259–267.
23 Louviere J, Hensher D, Swait J: *Stated Choice Methods: Analysis and Applications*. Cambridge University Press: Cambridge, 2000.
24 Potoglou D, Burge P, Flynn T *et al*: Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data. *Soc Sci Med* 2011; **72**: 1717–1727.
25 Lancsar E, Louviere J: Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? *Health Econ* 2006; **15**: 797–811.
26 Marley AA, Flynn TN, Louviere JJ: Probabilistic models of set-dependent and attribute-level best–worst choice. *Mathemat Psychol* 2008; **52**: 281–296.
27 Swait J, Louviere J: The role of the scale parameter in the estimation and comparison of multinomial logit models. *Marketing Res* 1993; **30**: 305–314.
28 Guindo LA, Wagner M, Baltussen R *et al*: From efficacy to equity: Literature review of decision criteria for resource allocation and healthcare decisionmaking. *Cost Eff Resour Alloc* 2012; **10**: 9.
29 Flynn TN, Louviere JJ, Peters TJ, Coast J: Best—worst scaling: what it can do for health care research and how to do it. *J Health Econ* 2007; **26**: 171–189.
30 Najafzadeh M, Lynd LD, Davis JC *et al*: Barriers to integrating personalized medicine into clinical practice: a best-worst scaling choice experiment. *Genet Med* 2012; **14**: 520–526.
31 Rottenkolber R: Discrete-Choice-Experimente zur Messung der Zahlungsbereitschaft für Gesundheitsleistungen - ein anwendungsbezogener Literaturreview. *Gesundh Ökon Qual Manag* 2011; **16**: 232–244.
32 Richardson J, McKie J: Empiricism, ethics and orthodox economic theory: what is the appropriate basis for decision-making in the health sector? *Soc Sci Med* 2005; **60**: 265–275.
33 Daniels N, Sabin J: The ethics of accountability in managed care reform. *Health Aff (Millwood)* 1998; **17**: 50–64.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)