HEDS Discussion Paper 04/05


ScHARR
School of Health and
Related Research

The University of Sheffield
ScHARR
School of Health and Related Research

Sheffield Health Economics Group

Discussion Paper Series

May 2004

Ref: 04/5

Public preferences for participation in a large DNA cohort study: a discrete choice experiment.

Rhydian Hapgood,[1] Chris McCabe,[2] Darren Shickle.[1]

1 Department of Public Health, ScHARR, University of Sheffield
2 Sheffield Health Economics Group, ScHARR, University of Sheffield

Corresponding author:

Darren Shickle

Department of Public Health

The University of Sheffield,

Regent Court, 30 Regent Street, Sheffield S1 4DA, UK

Email: d.shickle@sheffield.ac.uk

Tel 00 (44) 114 222 5454

Abstract

Objectives. To determine the general public's preferences over the design and use of UK Biobank; and the design for optimum recruitment.

Design. Discrete choice face-to-face interviews using a fractional factorial design and multinomial logit regression modelling.

Setting. 180 sampling points across 11 regions of the UK.

Participants. Members of the public.

Main outcome measures. Relative risks of people's preferences for project design and use.

Results. 34.4% of respondents were willing to take part in UK Biobank (n=1283). The most highly preferred scenario was: individual feedback from the study; consent every time new data is requested; DNA and information destruction on withdrawal; and **access to the data by the NHS and Universities but not other third parties**. The single most important attribute was access to data. If individual's insurance companies were to be given access to the data this would be the largest single impediment to recruitment to the study. Extra resources are likely to be needed to counter the reduced recruitment rate if pharmaceutical companies are allowed access to the data.

Conclusions. The general public do have clear preferences regarding the design of biobanks. Whilst designing the study to meet the most preferred scenario *may* not be practical within available resources, biobanks can use the type of information provided here to compare the costs and benefits of different study designs. The 'price'of discounting public preferences in terms of reduced recruitment should be an important part of the 'weighing' process. Pilot studies of recruitment under alternative study designs may be justified.

Introduction

The Human Genome Project (HGP) has established the complete sequence of human DNA.[1 2] Elucidation of the roles that genes play in health and disease requires large, prospective cohort studies that investigate the interactions between genes, lifestyle and the environment.[3] With this in mind, many countries are setting up national cohorts, including Iceland, Estonia, Australia and the UK.[4-7] If the public do not consent to these large DNA cohorts, then the potential of the HGP may never be realised.

UK Biobank is a DNA cohort study funded jointly by the MRC, the Wellcome Trust and the Department of Health.[7] It aims to recruit half a million volunteers aged 45-69 years, through primary care. It will take DNA, medical and lifestyle information at baseline, and is likely to follow up the cohort for 10-20 years. There are a number of unique features that warrant research. For example, DNA will be taken and stored, many of the tests to be performed on the DNA will not be known at recruitment, and insurance, biotechnology, pharmaceutical companies and the police (amongst others) may be interested in the information produced. Whilst it is not proposed that insurance companies have access, concern that they may do could have an adverse effect on recruitment.

The UK Government sees UK Biobank as "an invaluable resource for researchers seeking to establish the effects of our genes". [8] As recently acknowledge by the House of Commons Science and Technology Select Committee, understanding individual's preferences regarding the study design is a research priority.[9]

Conjoint analysis (CA) is a method for disaggregating individual's preferences in multifactorial decision making environments, rooted in Lancaster's theory of value.[10] CA presents individuals with one or more pairwise choices, and asks them to choose which scenario they prefer. A scenario is constructed from a number of attributes, where each attribute is considered to be a potentially important determinant of the decision. An attribute can have two or more levels. A scenario consists of one level on each of the attributes. An example of a pairwise choice is given in Figure 1. Whilst the interview process can be time consuming and the task cognitively demanding, the approach provides more information than Likert and ranking questionnaires.[11]

CA was developed in mathematical psychology, and has been widely used in transport and environmental economics.[11] Subsequently, it has been used in healthcare including eliciting patient preferences in the delivery of health services and doctor preferences regarding characteristics of their job.[12-14] This study used CA to examine the public's preferences for the study design of UK BioBank. However the methodological issues of consent, feedback, withdrawal and access are common problems faced by those establishing biobanks elsewhere.

Methods

Public attitudes to participation in a large DNA cohort study were measured by face-to-face conjoint interviews. Interview schedules were designed on the basis of known concerns with the UK Biobank protocol.[15 16] These concerns were used to formulate attributes of paired scenarios, each of which consisted 2-4 levels (see table 1).

The attribute levels generate 72 scenarios (the product of the number of levels for each attribute; 3x2x3x4). An orthogonal array consisting of 16 scenarios produced 120 pairwise choices. As no one individual could provide data on all the choices, the scenarios were allocated across 10 questionnaires.

Interviews were conducted with 1283 members of the public in 180 centres across 11 regions of mainland Great Britain, using stratified sampling representative of the British population. A market research company was hired to administer the questionnaires as they have a network of interviewers that facilitates nation-wide sampling. The interviewers underwent common training. The interviews were administered face-to-face in people's homes. They were preceded by general demographic questions and followed by questions commissioned by commercial clients. Each interview began with an introductory explanation of UK Biobank, and an explanation of the nature of conjoint questionnaires, in particular how the scenarios are hypothetical and how participants are asked to imagine they really have to choose between two options (see figure 2). Respondents were also asked about their age, sex, ethnic group, social grade, terminal educational age, income, lifestage, marital status, children, employment status and housing. In addition, respondents' willingness to

participate in the UK Biobank study was measured at the end of the interview using a visual analogue scale. The questionnaires were piloted on a selection of researchers for comprehension and for time taken to complete the interview. Fieldwork was conducted during the third week of January 2002. Each interviewer was given a set of around 120 addresses, selected by ACORN profiling, and had to achieve a quota of 12-15 interviews from these 120 addresses. Show cards were used to display conjoint pairs of scenarios to choose from, and likert scales where appropriate. Interviewers asked for the respondents to give their first impression as to which scenario they prefered, although a "no preference" option was allowed. Responses were entered directly onto a laptop computer.

A multinomial logit regression model clustered on the individual was constructed in STATA v7, as the discrete dependent variable had three possible outcomes with no natural ordering. A separate model was estimated for the over 45 years of age sub-sample (the population from which UK Biobank will recruit).Within sample predictive performance was assessed by the level of agreement between observed modal responses and predicted responses, with non-dominant pairs (where the model cannot be expected to predict) removed. External predictive validity was established by testing the model on another conjoint dataset constructed as part of a parallel study on UK Biobank performed in North East Derbyshire. This dataset comprised 665 responses to a postal survey of 2000 people sampled from the electoral register.

Results

The responses per pairwise choice ranged from 95 to 155 (n=1283). Thirty four per cent (n=441) were willing to take part in UK Biobank after having completed the questionnaire.

In order to exclude noise from the model (as indifference may mean true indifference between two options that are equally liked or disliked, or inability to decide because of lack of motivation or time) an incremental exclusion of respondents based upon proportion of answers being indifferent was carried out. This was performed iteratively until modelling preferring A over B gave the same model as preferring B over A, but with the opposite sign. The models became stable when those individuals who answered 9 or more questions (out of a possible 12 or 13) as indifferent were excluded. The final model included 79.4% of the initial dataset (n=1019). The excluded group was not significantly different from the final dataset for sex, ethnic group or social grade. There was a significant difference for age due to a cohort effect (percentage over 65 years in final dataset=16.5%, in the excluded group=26.4%; χ^2 excluding over 65 category, χ^2=2.18, p= 0.70, df=4).

Multinomial logit regression produced two sets of coefficients; first, those explaining respondents' choice of A over B and second, those explaining respondents' preference for indifference over choosing scenario B. The first set is presented in Table 2. A positive coefficient means that the attribute level contributes to a choice between scenarios by increasing the probability of the individual preferring the option

containing the attribute level to the reference level. When exponentiated, this translates to a relative risk (RR)>1.

The population was half as likely to choose a scenario with no feedback as a component compared to a scenario with individual feedback. There was no significant difference between preferences for feedback to GP compared to individual feedback.

Participants were 23% less likely to choose a scenario if consent was obtained only at recruitment.

Respondents did not exhibit strong preferences for the arrangements on withdrawal from the study. The most popular option was to destroy the DNA but to use the information collected. However, the RR was small, and the 95% confidence intervals overlapped with alternative options.

Preferences were much more distinct regarding access to data. The preferred option was for the NHS and Universities only to have access. This was strongly preferred to biotechnology companies and the police (RR 0.71 and 0.74 respectively); which were in turn strongly preferred to insurance companies (RR 0.39).

The coefficients for preferring A over B therefore showed that the optimum scenario for patient participation is likely to be:

- feedback direct to individual or via GP;
- consent every time new data is requested;

- DNA destroyed but information retained on withdrawing from the study or all DNA and data retained for use by the study; and
- access by NHS and Universities only to the data.

The most influential items on preferences as indicated by the relative size of the coefficients were, in descending order: insurance company access, not receiving feedback, drug and biotech access, police access and consent only at the start.

Preferences of 45yrs and over participants (i.e. the UK Biobank age group) were not significantly different (see table 2).

Table 3 shows the most and least preferred scenarios as predicted by the model. These are a product of *both* sets of coefficients and can be slightly different to when one set only is used. The 18 most popular scenarios all contained consent every time. There was some trading of access by biotechnology companies and the police (but not insurance companies) in order to retain this consent design. Consent just at the start of the study became acceptable provided that access was only given to NHS and universities, or if participants had the option to have everything destroyed on withdrawal (scenario ranks 19-28).

The least preferred scenarios contained no feedback (ranks 59-72) and access by insurance companies (ranks 67-72). Participants chose a scenario containing no feedback if offered consent every time new information is requested (ranks 44-51). There was evidence of trading of attributes, where the opportunity of including some attributes in the study design was given up in order to retain more strongly preferred

options. The highest rank achieved by a scenario containing no feedback (rank 29) involved trading of feedback from the study against choosing consent every time, destroying everything on withdrawal, and limiting access to NHS and Universities. The highest probability achieved by a scenario containing access by insurance companies (rank 36) involved trading of preferred access to the data against choosing feedback coupled with consent every time new information is requested together with the option of destroying everything on withdrawal. Respondents may have felt that these characteristics would give them a veto over access by insurance companies.

The model predicted internal and external datasets with 76% and 80% accuracy respectively, as measured by percentage agreement of observed modal responses with predicted values (with non-dominant pairs removed). The model predicted correctly 97% of the time where percentage difference between the most common and second most common observed responses was greater than 20%, but was no better at predicting than tossing a coin when the difference fell below 20%.

Discussion

Main Findings

Approximately a third of those approached are likely to be willing to take part in Biobank. This is likely to be optimised by designing the study using the most preferred scenario identified by the conjoint analysis (individual feedback with consent every time new data is requested, DNA and information destruction on withdrawing from the study, together with access by NHS and Universities only to the data). Replacing individual feedback with GP feedback may reduce costs without impacting significantly upon recruitment. Substituting destruction of DNA but retention of information is likely to increase the long term value of Biobank without damaging recruitment.

Strengths and weaknesses

Given the imminence of and investment in DNA cohort studies world-wide, then this is timely research. It could be argued that this research may not reflect real preferences. There is early evidence from research into Chlamydia testing that expressed stated preferences using a conjoint study design do reflect subsequent revealed preferences.[17] Indifference created noise which affected the stability of models and data had to be excluded to remove the noise. This was not a fatal flaw in this study as the sample size was large, but may be an issue for smaller conjoint surveys.

What does it mean in practice?

The organisation responsible for administering UK Biobank will have to weigh up the costs and benefits of designing the study based on participant preference. Not taking preferences into consideration may affect recruitment rate and consequently cost and value of biobanks. However, ceding to participant preferences may be impractical or undesirable.

People preferred some feedback (but had no preference as to whether this is direct to the individual or via the GP) but accurate and reliable individual feedback to half a million people represents a significant logistical challenge. General as well as individual feedback was previously seen as a crucial motivator to participation.[15 18] GPs have expressed a wish to not have access to UK Biobank data, to avoid patients' concerns about insurance company access.[16 18] Individual feedback is not common in studies and is unlikely to have clinical significance for this study. It may be sensible not to arouse any expectations of feedback of this nature.

Respondents preferred consent at every new data collection, but consent once for the entire period is more pragmatic and cheaper. Information about the uses to which the samples would be put, the unacceptability of mounting studies on diseases not named in the initial consent and assurance on confidentiality were seen as important in previous research. [15-16 18] GPs were very reluctant about releasing patient information without the patient providing consent on each occasion morbidity data is requested.[16] This may be especially important if the data is particularly sensitive or qualitatively different to that described at recruitment. Four out of five people in the People's Panel

thought that specific consent should be sought for each test carried out on their DNA.[18] BioBank participants will have their DNA examined on only a small number of occasions during the follow-up period. It may therefore be feasible to contact a participant on each occasion to check they are still willing to provide consent.

A single consent at the start will have significant advantages for biobanks. It will cost less, will not depend on active participation by participants and will minimise non-response bias. The impossibility of potential research subjects knowing what information biobanks would access or what DNA tests they will do in the future does present an ethical concern, especially if a disease that recruits subsequently developed is associated with stigma, such as a mental disorder or a sexual problem.

All three options for events on withdrawal were similar in their desirability to participants. Thus, given that destruction of information on withdrawal from the study is wasteful, then retention of all data should be the default design. Individuals who decide to withdraw may have strong views however.

People preferred the NHS and universities only to have access to the data. Access for insurance companies was particularly unpopular in previous research.[18] The latter has not been suggested for Biobank, but the strong opposition within this consultation suggests that this issue should be specifically addressed. Access by the police was unpopular but acceptable in certain circumstances. However, over 90 per cent of a MORI sample believed the police should have access to Biobank to enable their investigations into a murder or sexual assault.[19] This may reflect a general wish for a police database, rather than allowing police access to a medical database.

Drug and biotechnology company access was also unpopular, but acceptable under certain circumstances. Commercial access caused concern in previous research.[15;18] Income from commercial sources is likely to be actively sought. It might be argued that only these companies have the infrastructures necessary to translate the information provided by biobanks into health care interventions. Encouragingly, previous research samples recognised and accepted this arguement.[15 18]

However, recruitment is likely to be impaired to some degree by the explicit involvement of commercial organisations. If we assume that Biobank requires 500,000 participants to meet the power requirements for the epidemiological hypotheses to be tested, then the numerical impact of allowing private sector access is a requirement to approach an additional 594,015 people. Also if the preference against pharmaceutical company involvement is stronger in certain socio-economic groups, allowing access may lead to a non-representative genetic database.

Future Research

Pilot studies of recruitment are required to test out prospectively the hypotheses that have been generated by our analyses, as well as to identify any barriers to recruitment not apparent from this research.

We thank Aki Tsuchyia, Phil Shackley and Ann Morgan for advice on analysis; Lynne Hazlehurst and Susan Wallace for help in conducting the study (questionnaire administration and data entry); and Liddy Goyder and Julie Ratcliffe for helpful comments on earlier drafts of this paper.

Contributors: DS had the original idea for the study. DS, CM and RH designed and wrote the protocol submitted for funding. DS and RH designed the conjoint questionnaires and supervised data entry. CM designed the orthogonal array. RH undertook the data manipulation and analysis and primary interpretation; and drafted the paper. DS and CM advised on data analysis and contributed to the interpretation. DS is guarantor and accepts full responsibility for the study, had access to the data, and controlled the decision to publish.

Funding: The Wellcome Trust
Competing interests: RH attended two meetings of the UK Biobank Protocol Development Group as a GP representative. RH and DS are members of the Fosse Way Cohort Spoke of UK Biobank.
Ethical approval was not required for this study.

Figure 1: Formulation for pairwise choices within conjoint analysis

Which situation would you prefer? *(please tick box below)*

	Research Project A	**Research Project B**
Who should get feedback about general health information found during the initial health check?	GP	No feedback
How often would a person need to give consent?	Just at the start	Every time
What should happen when someone wants to pull out of the study?	Destroy all DNA and information	Destroy DNA but information already collected could be used
Who should be allowed to conduct research on the DNA and information?	In addition to use for research, the police would also have access	All researchers including drug and biotech companies
	Prefer A ☐	**Prefer B** ☐

Please tick one or both boxes

Table 1: Attributes and levels used to construct the conjoint scenarios.

Attribute	Level 0	Level 1	Level 2	Level 3
Who should get feedback about general health information found during the initial health check?	Individual participants	Participant's GP	No feedback	n/a
How often would a person need to give consent?	Every time new information is gathered	Just at the start of the project	n/a	n/a
What should happen when someone wants to pull out of the study?	Destroy all DNA information	Destroy DNA but information already collected could be used	Use DNA and information but no new information to be collected	n/a
Who should be allowed to conduct research on the DNA and information?	NHS and Universities	All researchers plus drug and biotech companies	All researchers plus the individual's insurance company	All researchers plus the police

Table 2: Conjoint analysis of people's preferences concerning participation in UK Biobank using multinomial logit regression (n=1019)

Prefers A over B	**First model**				**Final Model**		
Attributes	**Coefficient**	**Std. Error**	**P**	**RR (95%CI)**	**Coefficient**	**Std. error**	**RR (95%CI)**
Feedback							
Feedback to GP*	-0.0027	0.0525	0.96	1.00 (0.90-1.11)	n/a	n/a	n/a
No feedback*	-0.6964	0.0454	0.00	0.50 (0.46-0.54)	-0.6954	0.0403	0.50 (0.46-0.54)
Consent							
Just at the start**	-0.2657	0.0377	0.00	0.77 (0.71-0.83)	-0.2658	0.0378	0.77 (0.71-0.83)
Wishes on withdrawal from the study							
Destroy DNA but use information***	0.1608	0.0441	0.00	1.17 (1.08-1.28)	0.1606	0.0440	1.17 (1.08-1.28)
Use both DNA and information already collected***	0.0832	0.0458	0.07	1.09 (0.99-1.19)	0.0827	0.0458	1.09 (0.99-1.19)
Access to the data							
All researchers including drug and biotech companies****	-0.3485	0.0503	0.00	0.71 (0.64-0.78)	-0.3466	0.0504	0.71 (0.64-0.78)
All researchers plus individual's insurance company****	-0.9452	0.0583	0.00	0.39 (0.35-0.44)	-0.9431	0.0581	0.39 (0.35-0.44)
All researchers plus police****	-0.2957	0.0585	0.00	0.74 (0.66-0.83)	-0.2955	0.0583	0.74 (0.66-0.83)

Reference level is * "Feedback to the individual" ** "Consent required every time"*** "Destroy all DNA and information" ****"NHS and Universities"

n/a = not applicable (as not in final model).

Table 3. Most and least preferred scenarios for UK Biobank design as predicted by a multinomial logit regression model (n=1019)

Rank	Feedback	Consent	Withdrawal	Access
Ten most preferred				
1	Individual feedback	every time	destroy everything	NHS and universities
2	Feedback to GP	every time	destroy everything	NHS and universities
3	Individual feedback	every time	Use DNA and info	NHS and universities
4	Feedback to GP	every time	Use DNA and info	NHS and universities
5	Individual feedback	every time	destroy DNA use info	NHS and universities
6	Feedback to GP	every time	destroy DNA use info	NHS and universities
7	Individual feedback	every time	destroy everything	biotech
8	Feedback to GP	every time	destroy everything	biotech
9	Individual feedback	every time	destroy everything	police
10	Feedback to GP	every time	destroy everything	police
Ten least preferred				
63	No feedback	just at the start	Use DNA and info	biotech
64	No feedback	just at the start	destroy DNA use info	biotech
65	No feedback	just at the start	Use DNA and info	police
66	No feedback	just at the start	destroy DNA use info	police
67	No feedback	every time	destroy everything	insurance
68	No feedback	every time	destroy DNA use info	insurance
69	No feedback	every time	Use DNA and info	insurance
70	No feedback	just at the start	destroy everything	insurance
71	No feedback	just at the start	destroy DNA use info	insurance
72	No feedback	just at the start	Use DNA and info	insurance

Reference List

1. Ventner JC,.et al. *Science* 2001;**291**:1304-51.

2. International Human Genome Sequencing Consortium. *Nature* 2001;**409**:860-921.

3. Little J, Khoury MJ, Bradley L, Clyne M, Gwinn M, Lin B *et al*. The human genome project is complete. How do we develop a handle for the pump? *American Journal of Epidemiology.* 2003;**157**:667-73.

4. Chadwick R. The Icelandic database- do modern times need modern sagas? *BMJ* 1999;**319**:441.

5. Frank L. Population genetics. Estonia prepares for national DNA database. *Science* 2000;**290**:31.

6. Prof. David Weisbrot, President Australian Law Reform Commission. 20-11-2002. Personal Communication

7. http://www.biobank.ac.uk (accessed 28th July, 2003).

8. Department of Health. Our inheritance, our future: realising the potential of genetics in the NHS. London: TSO, 2003.

9. http://www.publications.parliament.uk/pa/cm200203/cmselect/cmsctech/uc132/uc13202.htm. (accessed 28th July, 2003).

10. Lancaster K. Consumer demand: a new approach. New York: Columbia University Press, 1971.

11. Ryan M, Scott DA, Reeves C, Bate A, van Teilingen ER, Russell EM *et al*. Eliciting public preferences for health care: a systematic review of techniques. *Health Technol Assess* 2001;**5**.

12. Ryan M. Using conjoint analysis to take account of patient preferences and go beyond health outcomes: an application to in vitro fertilisation. *Social Science & Medicine.* 1999; **48**:535-46.

13. Ubach C, Scott A, French F, Awramenko M, Needham G. What do hospital consultants value about their jobs? A discrete choice experiment. *BMJ* 2003;**326**:1432-5.

14. Vick S,.Scott A. Agency in health care. Examining patients' preferences for attributes. *J Health Econ* 1998;**17**:587-605.

15. Cragg, Ross, Dawson. Public perceptions of the collection of human biological samples. London: The Wellcome Trust/MRC, 2000. http://www.wellcome.ac.uk/en/1/bioenpopethpub.html. (accessed 28th July, 2003).

16. Hapgood, R., Shickle, D., and Kent, A. Consultation with primary care health professionals on the proposed UK Population Biomedical Collection. London: The Wellcome Trust/MRC, 2001. http://www.wellcome.ac.uk/en/1/bioenpopethpri.html. (accessed 28th July, 2003).

17. Watson V, Ryan M, Watson E. Testing the validity of discrete choice experiments: evidence from a chlamydia screening study. *Personal Communication.* 2003.

18. People, Science and Policy Ltd. Biobank UK; A question of trust. Report prepared for the Medical Research Council and the Wellcome Trust. London: People, Science and Policy Ltd. 2002.

19. Public attitudes to human genetic information. Report prepared by MORI on a quantitative study of the People's Panel. 2001. London, Human Genetics Commission http://www.hgc.gov.uk/business_publications_morigeneticattitudes.pdf. (accessed 28th July, 2003).