Report to the NHS Sickle Cell & Thalassaemia Screening Programme

Analytical Report on The Feasibility of Using Ethnic Questions for Risk Status Ascertainment in Antenatal Selective Screening for Sickle Cell and Thalassaemia:

The Findings of a Formal Trial of Candidate Questions

Peter J Aspinall

Senior Research Fellow (Public Health)
Centre for Health Services Studies
University of Kent,

&

The Trial Collaborative Group

November 2003

'Analytical Report….' (76 pp).
©Peter J Aspinall

ISBN: 1-904236-09-X

For copies of this report please e-mail P.J.Aspinall@kent.ac.uk

Contents

Preface and Acknowledgements		4
Executive Summary		6
1	Introduction	9
2	Aims of the trial	11
3	The design of the two candidate questions	16
4	Analytical methods	27
5.	Results of the trial	28
5.1	Participation rates & refusals	28
5.2	Randomisation	29
5.3	The time taken to ask the ethnic questions	29
5.4	Self-ascribed ethnicity using the two questions	37
5.5	The congruence of language and ethnic/family origins	50
5.6	Test-retest on the ethnicity questions	53
5.7	What happens in a selective screening programme: the case of Exeter	59
5.8	Did the ethnicity questions identify those with clinically relevant haemoglobinopathies: the results of the universal screening programme (Birmingham, King's London, and Leicester)	62
5.9	How did the ethnicity questions identify those with abnormal MCH results, possibly indicating thalassaemia traits	65
6	Conclusion and recommendations	68
References		71
Appendices		72
(1)	Ethnicity questions	72
(2)	Statistical power calculations	74

Preface and Acknowledgements

This *Analytical Report* which focuses specifically on the trial findings represents the work the Centre for Health Services Studies, University of Kent, contracted to undertake in a joint proposal submitted to the NHS Sickle Cell and Thalassaemia Screening Programme in 2002. It forms part of a collaborative research programme to assess the efficacy of the two candidate questions developed in *The Secondary Review* which has involved many people, including researchers, midwives, laboratory staff, the women presenting for antenatal booking in the four sites who agreed to participate in the study, and others. Acknowledgement is also due to Dr Colin Cryer, Senior Research Fellow, Centre for Health Services Studies, for undertaking statistical power calculations for the sample. This report is one of a number arising from this programme (see NHS Sickle Cell & Thalassaemia Screening Programme website for full list). It does not report on laboratory results other than those recorded on the trial schedules nor on the programme of cognitive research initiated by CHSS in May 2002. These and others have been prepared by different members of the Trial Collaborative Group, whose full membership is:

Peter Aspinall, Centre for Health Services Studies, University of Kent at Canterbury
Keith Chambers, University Hospitals of Leicester NHS Trust
Claire Chapman, University Hospitals of Leicester NHS Trust
Fiona Cochran, Royal Devon and Exeter NHS Trust
Suzy Crawford, Sickle Cell/Thalassaemia Counselling Centre, Birmingham
Lorraine Culley, Health Policy Research Unit, De Montfort University
Pam Dobson, Kings College Hospitals NHS Trust
Simon Dyson, TASC Unit, De Montfort University
Sue Dyson, School of Nursing and Midwifery, De Montfort University
Lucille Fifield, Sickle Cell/Thalassaemia Counselling Centre, Leicester
Cynthia Gill, Freelance Haemoglobinopathy Specialist Worker, London
Vanita Jivanji, Sickle Cell/Thalassaemia Counselling Centre, Leicester
Katherine Hooper, Health Policy Research Unit, De Montfort University
Ann Kennefick, Sickle Cell/Thalassaemia Counselling Centre, Birmingham
Mavis Kirkham, University of Sheffield
Janet Lawrence, Sickle Cell/Thalassaemia Counselling Centre, Birmingham
Luriteen Miller, Sickle Cell/Thalassaemia Counselling Centre, Birmingham
Patsy Morris, Kings College Hospitals NHS Trust
Faye Sutton, Royal Devon and Exeter NHS Trust
Sukhjinder Marwah, City Hospital Birmingham
David Rees, Kings College Hospitals NHS Trust
Collis Rochester-Peart, Sickle Cell/Thalassaemia Counselling Centre, SE London
Patricia Squire, University Hospitals of Leicester NHS Trust
Barbara Wild, Kings College Hospitals NHS Trust
Maureen Williams, Sickle Cell/Thalassaemia Counselling Centre, Birmingham
Christine Wright, City Hospital Birmingham
Scott Yates, TASC Unit, De Montfort University

Additionally, the following were members of the trial's Steering Group:

UK Thalassaemia Society
Vincent Cox, Sickle Cell Society
Phil Darbyshire, Birmingham Children's Hospital
Mark Johnson, De Montfort University
Carol King, Leicester OSCAR
Lynne Mathers, Birmingham Children's Hospital
Naresh Rati, Heart of Birmingham PCT (until December 2002)
Rosaline Steele, Royal College of Midwives (until December 2002)
Janet Fyle, Royal College of Midwives (from January 2003)
Josh Wright, Sheffield

The design of the study's research instruments – the questionnaire schedules – was a collaborative effort, Peter Aspinall contributing the two candidate questions to be tested (as reported in Aspinall 2002), ethnicity timing, and the test-retest components, Simon Dyson the components on language, and David Rees the classification of pathology laboratory results. Simon Dyson was responsible for initial dissemination of the study instruments, training the midwives who collected the data, and managing the project in the field during the ten months the trial was active (1st September 2002 to 30th June 2003). Peter Aspinall designed the database for the trial data, coded and entered the full dataset of trial schedules, ran quality checks, derived secondary variables, compiled the analysis specification and undertook the analysis, and wrote this final *Analytical Report*.

PJA

28th November 2003

Executive Summary

1. Two candidate evidence-based ethnic/family origin questions were tested in this formal trial to assess their efficacy in ascertaining risk status in selective antenatal screening for sickle cell and thalassaemia.

2. One of these questions (Question A) is a classification question similar in structure to the 2001 Census England and Wales question but with extended categorisation to capture all appropriate risk groups and a "tick all that apply" method (as opposed to categories) to capture mixed heritage; the other (Question B) is an open response ancestry or ethnic/family origins question, similar in type to those used in US and Canadian Censuses, but comprising an initial "screening" question to identify those with ancestors from areas of the world outside of the UK or Republic of Ireland followed by free text provision to write in countries of ethnic/family origin.

3. The conceptual base of both questions – ethnic/family origins – used in large scale social including government surveys (such as the PSI Fourth National Survey and the 1999 Health Survey for England) was understood by women in a cognitive research exercise. However, some women did not understand the meaning of the term 'ancestors', a finding common to that reported in cognitive research for the 2001 Census Development Programme, raising doubts about its utility as a concept for a screening question.

4. A total of 4,775 women were formally invited to participate (that is, their responses to invitation were recorded) in this trial from 4 centres: Birmingham (n=240), Exeter (n=1059), Kings, London (1197), and Leicester UA & County (2279). These sites represented areas of varying ethnic density, low in Exeter and high in King's, London, Birmingham and Leicester UA. Refusal rates varied from 7.9% (Leicester) to 28.7% (Birmingham) on question A and 7.3% (Leicester) to 38.5% (Birmingham) on question B. Total achieved valid responses numbered 4232.

5. The mean time taken to ask the ethnic/family origin question in the four centres varied from 2.2 minutes in Birmingham to 4.8 minutes in Leicester (median range, 1.1-4.0). The time taken to ask question B was longer than that for question A, although the differences as measured by the median were significant in only two of the four sites (Birmingham and Exeter). Moreover, in two of the sites (Birmingham

and Kings) there was much greater variability in the times recorded for question B than question A. The reasons for these differences are unclear but the lack of context in question B in contradistinction to the detailed categorisation provided in question A and possibility some confusion concerning the conceptual base of question B could have contributed to some of the differences.

6. Question A performed very successfully in all four sites. There was missing data for only 12 respondents (that is, people recruited to question A who did not tick a box) out of a total of 2147 who answered question A, a rate of 0.6% that is tolerable. Moreover, the misinterpretation rate (n=2, 0.1%) related to women who ticked a box for their partner, not a misinterpretation of the ethnic question as such but one of subject status.

7. On part one of question B a gross error rate of 5.53% was incurred amongst the 2081 women recruited into the trial who answered this question. Most of this error (4.18%) was attributable to 'Don't knows' but 1.30% to missing answers. This is a substantial error rate – almost 5% - bearing in mind the kind of error ranges that Zeuner *et al.* (1999) were modelling. This error rate, of course, excludes errors that arose in part 2 of the question.

8. Errors in part two of question B comprised missing data and misinterpretations. There were a total of 22 missing responses for respondents who had ticked 'yes' in the screening question out of a total of 709 responses (or 3.1%, over five times the rate for question A). Moreover, misinterpretations numbered 4 (or 0.68%, over seven times the rate in Question A). Further, there were consequential categorical ambiguities in Question B (for example, five in the Leicester data).

9. Language data was missing on 119 cases of Question A and 91 of Question B (the identical language question set was asked on both schedules). There was evidence of a lack of congruence (strong or equivocal) between assignment on the ethnic questions and language in a total of 52 cases. Ethnicity/language incongruence was much higher on Question B (2.43%, 47/1934) than Question A (0.24%, 5/2063). Language incongruence on Question B was highest in Birmingham and Leicester (at 3.1%), both centres with significant numbers of respondents of South Asian ethnic origin. Part 1 of Question B may have been subject to misinterpretation (as invoking, say, conceptualisations of 'Britishness' or nationality). It is notable that language

incongruence on Question B was much lower in Kings (indeed, the lowest across the 4 sites) whose minority ethnic group population is primarily Afro-Caribbean.

10. On the test-retest component of the trial – a measure of reproducibility or reliability of results and, consequently, the stability of the questions – the consequential error rate for question A was 3 and for question B 25, an extremal quotient of 8.3.

11. In Exeter, the only site utilising selective antenatal screening for sickle cell and thalassaemia, the evidence from the trial suggests that some 45 women or around 5% of those recruited into the study were at risk but not tested.

12. The 3 sites in which universal screening took place – Birmingham, Kings, and Leicester – reported a total of 174 clinically relevant haemoglobinopathies. Of these there were 14 incongruent responses, all but one relating to Question B or 8.0% (18.8% of all Question B responses [n=69] in which clinically relevant haemoglobinopathies occurred). This high rate of incongruent responses on Question B falls well outside the tolerance limits suggested by Zeuner *et al*. (1999), regardless of all the other cumulative source of errors.

13. Findings for MCH values suggestive of thalassaemia and ethnicity are consistent across both questions: a high number of cases of MCH >25<27 in the White-English/Scottish/Welsh/Irish group on question A and the equivalent group on question B ('B – No') and amongst women of Indian ethnic origins and somewhat smaller numbers across the different categories in women of Black-African, Black-Caribbean, or other black origins.

14. On all the measures used in the trial to assess validity and reliability (including test-retest reproducibility), Question A outperformed Question B. Cumulative gross errors on Question B substantially exceeded those on Question A.

15. **It is recommended that Question A be adopted as the candidate question for use in selective antenatal screening for sickle cell and thalassaemia.**

1. Introduction

A comprehensive programme of work has been commissioned by the NHS Sickle Cell & Thalassaemia Screening Programme to assess the feasibility of using questions on ethnic origin as a primary tool for antenatal selective screening for sickle cell disorders and thalassaemia. In the initial phase of this work a comprehensive review of the literature was undertaken in the *Secondary Review of the Ethnic Question* which resulted in the design of two evidence-based questions for testing (Aspinall 2002). These comprised: (i) a conventional ethnicity classification question, similar in structure to the 2001 Census question but with additional categorisation to reflect those ethnic origin groups at risk for the haemoglobinopathies and also a different method for capturing mixed origins (multi-ticking rather than pre-designated categories) [Question A]; (ii) an open response ethnic origin/ancestry question similar to those used in the US and Canadian decennial censuses but with a two part structure – part one asking if the respondent or any of her known ancestors, as far back as she can recall, have ethnic/family origins from areas of the world outside of the United Kingdom or Republic of Ireland & part two asked her to write in all the countries if she answers affirmatively in part one [Question B].

Phase two of this programme has involved both Quantitative and Qualitative elements. The former has encompassed the testing of these two questions in a formal trial and a programme of cognitive research, initiated late in the programme, to explore women's understandings of the conceptual bases to these two questions and how they went about answering them. The trial was undertaken in four sites – King's College Hospital (London), Birmingham, Leicester City and County, and Exeter, areas of varying ethnic density. Exeter represented an area of low ethnic density (around 5%) and King's College catchment an area of high ethnic density (43-50%), the other areas being intermediate to high (Birmingham, 34%) and Leicester UA, 39% [but Leicester County, 7%] See table 1 below. The aim of the trial was to assess the efficacy of the two candidate questions at identifying persons at risk of the haemoglobinopathies in areas of varying ethnic density (so as to include the effect of related contextual factors).

Trial sites – percentage of people in ethnic groups

Ethnic Group	***Birmingham***	***Exeter***	***Kings, London***			***Leicester***	
			Lambeth	*Lewisham*	*Southwark*	*UA*	*County*
All people	977,087	111,076	266,169	248,922	244,866	279,921	609,578
	% of population in ethnic groups						
White British	65.54	94.74	49.57	56.97	52.17	60.54	92.75
White Irish	3.22	0.60	3.26	2.81	3.13	1.29	0.67
Other White	1.49	2.30	9.55	6.14	7.71	2.03	1.29
White & Black Caribbean	1.60	0.16	2.00	1.91	1.37	1.01	0.26
White & Black African	0.15	0.10	0.81	0.64	0.80	0.19	0.06
White & Asian	0.65	0.32	0.79	0.63	0.55	0.68	0.28
Other Mixed	0.47	0.26	1.23	0.99	1.02	0.44	0.15
Indian	5.71	0.26	2.00	1.40	1.49	25.73	3.04
Pakistani	10.65	0.06	0.99	0.44	0.46	1.53	0.15
Bangladeshi	2.13	0.13	0.18	0.49	1.49	0.69	0.23
Other Asian	1.03	0.21	0.77	1.46	0.63	1.97	0.28
Caribbean	4.90	0.06	12.07	12.27	7.99	1.65	0.18
African	0.64	0.13	11.59	9.07	16.07	1.23	0.11
Black Other	0.59	0.02	2.10	2.07	1.84	0.20	0.03
Chinese	0.52	0.34	1.26	1.38	1.83	0.51	0.36
Other ethnic group	0.63	0.33	1.19	1.32	1.45	0.32	0.17

Source: ONS. Key Statistics, table KS06.

The trial was active over a ten month period (1st September 2002 – 30th June 2003), data collection being co-ordinated by local contacts in each of the participating NHS Trusts (City Hospital NHS Trust, Birmingham; Royal Devon & Exeter Healthcare NHS Trust; King's Healthcare NHS Trust; and University Hospitals of Leicester NHS Trust).

The management of the fieldwork – from initial dissemination of the trial instrumentation to the collection of completed schedules was managed by De Montfort University and is not further reported on here, other than to record that: (1) The research protocols were submitted to, and approved by, the Trent Multi-Centre Research Ethics Committee; (2) Amended protocols were submitted to: increase the sample size to satisfy the requirements of a two-tailed hypothesis (that either Question A out-performed B or vice-versa, rather than a unidirectional hypothesis (that A outperformed B), rejecting the notion that the evidence based nature of Question A and the untried nature of Question B merited the assumption of a unidirectional hypothesis; to permit the recruitment of carriers to the study at the point of contact with the haemoglobinopathy counsellor, if they had not been invited to take part in the research by the community midwife at the initial booking-in interview; and to permit the study to run for 9 months rather than 6 months; (3) A total of 57 workshop training sessions were run for midwives, of which 13 were outreach. 88% of midwives attended.

2. Aims of the Trial

The two recent health technology assessment reports (Zeuner *et al*., 1999 & Davies *et al*., 2000) reviewed the two alternative options for antenatal screening: a universal and a selective approach. Selective screening limits screening to those women from ethnic minority groups who are considered to be at high risk. Both reports concluded that universal screening for all women (including those of North European origin) would not be cost effective in areas of low minority ethnic group density, although their specific conclusions differed.

Because of the major costs incurred in a universal approach in geographical areas with a small population at risk, the NHS Sickle Cell and Thalassaemia Screening Programme commissioned research to identify the most effective ethnic question to identify at risk women in these areas. Previous research had indicated a need for a focus on ancestry or ethnic origin as the problems of selective attribution associated with self-reported ethnic group made it too unstable for use in a screening setting. Moreover, while the widely recognised distinction between women of north European and non-north European origin was regarded as important, studies had demonstrated the need for more refined classifications or capture. In particular, identification of groups at risk for α^{0thal} trait was required, that is, Chinese, south-east Asians, and those from the eastern Mediterranean, so as to guide interpretation of laboratory results and inform the need for further investigation of those results judged ambiguous. Further, an increasing prevalence of inter-ethnic unions – including groups at particular risk of the haemoglobinopathies – was rendering traditional approaches to ethnic ascertainment, capture by identification with a single group, of diminishing value for screening purposes.

In the *Secondary Review* (Aspinall 2000) the literature on the conceptual basis of ethnicity questions and on the classifications used in national censuses and specific screening programmes was systematically reviewed. These included the 2001 Census ethnic group question and classification which was adopted for use across Government and for routine monitoring in the NHS (including the mandatory monitoring of the ethnic group of hospital inpatients). However, the conceptual basis of this question – ethnic group – made it unsuitable for selection, as ethnic group is a measure of membership in groups and depends on such matters as group allegiance and psychological needs. Further, the categorisation excluded several important Mediterranean groups at risk for haemoglobin disorders and did not capture all those of either mixed race or of mixed disparate *ethnic* origins where the mix included one or more high risk groups. Consequently an amended ethnic origin question based on the 2001

Census was developed as one of the candidate questions for testing in a formal trial. The other question selected for testing was more strongly based on the traditional distinction judged relevant for screening for the haemoglobinopathies, that is, between women of non-North European and north European ethnic origin, the former identifying women at an increased risk of haemoglobinopathy carrier status. The actual distinction used in the question was that between women having known ancestors from either the United Kingdom or Ireland or from areas of the world outside these isles. As with the classification question, the importance of ancestry and ethnic/family origins was emphasised as both questions were designed to ascertain ethnic status through self-report. However, the more refined identification of some groups at risk for α^{0thal} trait was still required, as was capture of the mixed population where at least one of the mixes was at increased risk for haemoglobinopathy carrier status. Consequently, five free text fields were offered to women who had ethnic/family origins from areas of the world outside of the UK or Republic of Ireland to identify the specific countries from which the woman's ancestors originated.

The two questions were clearly very different although both utilised the conceptual base of 'ethnic/family origins' and the open response question also referred to 'ancestors'. Both questions also contained an explanation of the concept of high-risk groups based on ethnic status. One of the main distinctions – and perhaps crucial for selective haemoglobinopathy screening - is that the classification question (A) is structured around five broad "pan-ethnic" groupings: 'White', 'Mixed', 'Asian or Asian British', 'Black or Black British', and 'Chinese and Other'. These are the same broad groupings as used in the 2001 Census. They may be said to be groupings that describe the broad historical processes of colonialism, migration, and discrimination and substantially reflect societal perceptions of ethnic differences. As summary categories they provide a point of access to the longer-term historical processes that have influenced and shaped the nature of ethnic relations. This is in contradistinction to either specific terms of self-identification (frequently immediate and idiosyncratic) or specific ethnic origins (often synonymous with countries or nationalities) which, nevertheless, are still potentiated by these wider processes.

One of the drawbacks of structuring a classification by using broad pan-ethnicities built on discourses relating to the consequences of these historical processes of colonialism, is that they can be interpreted as carrying strong racial connotations, based on the binary of white and black groups. However, in antenatal haemoglobinopathy screening, the use of such groupings as a structuring mechanism may be an advantage. A focus on ethnic/family origins or ancestry that is accessed only via countries of origin may give no clue to the person's phenotype. For example, a person who identifies 'South Africa' or the 'United States of

America' could be black, white, or a member of any of a further dozen or so ethnic/racial groups covering the full spectrum of risk statuses. By its reliance on the identification of countries of ancestral origins, responses to Question B can easily obscure the person's ethnic/racial origins. In question A, however, these broad pan-ethnic groupings are not the ethnic/family origin response options: rather, these are listed under each of the broad groupings. Thus, what question A offers is a list of 18 options collected under four broad pan-ethnic groups which identify *socially* constructed 'race' groupings and, in general terms, the person's phenotype. Additionally, respondents answering question A can tick as many of these 18 options as they wish.

The second main distinction is that question A provides context but question B virtually none. The context in question A is located in the five pan-ethnic groups and 18 ethnic/family origin categories. People are familiar with this kind of question as it is similar in structure to the 2001 Census question and to those now used on ethnic monitoring forms. Respondents can find the category that fits them best by interpreting the category range. Further, the categories, in themselves, offer strong clues about the underlying conceptual base. The only *specific* context in question B is reference to the countries of the United Kingdom and the Republic of Ireland. Respondents have to infer the meaning of ethnic/family origins and ancestors without access to the response options offered in question A as a "navigational tool". These two clear distinctions suggest potential areas where differences may arise in the accuracy of ascertainment of ethnic status.

The key aim of the trial is to identify which of the two questions best ascertains the true ethnic family origins of the woman booking and hence whether the woman is in a high risk group and would therefore qualify for screening on ethnic grounds. There is clearly no "gold standard" test of a person's ethnic/family origins. Survey evidence shows that relatively few people are knowledgeable in detail about their ancestry beyond their grandparents. However, based on *known* ancestry, most people are able to make a distinction between whether they are of north European or non-north European ethnic origin and are also able to assign themselves to one or more of around 15-20 ethnic origin groups of the kind commonly used on Census/ethnic monitoring forms. Thus, any question soliciting information from a respondent on her ethnic/family origins will (of itself) always be a judgement based on incomplete knowledge and subject to error. The aim of ethnic ascertainment in the screening context is to minimise the various sources of error.

There are a number of these sources of error in ascertaining ethnic status. The woman may not understand the conceptual basis of the question, be it 'ethnic origins', 'family origins', or

'ancestry'. If she does not understand the conceptual basis, she may misinterpret the categories. Further, she may treat different conceptual bases – such as 'race', 'ethnicity', 'national origins, etc. – as all accessing the same semantic domain and therefore answer the question in a way that was not intended. Secondly, the respondent may not understand the categories or interpret them incorrectly. These may be errors of omission (where the woman does not find her *group* or a *group she can identify with*) or commission (for example, where 'Asian' is intended to mean 'Indian subcontinent' rather than 'East Asian' or 'South-East Asian'). Finally, there may be external errors (those external to the question), related to language barriers, a failure to answer the question, or a misconstruction of the purpose of the question.

Identifying the presence and prevalence of these different sources of error and the contribution they make to the accuracy of ethnic ascertainment is a highly complex technical exercise. With respect to conceptual misunderstandings, these are best accessed through *'cognitive research'*, a recognised tool in census testing programmes. Additionally, inferences can sometimes be drawn from the systematic evaluation of response data routinely collected from subjects. Categorical errors can also be identified through cognitive research. However, such methods can only be used with a small number of subjects as they are resource intensive. Another approach is through *survey validation* programmes, sometimes called *test-retest evaluations.* These approaches are constructed around the retest reproducibility of the respondent's initial test response or answer, enabling indices of error rate (such as the 'net error rate' and 'gross error rate' to be compiled from the two sets of responses). A third method that may be available is that of the *'internal consistency'* of a person's individual record in the dataset, that is, whether there is concordance between, say, the respondent's ethnic ascertainment and her preferred language (although this may only be indicative as there is clearly no read-across from language use to ethnic group/origin). The relationship between ethnic ascertainment and laboratory results may provide some additional validation, but the variation of sickle gene frequency across the non-North European population and the dissociation of ethnic group from genetic risk through population mixing diminishes the utility of test results in this context beyond the north European/non-north European population divide. External errors are perhaps the most difficult to detect, for example, the contribution of language ability to the misinterpretation of the ethnic question. Again, the evidence may be incomplete and circumstantial.

All these methods are used in the evaluation of questions A and B. As the number of women recruited into the trial fell well short of that required by the statistical power calculations, such evidence must be judged as indicative rather than statistically significant. However, the

finding of *consistent* differences in effectiveness between the two questions may point to the preferred candidature of one over the other.

2. The Design of the Two Candidate Questions

The conceptual basis of ethnicity classifications and respondents' behaviour in assigning their ethnic origin: research studies and programme evidence

In the secondary review of ethnicity questions (Aspinall 2002) the various conceptual bases of different ethnicity questions were comprehensively reviewed and two evidence-based candidate questions proposed and adopted as candidates for testing in the formal trial (Question A & Question B: see Annex 1). Towards the end of the trial a programme of cognitive research was initiated by Aspinall to explore the utility of questions A & B in the light of equivocal findings about women's understandings of these two questions and the concepts underlying them that were becoming available from the trial and the qualitative research.

Findings are available from two sources: (i) Cognitive and focus group research exercises undertaken by the Office for National Statistics' Social Survey Division as part of the 2001 Census Development Programme; (ii) the cognitive research interviews initiated as part of the testing of the two questions under the auspices of the haemoglobinopathy screening programme. There is also some evidence from census testing programmes in the USA but the different traditions with respect to ethnicity data collection (reflecting different underlying processes of ethnogenesis) have only limited applicability to the UK context.

Cognitive research is a method that is frequently used by census and survey agencies as a way of developing sensitive and appropriate survey questions. Cognitive interviews are sometimes referred to as laboratory thinkaloud interviews with structured probes, the method being akin to an in-depth interview that makes particular use of prompts and probes to gain an understanding of the respondent's cognitive processes. In some applications of the method, the respondent is invited to think out aloud about what is happening cognitively during the process of answering a survey question. Examples of the use of this method in a survey setting include Jobe and Mingay (1990) and Willis *et al.* (1991). Cognitive research has also been extensively used in census development programmes in North America and Britain. In the UK 2001 Census Development Programme (in which Aspinall participated), several cognitive research (Rainford 1997) and focus group (Mortimer & White 1996) studies were undertaken of conceptual issues and specific ethnicity questions. Similar research has been undertaken in the United States in the development of questions on race.

Based on the systematic review of the literature, a number of conceptual bases were reviewed as suitable for questions on ethnic origin for use in an antenatal screening context. Since risk status is determined by origins in certain countries which now are, or were in the past, areas of malarial endemicity, the need is for questions that will capture ethnic/family origins in those areas. Clearly, the focus on origins is key to the capture of groups at risk, rather than conceptualisations of ethnicity that focus on current perceived membership of an (ethnic) group or current affiliation in a cultural sense. Such measures are subject to contextual and situational instability resulting from processes of selective attribution, as demonstrated in census validation surveys that have used test-retest methodologies. Ethnic origins, on the other hand, suggest a focus on the past and on the particularity of place and geography. They also invoke reference to ancestry, descent, forebears, lineage, and the like.

Consequently, the challenge in designing the questions was to identify appropriate concepts that would capture this sense of ethnic origins in particular parts of the world that are/were areas of malarial endemicity. The candidate conceptual bases identified were ‘ethnic origins’, ‘ancestry’, and ‘ethnic family origins’. The concept of ethnic origin has also been widely used in ethnicity questions, although the term appears frequently to be employed as a synonym for ethnic group (Aspinall 2002, pp 29-30). Although not used in the major census tests of 1997 and 1999, interviewers in the 2001 Census development programme’s cognitive research interviews probed individual respondents' understandings of the term ‘ethnic origin’ and how it compared to ethnic group and ancestry (Rainford 1997). 'Ethnic origin' was defined by the majority of respondents as where a person came from and some linked this back to a person's parents and where they originated. Rainford reports that respondents were divided over whether ethnic origin meant the same as the other terms: 'Some thought that it meant the same as ethnic group and others thought that it was different and was instead similar to ancestry. Some respondents also thought that ethnic origin had a wider meaning than ancestry, and referred to more than just parentage'. In addition, as part of the 2001 Census development programme the difference between ethnic origin and ethnic group was explored in a series of focus groups with members of minority ethnic groups. Mortimer and White (1996) reported the finding that ethnic origin referred to 'an individual's parental background, including their own and their parents' place of birth', including 'a sense of history, of linking to the past', while ethnic group referred to 'people from different origins but who have the same social or economic needs'. These participants were happy to talk in terms of their ancestral backgrounds and spoke of the importance of recognising them but did not see the term group in the same way.

Clearly, the sense respondents had of a focus on the past made this concept attractive for use in a haemoglobinopathy screening context, especially as the term is familiar to most and has a history of common usage going back to the 1960s and earlier. However, not all respondents understood the term in this way, some seeing it as a synonym for ethnic group. The possibility, therefore, of the misinterpretation of this commonly used term as 'ethnic group' needed to be considered.

The term *ancestry* has limited saliency in Britain as a measure of ethnicity used in data collections (Aspinall 2002, pp 30-33). The 1991 Census Ethnic group question referred to 'the person's ancestry' in a detailed question instruction but it was not central to the question itself. In terms of usage by the wider society, 'ancestry' is a term generally used to refer to one's (especially remote) family descent or distant lineage. In the individual cognitive interviews conducted as part of the 2001 Census testing programme (Rainford 1997), respondents' understanding was sought of a revised ethnic group question in which those ticking a 'Black-British' category were asked to indicate their ancestry below as 'Of Caribbean ancestry' or 'Of African ancestry'. Similarly, those ticking 'Asian-British' were asked to indicate if they were 'Of Indian ancestry', 'Of Pakistani ancestry', or 'Of Bangladeshi ancestry'. Respondents' understandings were also sought of the term 'ancestry'. Respondents had different ways of understanding this term, including 'their parents', 'their grandparents', and 'generations before their grandparents', and a few did not comprehend the term at all and declined to hazard a guess. One respondent commented: 'Ancestry seems very far away, far removed, origin is just the place where you were born'. Rainford reports that some of the respondents thought of ancestry as generations before their grandparents then applied this definition to their answers. For example, in the Black-British category, there were examples where respondents ticked 'Of African ancestry', although their parental background was in the Caribbean, because they thought of several generations back in their family history. This also led to some multi-ticking of the ancestry subgroups. Rainford cites one respondent: 'I ticked Black/Caribbean and I ticked Black-British, I've got a British passport. And plus, I tick 'Caribbean' and..um.. 'Of African', which my foreparents are from Africa, my ancestors are from Africa, so I have to tick it'. Amongst those who did not understand the term, some were wary of using the 'Black-British' or 'Asian-British' categories in case they were mistaken. Alternative terms for ancestry suggested by respondents included 'descent' and 'parentage'.

However, analysis of responses in the ONS 1997 Census Test found no statistically significant differences in form and question response rates between questions asking for 'ancestral origin' and 'ethnic group' using identical categories (ONS, 1998). Moreover, an interviewer based follow-up survey reported no overall differences in discrepancy rates

(between test form and follow-up answers) and that participants responded to the categories rather than the wording of the questions. While ancestry appears to tap precisely what would be useful in assessing risk of variant haemoglobin in the screening context, that is, lineage, its possible lack of stability in different situations and ONS's decision to eschew the term in the 2001 Census question suggested the need for caution in adopting it as the term of choice. It was suggested that qualitative research could focus on persons' understanding of this term and the extent of non-response attributable to lack of understanding.

The lack of an evidence base on the public's understanding of the term 'ancestry' posed a risk if this concept was going to be used in one of the candidate questions used in the haemoglobinopathy trial. While it has been used in other national censuses, including the 1990 and 2000 US censuses (where it is used synonymously with 'ethnic origin'), it is probable that the public's understanding of the term is different in the United States. For example, the US Census Bureau defines 'ancestry' for respondents as '…. the person's ethnic origin or descent, "roots", or heritage. Ancestry also may refer to the country of birth of the person or the person's parents or ancestors before their arrival in the United States'. Thus, it is a concept that refers to a person's very recent forebears (and even their own country of birth) rather than one's fairly remote family descent. Again, in the 2000 US Census, ancestry was asked of all persons '… no matter how many generations they have been in this country'. The Canadian 1996 and 2001 Census asked about the origins of the person's ancestors ('To which ethnic or cultural group(s) did this person's ancestors belong'), an ancestor being defined in guidance as 'someone from whom the person is descended and is usually more distant than a grandparent'. Thus, the Canadian use of this term is more akin to the colloquial understanding of 'ancestry' in Britain. The term 'ancestry' has also been used in Australian Censuses (1986) but, because of doubts about data quality and insufficient evidence of the value of the ancestry data produced, it was not repeated in the 1991 Census. Although two questions were tested in the 1993 Australian Census Test, it did not prove feasible to include either an ancestry or other direct ethnic origin question in the 1996 census. However, a question using concepts similar to those used in 1986 was included in a census test conducted in May 1997 and the 2001 Census also asked the open response question 'What is the person's ancestry?'.

The usage of the term ancestry in these national censuses reveal important variations in the definition of the term (very wide in the case of the US census, narrower and more distant in the Canadian Census) and that open response questions with examples are now the preferred format, with provision for multiple reporting. Because 'ancestry' is exactly the dimension of ethnicity that needs to be captured in identifying the population at risk, this made the concept a strong candidate for use in the formal trial, in spite of evidence from the 2001 Census

testing programme that some respondents may not understand the term and even fail to answer the question based on such misunderstanding.

The third concept identified as a possible candidate was *ethnic/family origins* (Aspinall 2002, pp 33-34). Although little used in survey settings before the mid-90s, more recently the useful term 'family origins' has been used in two large national surveys (the PSI 4th National Survey of Ethnic Minorities [Modood, Berthoud, Lakey, Nazroo, Smith, Virdee, *et al*., 1997] and the 1999 Health Survey for England [Erens, Primatesta, & Prior, 2001]) and 'family's ethnic origins' (of mother and father) in a suggested but untested question for the 2001 Census (Berthoud, 1998). 'Family origins' is a fairly explicit reference to descent or origins that appears to lack the level of misunderstanding associated with the term 'ancestry'. The term is used ('Origine des familles') in the French neonatal sickle cell screening programme (Galactéros, 1999). Additionally, the concept of family origins was adopted in one of the NHS Executive's primary care pilot sites for the collection of ethnic group data (Princes Park, Liverpool). No adverse findings were reported from these two surveys that suggested that there were difficulties amongst the public in understanding the term. Moreover, the focus on *family* origins invoked the concept of ancestry without actually using that term, making the term attractive for use in the selective antenatal screening setting.

A fourth concept - *Country of origin and national origins* – was ruled out as these generic terms are not used in ethnicity classifications in Britain, although the latter has some saliency in the USA (Ahdieh & Hahn, 1996), and neither are ethnic specific. There is evidence from the United States that Asian and Hispanic migrants primarily identify with their countries of origin and 'national origin' is defined in terms of country of birth by most people (although occasionally being defined in terms of citizenship rather than place of birth). In the development and cognitive testing of race and ethnic origin questions for the US 2000 Census, Gerber and de la Puente (1996) found that for some respondents, 'ancestry' and 'national origin' were two quite different concepts. For example, White respondents often saw their 'ancestry' as a European country but their 'national origin' as 'American'. This was also the case for Hispanic and Asian migrants whose children were born in the United States or who identified themselves as 'American' because they were citizens. The term 'national origin' was dropped because the interest of the census was in ancestry rather than country of origin. The ambiguity in this term's meaning - its potential for confusion with nationality or citizenship, its uncertain relationship with ancestry, and its lack of saliency in Britain - disqualifies it as a candidate conceptual base for haemoglobinopathy screening.

Following a process of evaluation and appraisal of the various candidate conceptual bases (described in *The Secondary Review of the Ethnic Question*, Aspinall 2002), it was decided to use the concept of 'ethnic/family origins' as the primary candidate of choice in the two questions selected for testing in the formal trial. The classification question (Question A) asked respondents: 'Do you have *ethnic/family origins* that are….' and the open response question (Question B, part 1): 'Do you or any of your known ancestors, as far back as you can recall, have *ethnic/family origins* from areas of the world outside of the United Kingdom or Republic of Ireland?'. Both question also made use of the concept of *ancestry* in a common question instruction ('We are asking about ethnicity because we want to know who is at risk from sickle cell/thalassaemia. These are serious inherited blood disorders that are more common in peoples whose *ancestors* lived in malarial areas of the world such as Africa, Asia, the Middle East, and the Mediterranean. Bearing this in mind…'). Moreover, parts 1 and 2 of question B also referred to *known ancestors* (part 2 in: 'If Yes, then for you or for any of your *known ancestors*, as far back as you can recall, please write in all the countries in the spaces below' [5 free text boxes].

Although both these questions are evidence-based (drawing on and presented in *The Secondary Review of the Ethnic Question*), there were uncertainties regarding how they would perform as questions amongst a cross-section of women of reproductive age in the antenatal booking-in setting. These uncertainties included how these women would understand the concepts used in the questions ('ethnic/family origins', 'known ancestry'), their ability to recall known ancestry and to 'operationalise' this knowledge in terms of countries of origin (question B), their ability to declare multiple ethnic/family origins through the 'tick one or more boxes' instruction (question A), the women's understandings of the particular category labels used in question A (some of which, like 'Italian, Maltese, or other Mediterranean' and 'North African, Arab or Iranian' had not been tested in other studies), and whether these women differentiated between the different semantic domains referenced in these questions and others to which they were accustomed to answer (for example, in the census).

As the trial progressed in 2002 and the data from the schedules was coded and entered on to a customised proprietary database with statistical functionality (Statistical Analysis Software [SAS] Copyright © 2003 SAS Institute Inc), it became evident that there were some recurrent responses that were indicative of respondent misunderstandings of the conceptual bases of the questions (especially with respect to question B) and (with less frequency) of the categories used – notably, the term 'Asian' – in question A. There is clearly a difficulty in assessing the performance of candidate questions (and especially complex issues like peoples' understandings of ethnicity concepts) from respondents' answers as there may be systematic

errors in the interpretation of the concepts that do not manifest themselves in – or cannot be inferred from - the responses given. By the early summer of 2002 capture of such issues was limited to findings emerging from a parallel programme of qualitative research being conducted under the NHS Haemoglobinopathy Screening Programme and interpretation and statistical analysis of the data from the questionnaire schedules.

In mid-May 2002 a programme of cognitive research of the type that had been used in the 2001 Census Development Programme was proposed to members of the project team as a means of gaining insight into some of the issues that were emerging from the trial data. Besides the systematic review of the evidence that had been undertaken in *The Secondary Review of the Ethnic Question,* there had been no programme of qualitative research to investigate some of the issues raised (such as respondents' understandings of the different conceptual bases). For example, in the work undertaken ahead of the PSI 4th national survey on minority ethnic groups, Modood *et al*. (1994) had undertaken structured depth interviews with 47 persons and group discussions with 27 (74 persons in all) to explore issues relating to ethnic identity.

A specification for a programme of individual cognitive interviews was drawn up that would involve between 20 and 50 face to face cognitive interviews with women consecutively presenting who had participated in the main trial. Clearly, the higher the number, the more generalisable would be the findings although the main purpose of such interviews would be to gain insight into interpretative issues. Moreover, the need was identified for a structured set of data in which responses to common questions on the cognitive schedule could be compared across the two question formats. The author prepared a short proforma, drawing upon a personal collection of literature on cognitive research in census and survey settings, including that relating to the 2001 Census Development Programme and similar work in the United States. A number of heads of enquiry were initially identified, including the woman's understanding of the conceptual base, whether she reads the question instruction or goes straight to the categories, how the woman 'screens' the categories (whether she previews all or works down the list), whether there were any ambiguities in the category listing, whether there were categories that she would have liked to see in the listing, whether her final choice was 'forced' or 'satisfactory', and how, ideally, she would have liked to respond. These were prepared as a structured list with space on the data collection instrument to record the woman's response in full. As in the cognitive research method, the schedules would contain a number of prompts for the interviewer against each head of enquiry. With a targeted response rate of 30-50 women split between question A and question B, the research might provide some clues to the way in which the women were interpreting the questions and whether there

were some systematic sources of error relating to their design and content that might be contributing to erroneous responses. Obtaining a response from the woman on each of the 'heads of enquiry' (even if it was 'negative', non-committal, etc.) through a structured data collection approach would enable some strong inferences to be made about the effectiveness of the questions and what are, and are not, aberrant (one-off) reactions. Further, emphasis was placed on capturing a spectrum of ethnic groups but weighted in favour of minority ethnic groups.

This programme of cognitive research was seen to complement findings emerging from the parallel programme of qualitative research. Two issues have faced the team in assessing the feasibility of asking candidate ethnic origin questions to assess risk of haemoglobinopathy carrier status in a selective screening context: (i) Is it feasible to obtain robust measures of ethnic family origin in a booking-in interview with all the other demands on the midwife in that setting, that is, an operational issue; (ii) Can ethnic family origin be used as a basis for risk ascertainment/selectivity for screening by capture through candidate questions in a formal data trial, i.e., an outcome issue geared to technical criteria. To answer the latter, some measure of the gross error rates incurred by the two questions and an understanding of the factors driving those error rates is needed. With respect to the second, the midwifery skills available in the qualitative team have been successfully deployed to assess the practical issues of operationalising data collection within the context of a frequently time-limited antenatal booking-in interview.

By definition direct observational time sampling can only give us limited insight into the cognitive processes involved in a decision-making process (although there were debriefing interviews as well). To obtain answers about the effectiveness of questions A & B, some way of getting inside the woman's cognitive or thought processes during her receipt, reading, and formulation of a response to the ethnic question was needed. This can only really be successfully undertaken through the kind of cognitive research methods used by census and survey researchers, that is, highly focussed and rigorous questioning about what is actually going on at a cognitive level (including the use of think aloud methods). There may be a lot of fuzziness in the thinking of the respondents around these questions and their development of a response and it is important to systematically capture this. Moreover, some of the comments captured in the qualitative research programme would be subject to reporting/recording biases and cannot be considered to be a systematic data set of utility in evaluating the effectiveness of the two questions. That is, it would have been difficult to have made a judgement on this basis about where the main weights (of both benefit & drawback) were located with respect to the two questions. It is easy in such a situation for a set of two or three common points to

assume disproportionate importance when they are, in fact, a small subset of depth interviews drawn from a trial population approaching 5,000 women.

Data collected around a set of "heads of enquiry", systematically recorded (so that for each item there is a recorded response), will give a stronger evidence base. One of the difficulties of interpreting responses recorded in a structured proforma used for data collection in a trial is that they will only occasionally (and sometimes incidentally) provide insight into what the woman's understanding is. The questionnaires may be 'working' as questionnaires - in that the woman is giving what appears to be a valid answer - but this may provide few clues to the validity of those answers (except where there is corroborating evidence such as language congruence with reported ethnic family origins or test-retest reproducibility). There may be substantial, significant, and consistent error built into the data if women are interpreting the conceptual base, question instruction, categories, and other apparatus of the question in a discrepant way or one which was not anticipated or intended. Cognitive research provides a point of access to these matters.

Evidence from the PSI 4th National Survey suggested that the question on family origins used in that survey was more discriminant in detecting disparate and multiple ethnic origins than the question on ethnic group membership, in tables that cross-tabulated responses from both questions (Nazroo 1997). The women interviewed in the cognitive research (25 women for schedule A and 25 for schedule B) were asked what the term 'ethnic/family origins' used in the question meant to them in their own words. A conceptual map was drawn from the set of 50 responses (available from Aspinall) and showed no substantial difference between the two question sets. In the cognitive research amongst Question B respondents, women were asked what the term ancestors meant to them in their own words. Two of the respondents (8%) did not know and a third woman pointed out that other people may not know its meaning because of the length of the word. Such a high proportion in a screening programme immediately raises concerns about the utility of this term. Five other respondents offered synonyms for the word (such as 'your forefathers', 'forefathers, relatives', 'foreparents', 'your forefathers, foreparents', and 'my forefathers'). Two other respondents made general reference to past family members ('where the family members have come from in the past'; 'past family members/relatives'). However, the vast majority of respondents (16, almost two-thirds) referred to one's especially remote family descent or one's ancestors collectively, the interpretation that is useful in a screening context (e.g., 'could be your great grandmother, further down your history line'; 'I'm not sure but I think its about grandmother, great grandmother, great great grandmother & down the line'; 'all my direct ascendants, as far back

as I can recall'; 'anyone, your mother, father, grandmother, grandfather and anyone your're descended from'; 'my great great great grandparents'; 'my family as far back as I can remember or cannot remember - even though I don't know my great grandparent – they are still my ancestors'; 'I would think back to my grandparents & the family tree, e.g. if there was heart disease in the family prior to my grandparents I would be aware of it'; 'that's like your distant, distant, very distant relatives'; 'parent, grandparents, and before grandparents'; 'your forebears…generation after generation of your family. Important in making up who you are at a biological/genetic level'; 'if I really wanted to be difficult I would say my ancestors came from Africa but I don't know who they are before my grandparents'; 'my ancestors I can't see & my grandparents I can see'; 'your family, but years & years & years ago'; 'my great grandfather & great grandmother - that as far back as I would go'; 'your relatives from a very long time back'; 'your family going back a long way'; 'mother's mother's, mother's etc.').

While the majority of interviewees had an accurate understanding of the term 'ancestry' that would make it useful to employ in a selective haemoglobinopathy screening context, the relatively high proportion of women who didn't understand the term or thought others might not (3 or 12%) renders its use in this context problematic. The 2001 Census development programme's cognitive research on the conceptual base of ethnicity questions came to the same conclusion about the term 'ancestry' with respect to its employment in the 2001 Census question, that is, the lack of understanding of the term amongst some participants ruled it out as a candidate.

The cognitive research schedules were designed to provide question sets specific to the two trial questions but also three questions common to both (see annex): how well the question met the needs of the respondent in describing their ethnic/family origins; the level of comfort experienced by the respondent in using the question to describe their ethnic/family origins; and whether there were any ways in which the question could be improved in terms of design, layout, write-in provision, or anything else. These three questions enable direct comparisons to be made between the two trial options.

Cognitive research findings on questions common to options A & B

Topic	Question A	Question B
Did the question fully meet the respondents' needs?		
The question fully met my needs	19 (76%)	17 (68%)
I would have liked other provision	6 (24%)	7 (28%)
Don't know	0	1 (4%)
Level of comfort in using the question?		
Very comfortable	21 (84%)	16 (64%)
Somewhat comfortable	1 (4%)	4 (16%)
Somewhat uncomfortable	2 (8%)	4 (16%)
Very uncomfortable	1 (4%)	1 (4%)
Are there ways in which the question could be improved?		
Yes	9 (36%)	13 (52%)
No	15 (60%)	11 (44%)
Don't know	1 (4%)	1 (4%)

Based on these 3 questions, option A performed better on all three measures, especially in terms of respondents' comfort in answering the question but also in lack of need for improvement.

4. Analytical methods

A database to process the data was designed using the proprietary statistical database package *Statistical Analysis Software* [Copyright © 2003 SAS Institute Inc] [see also *Documentation*: 83 variable data dictionary]. Data from the questionnaire schedules was precoded for selected fields: times were entered in minutes correct to two decimal places and weeks pregnant at the booking-in interview correct to two decimal places. The responses to the two ethnic questions were entered as generic, free text strings – in the case of multi-ticking in Question A, each response box was linked by the operand +; in the case of Question B the responses in each free text box were also linked by the operand + (individual descriptive components within a free text box being comma separated). The responses to the language components ('the client's preferred language for this booking-in interview' and 'how the interview was conducted if the client's preferred language was not English') and the pathology laboratory results were numerically coded. All free text (in open response questions) was entered as *verbatim* and uncorrected.

Range checks were performed on all the data and values out-of-range checked against original questionnaire schedules. A sample of data was re-keyed to validate data entry. The quality of some of the data was poor (missing values, non-standard reporting on the questionnaire schedules, reporting with annotation, etc.). A significant amount of additional coding has taken place subsequent to initial data entry to ensure consistency in the dataset.

Analysis was undertaken using standard one-way and two-way tabulations with descriptive statistics. Tests of statistical significance (Question A vs. Question B) are performed on all comparisons. Some of the outcome data - time taken to answer the ethnic question and weeks' gestation at booking-in interview - may be amenable to logistic regression modelling in further analysis of the dataset, although the number of explanatory variables are few.

5. Results of the Trial

5.1. Participation Rates and Refusals

Although all the participating centres were issued with sufficient schedules to enable all women presenting for antenatal screening in the four sites to participate during the active life of the trial, a substantial proportion of the women were not given this opportunity of participation. Some of the midwives chose not to become involved in the study so women booking through them were not asked. Some measure of the extent of midwives' decision not to participate in the study can be assessed from the difference between figures of all women booking during the active phase of the trial and the aggregate count of refusals and women recruited.

The trial population – from which women would be recruited – was defined in the study protocol as all women presenting for booking at the four sites – so assessment of the results must be made against this population. The number of women who were invited to participate but who declined is unknown as records were not kept of all such refusals. Thus, the figures of women who declined that are given in the results are only those who the midwife recorded as 'Declined' on the schedule and are an unknown proportion of all women who declined participation.

Overall, 4,775 women were *recorded* as having being invited to participate in the study, of whom 542 (11.4%) declined. Refusal rates were highest in Birmingham (33.8%), the site that also had the lowest participation rate (possibly because of the restricted time allocation for antenatal booking in the trust). In Exeter the refusal rate was 15.8% (less than half that in Birmingham) and was even lower in Kings (10.1%) and Leicester (7.6%).

Refusals for B (n=272, 11.6%) slightly exceeded those for A (n=269, 11.1%) (table 1). In Birmingham, however, refusals for Question B were 10 percentage points higher than for Question A (but with minor differences in the other 3 sites). Given the uncertainty with which midwives recorded refusals, no particular significance can be attached to these figures.

Table 1. Recruitment and Refusals

	Birmingham	**Exeter**	**Kings**	**Leicester**
Refused	A=35 (28.7%) B=45 (38.5%) 81	A=75 (14.5%) B=92 (17.0%) 167	A=67 (10.8%) B=54 (12.5%) 121	A=92 (7.9%) B=81 (7.3%) 173
Completed	A=87 B=72 159	A=442 B=450 892	A=552 B=524 1076	A=1071 B=1034 2105
Total	A=122 B=117 240	A=518 B=541 1059	A=619 B=578 1197	A=1163 B=1115 2279

Note: Leicester data includes 8 completed schedules which were retests only with no EQANS no. for linkage and one response is recorded as completed by a counsellor only. The King's data recorded that 7 women were recruited at counselling and for a further participant 2nd interview data only was available.

In Birmingham midwives had annotated the schedule in 4 cases of refusal ('Client already knows her status'; 'Declared she's Somalian'; 'Indian'; 'Reading information sheet and explaining sickle/thal took half an hour; and 'Reading information sheet took half hour'). In the Exeter data a few respondents who refused (<3%) annotated the schedules with reasons for refusing, such as: 'is moving away from Exeter during her pregnancy'; 'patient adopted – family history not known'. In the Kings data annotations relating to declines were given for 3 women ('no English; no translation available; GP didn't inform us of need for interpreter'; 'trait' (n=2). In Leicester 4 midwives recorded that 4 respondents who refused stated that they had had previous pregnancies in which they had been tested.

5.2. Randomisation

The random allocation of schedules A and B to the patient population appears to have worked satisfactorily, schedule A comprising 50.8% of the total in Birmingham, 48.9% in Exeter, 51.7% in King's, and 51.0% in Leicester.

5.3. The time taken to ask the ethnic questions

The research instrument encompassing question A or B contained instructions for the midwife to record in minutes and seconds the length of time it took the client to complete the ethnicity

question, including any time requirements generated by the ethnicity question for explaining sickle cell/thalassaemia to the client. Provision to time the start and end of the process was made available to midwives.

Table 2. Time taken to answer the ethnicity questions (see also figs. 1 & 2)

	Birmingham	***Exeter***	***Kings***	***Leicester***
Mean (average)	2.22	2.32	4.49	4.78
95% Confidence Level for Mean	0.41	0.16	0.32	0.20
Standard Error	0.21	0.08	0.16	0.10
Median	1.08	1.75	3.00	4.00
Mode	0.50	2.00	2.00	5.00
Standard Deviation	2.61	2.37	4.99	4.44
Sample Variance	6.80	5.61	24.94	19.72
Range	14.83	19.92	82.97	49.98
Minimum	0.17	0.08	0.03	0.02
Maximum	15.00	20.00	83.00	50.00
Count (valid cases)	157	829	959	1990

Note (a): (i)Times were converted into minutes, correct to two decimal places.
(ii) For Leicester 11 consecutive cases (60 [4], 40 [1], 38 [1], 30 [3], 25 [2]) are omitted from analysis – timings given by haemoglobinopathy counsellor appear to relate to entire encounter.
(iii) Only cases with a valid time have been reported.

Note (b) on descriptive statistics: The *standard deviation* is a measure of the variability between individuals in the level of the factor being investigated and is thus a descriptive index. It is more robust than *sample variance* (also a measure of variability from which the standard deviation is derived). *Standard error* is a measure of the uncertainty in a sample statistic. The standard error of the sample statistic, which depends on both the standard deviation and the sample size, is a recognition that a sample is most unlikely to determine the population value exactly. In fact, if a further sample is taken in identical circumstances it will almost certainly produce a different estimate of the same population value.

Overall, mean times were longest in Kings and Leicester sites and shortest in Birmingham (table 2 & figs 1 & 2). It is probable that the high proportion of ethnic minority women amongst participants in the Kings and Leicester sites (and the very low percentage in Exeter) account for these differences, although times were shortest in Birmingham (again, possibly because of the trust's policy to restrict the antenatal booking in interview to a relatively short time). These differences are also reflected in median (a better indicator than means) and modal times. Indeed, the median time in Birmingham (1.1 minutes) was significantly lower than that for Exeter (1.8 minutes). Consistency in times recorded (as measured by standard deviation and sample variance) was highest in Kings and Leicester and lowest in Birmingham. The range in times taken was particularly high in Kings and in Leicester was also well above that recorded for Birmingham and Exeter.

Fig. 1:

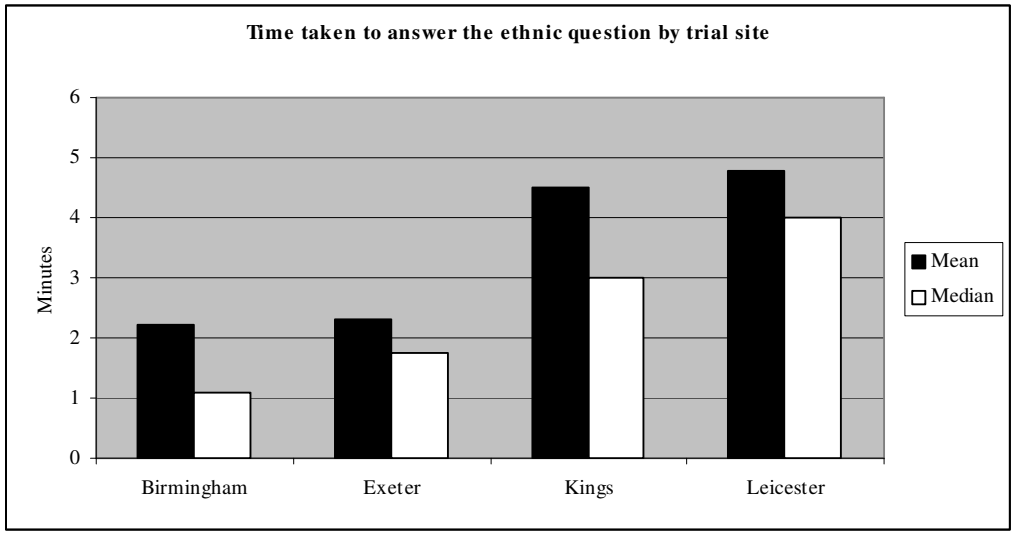
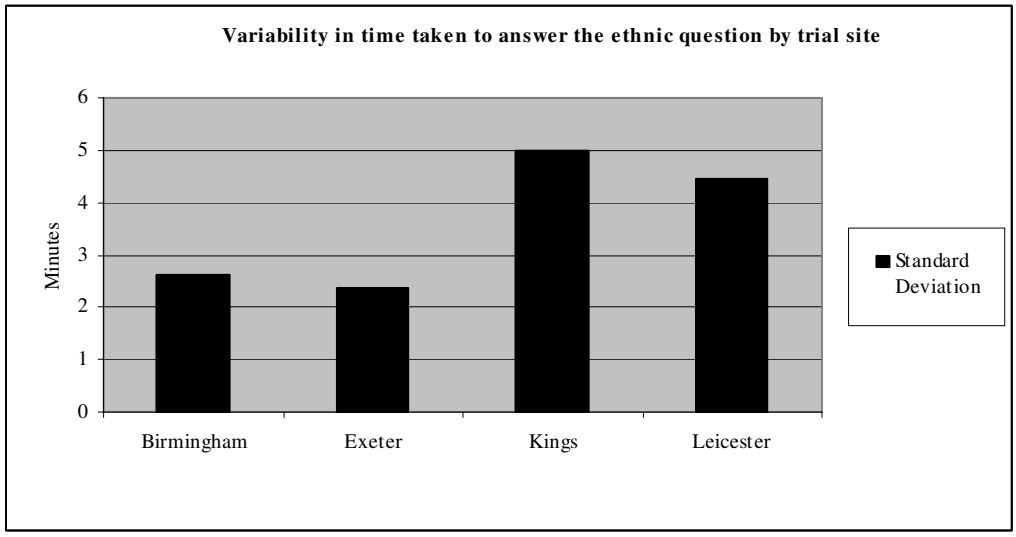

Time taken to answer the ethnic question by trial site
Minutes
0
1
2
3
4
5
6
Birmingham
Exeter
Kings
Leicester
Mean
Median

Fig. 2:


Variability in time taken to answer the ethnic question by trial site
Minutes
0
1
2
3
4
5
6
Birmingham
Exeter
Kings
Leicester
Standard Deviation

Although the question instruction asked midwives to include any time requirements generated by the ethnicity question for explaining sickle cell/thalassaemia to the client, the extent to which this instruction was consistently followed is unknown. However, in the Exeter data 10 midwives gave times in annotations on the schedules for asking the question broken down into asking the question itself and discussion. Discussion accounted for 60-91% of the time (>75% in 6 cases). Two Leicester respondents also gave separate timings, discussion accounting for 70% and 80% of the total time. Thus, the time taken for actually asking the question (exclusive of any discussion) is likely to have been around a third of those times recorded in table 2.

An analysis has also been undertaken of times taken across the four sites for the two Questions (A and B).

Table 3. Time taken to answer the ethnicity questions in Birmingham

Birmingham	***Qn A (classification)***	***Qn B (open response)***
Mean (average)	1.86	2.65
95% Confidence Level for Mean	0.42	0.75
Standard Error	0.21	0.37
Median	1.00	1.51
Mode	0.50	0.50
Standard Deviation	1.95	3.18
Sample Variance	3.79	10.12
Range	7.86	14.75
Minimum	0.17	0.25
Maximum	8.03	15.00
Count (valid cases)	85	72

In Birmingham the mean time to ask question B (2.7 minutes) was significantly longer than that for question A (1.9 minutes or 70.2% that for question B) (table 3). This was also reflected in the median times (question A taking two-thirds (66.2%) of the time for question B). There were also significant differences in the consistency of times recorded for question A and Question B (with Question B showing greater variability as recorded by standard deviation and sample variance).

Table 4. Time taken to answer the ethnicity questions in Exeter

Exeter	***Qn A (classification)***	***Qn B (open response)***
Mean (average)	2.29	2.35
95% Confidence Level for Mean	0.22	0.23
Standard Error	0.11	0.12
Median	1.50	2.00
Mode	1.00	2.00
Standard Deviation	2.33	2.41
Sample Variance	5.43	5.81
Range	19.92	14.92
Minimum	0.08	0.08
Maximum	20.00	15.00
Count (valid cases)	417	412

In Exeter the differences were less pronounced but consistent with respect to the two questions (table 4). The mean time taken to ask Question A and B was very similar, although the median shows that Question A was quicker (75% of the time taken to answer question B). This is what one would expect as question B had two parts and Part two required writing in an answer (a more time-consuming exercise than ticking a box). Also, again as one might expect, there was less consistency in recording times for question B (as revealed by standard deviation and sample variance), although the differences were not significant.

Table 5. Time taken to answer the ethnicity questions in Kings

Kings	***Qn A (classification)***	***Qn B (open response)***
Mean (average)	4.31	4.69
95% Confidence Level for Mean	0.35	0.54
Standard Error	0.18	0.27
Median	3	3
Mode	2	2
Standard Deviation	3.94	5.90
Sample Variance	15.54	34.77
Range	24.97	82.95
Minimum	0.03	0.05
Maximum	25.00	83
Count (valid cases)	491	468

The mean times taken to answer questions A and B in Kings were very similar (question A being 91.9% that for question B) and the median times were identical (3 minutes) (table 5). However, again there was significantly less consistency in the times recorded for B than A (as shown by standard deviation and sample variance).

Table 6. Time taken to answer the ethnicity questions in Leicester

Leicester	***Qn A (classification)***	***Qn B (open response)***
Mean (average)	4.64	4.92
95% Confidence Level for Mean	0.27	0.29
Standard Error	0.14	0.15
Median	4.00	4.00
Mode	5.00	5.00
Standard Deviation	4.32	4.56
Sample Variance	18.64	20.80
Range	39.98	49.97
Minimum	0.02	0.03
Maximum	40	50
Count (valid cases)	1005	985

Note: 11 consecutive cases (60 [4], 40 [1], 38 [1], 30 [3], 25 [2]) omitted from analysis – timings given by haemoglobinopathy counsellor appear to relate to entire encounter.

Finally, in Leicester, the mean time for asking question A was shorter than for question B (94.3% of the time for question B), although the median time (4 minutes) was identical (table 6). Again, there were differences in consistency of times taken to ask the questions although not significant in this site.

Fig. 3:

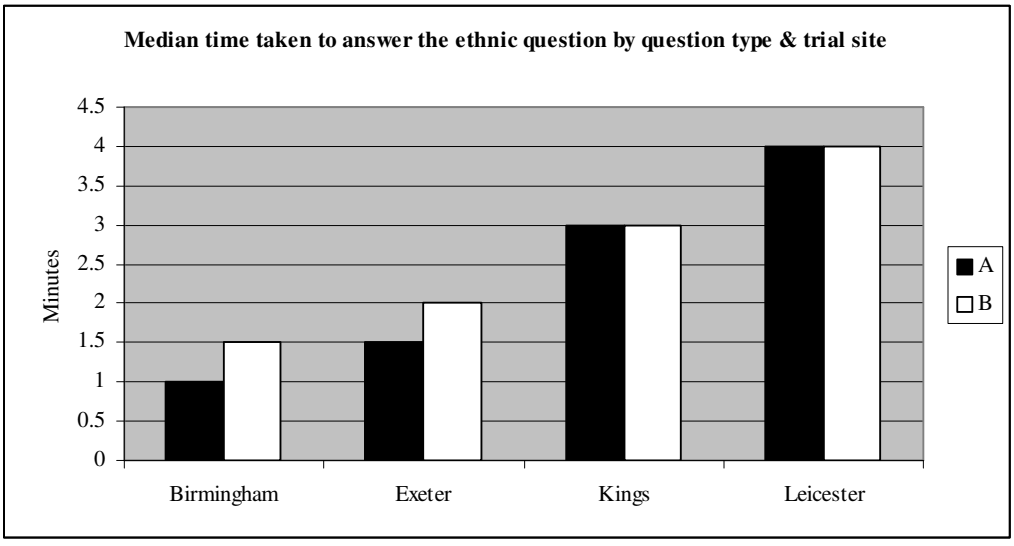
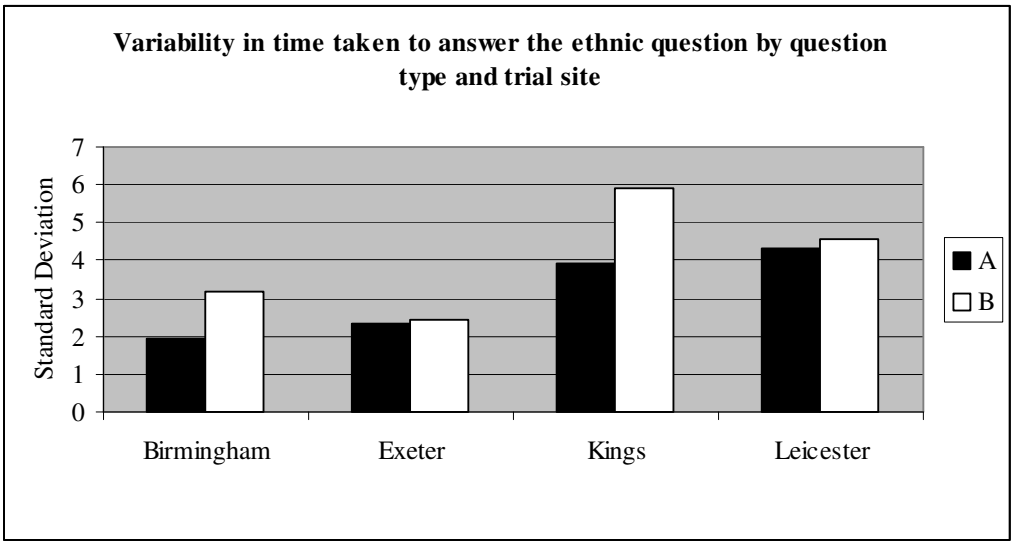


Fig. 4:



Discussion

The time taken to ask question B was longer than that for question A, although the differences as measured by the median were significant in only two of the four sites (Birmingham and Exeter) (fig. 3). Moreover, in two of the sites (Birmingham and Kings) there was much greater variability in the times recorded for question B than question A (fig. 4). The reasons for these differences are unclear but the lack of context in question B in contradistinction to the detailed categorisation provided in question A and possibility some confusion concerning the conceptual base of question B could have contributed to some of the differences.

The length of time it takes to ask the ethnic question is crucial to the operationalising of the use of a screening question. The introduction of the question into routine practice settings would incur costs for the additional time spent by midwives in determining the ethnic origin of the women at booking, the costs of education, training, guidance and other resources to support implementation, and the costs to hospital trusts and PCTs to monitor compliance with a selective screening strategy. The general practice feasibility study of ethnic monitoring revealed that for 2.1 per cent of patients (n=678) the staff member felt uncomfortable asking about ethnicity, a significant correlation existing between the time taken to ask the patient (over 3 minutes in 4.3 per cent) and level of discomfort (Pringle & Rothera, 1995, 1996). Almost half the primary care staff felt that 'thorough training' would be required to ask about ethnic group (Pringle & Rothera, 1995). From the midwife's point of view, the time taken to

administer the question within a busy schedule - assumed by Zeuner *et al*. (2000) to be part of routine antenatal booking but estimated at 1-2 minutes in sensitivity analyses and estimated by Davies *et al*. (1999) ('obtaining a family history') to take an average of 5 minutes - the motivation to elicit the relevant information rather than an answer that satisfies bureaucratic form-filling, and the confidence to ask the client and offer reassurance when needed are all likely to determine the level of successful implementation. Few research findings are available on time needed for ethnic ascertainment. A cost-effectiveness evaluation of newborn haemoglobinopathy screening in Alaska (Gessner *et al*., 1996) modelled race ascertainment costs at 0.50$, estimated from price quotes from the Oregon Public Health Laboratory, the newborn screening reference laboratory used by Alaska. Racial targeting costs were corroborated in this study with the personnel from the newborn screening laboratory in one other state. One author estimated the cost of ascertaining race in the hospital setting as $2.73 based on a trained nurse being able to accurately determine race for six infants per hour (Lane & Eckman 1992). Another investigator believes that in most instances the cost of racial ascertainment will be nominal (Tsevat, Wong, *et al*., 1992). As in the study by Zeuner *et al*. (1999), Gessner *et al.* (1996) argue that the cost is important to quantify because of its large impact on the ordering of selective follow-up options.

The timings recorded in the trial (both mean and median) are generally higher than those used by Zeuner *et al*. (1-2 minutes in sensitivity analyses): a range of mean times of 1.9-4.6 minutes for question A (median, 1-3 minutes) and of 2.4-4.9 minutes for question B (median, 1.5-4 minutes). They are closer to the time estimated by Davies *et al*. (1999) for 'obtaining a family history' (an average of 5 minutes). The mean 2 minute threshold was exceeded in all but one site for question A and in all sites for question B.

In the general practice study (Pringle & Rothera 1996), three quarters of the recording episodes for asking the ethnicity question took less than one minute (339/454 (74.7%) in Lincolnshire and 303/390 (77.7%) in Leicester), but in 34 cases (4%) it took more than three minutes. Given that around 75% of the time taken to ask the ethnicity questions in the haemoglobinopathy trial was probably spent in discussion with respect to sickle cell/thalassaemia (based on a very small number of timings), the trial findings are broadly consistent with those found in the primary care study.

It might be useful to use the timings recorded in the trial as new parameters in the Zeuner *et al*. (1999) model to see what the implications would be for decisions relating to selective vs. universal screening based on their cost-effectiveness analysis. The length of time taken to ask the ethnic question was critical to the balance of advantage with respect to screening options

in this model and the differences between question A and question B (as well as overall timings) may also be significant to outcomes even though they are small in magnitude.

5. 4. Self -ascribed ethnicity using the two questions

Respondents in the trial were randomised to receive schedules containing either question A or question B, questions developed in *The Secondary Review* stage of this work. As already noted, question A is a conventional classification question that has some similarities in structure to the 2001 Census question on ethnic group but with additional categories, a different way of capturing the mixed heritage population ('tick one or more boxes' rather than predesignated categories), and a different conceptual base (ethnic/family origins rather than ethnic group/cultural background). Question B is an open response (free text) question with a part one or initial screen (that seeks to identify whether the respondent or any of her known ancestors, as far back as she is able to recall) have ethnic/family origins from areas of the world outside of the United Kingdom or Republic of Ireland) and a part 2 (asking for such respondents to write in all the countries for themselves and their known ancestors in five free text boxes supplied).

At the end of both questions a box was supplied which the respondent could tick to indicate 'Ethnicity Information Refused'. None of the respondents across the four sites who agreed to participate in the study utilised this and this finding is consistent with evidence on refusals presented in *The Secondary Review* (Aspinall 2002, pp. 14-19).

Both questions could evoke free text responses: Question A in the four free text fields attached to the 'Any other….' Categories and question B (part 2) if the respondent utilised the write-in provision. Question A also had provision for multi-ticking. Because of the complexity of the responses (some on question A involving multi-ticking and free text) the data was entered into the proprietary statistical package (SAS) as a generic text string and was subsequently recoded depending on the type of analysis undertaken. The responses across all sites to question A, followed by B, are presented.

Question A

In Exeter of the 442 respondents who answered question A data was missing for none. A total of 382 (86.2%) gave 'White (English, Scottish, Welsh or Irish)' as their ethnic family origins

and 61 respondents other origins (see table 7). This question appears to have been very successful at capturing mixed responses (n=40) and also white groups at risk. For example, a total of 16 persons identified as 'Italian, Maltese, or other Mediterranean' (5 in single responses and 11 in mixed responses). Census tests and trials in the USA have consistently reported improved capture with the naming on a group (in both classification and open response questions).

Table 7. Responses to question A ascertaining ethnic/family origins

	Birmingham	Exeter	Kings London	Leicester
WHITE				
English/Scottish/Welsh/Irish	9	381	179	758
Other N European	1	3	16	19
Greek/Greek Cypriot & Turkish/Turkish Cypriot†	0	0	9	2
Italian, Maltese, or other Mediterranean†	0	5	9	9
Any other White background	1	4 [E3]	26[K1]	15
Mixed white – not at risk	0	12 [E1]	17[K2]	7
Mixed white – at risk†	0	12 [E2]	16[K3]	27[L1]
MIXED INVOLVING AT LEAST ONE MINORITY ETHNIC GROUP (MEG)				
Mixed MEG + White†	11[B1]	15[E4]	50[K4]	29[L2]
Mixed MEG only†	1[B2]	1	19[K5]	7[L3]
ASIAN OR ASIAN BRITISH				
Indian or African-Indian†	13	2	9	104
Pakistani†	23		3	14
Bangladeshi†	3		7	7
Any other Asian background†	2	1 ('Syrian')	4[K6]	11
BLACK OR BLACK BRITISH				
Caribbean†	18		63	9
African†	3		108	22
Any other Black background†	0		5[K7]	4
CHINESE AND OTHER				
Chinese†	0		4	3
Japanese†	0		1	1
Malaysian, Vietnamese, or Filipino†	0	3		3
N African, Arab, or Iranian†	0	1		
Other origins†	1		4 (Latin American=2; Ecuador=1; Cambodia=1)	6
MISINTERPRETATIONS	0	2[E5]	0	0
MISSING DATA	1	0	4	7
PAGE IN PROFORMA MISSING	0	0	0	2
TOTAL	84	442	553	1068

Note: The coding on the database was: *White* - English, Scottish, Welsh, Irish (A1); Other North European (A2); Greek or Greek Cypriot (A3); Turkish or Turkish Cypriot (A4); Italian, Maltese, or other Mediterranean (A5); Any other White background (A6); *Asian or Asian British* – Indian or African-Indian (C1); Pakistani (C2); Bangladeshi (C3); Any other Asian background (C4); *Black or Black British* – Caribbean (D1); Africa
n (D2); Any other Black background (D3); *Chinese and Other* – Chinese (E1); Japanese (E2); Malaysian, Vietnamese, or Filipino (E3); North African, Arab, or Iranian (E4); and Any other (E5).
Note: † = origins at risk of haemoglobinopathies

Birmingham

B1: A1+A6(French)+C3+D1 [1]; A1+C1+D1 [1]; A1+C2 [1]; A1+D1 [5]; A1+D1+D2 [1]; A5(?)unsure+D1+D3(British) [1]; A6 (not sure)+C4(not sure)+D1+D2+E5(not sure) [1]

B2: C1+C4(Jamaican Indian)+D1 [1]

Exeter

E1: A1+A2[9]; A1+A2 (mother half French) +6(Irish) [1]; A1+ A6 (Breton) [1]; A1+A6(Russian) [1]

E2: A1+A2(Holland)+A3 [1]; A1+A3 [1]; A1+A5 [8]; A1+A5+A6(American) [1]; A2+ A5 [1]

E3: Australia; German; Slovak, East European; South Africa

E4: A1+D3 (Maori) [1]; A1+A2+E5 (North American Indian) [1]; A1+C1 [4]; A1+C4 [Burmese] [1]; A1+D1 [1]; A1+D3 (Egypt) [1]; A1+E4 [1]; A1+E5 (Aborigine) [1]; A1+E5 (Jewish) [1]; A5+E4 [1]; A6 (Persian)+E4 [1]; A6+C4(Japanese)+E2 [1]

E5: Two women had ticked two boxes, one for their own origin and one for their husband ('A1+C4 [husband's father half Burmese]'; 'A1+D3 [partner]')

Kings

K1 Ecuador [2]; African; Australian [3]; East Europe; French; Lithuania; Middle Eastern; New Zealand [2]; Peru; Russian; Slovakian; South American [2]; Spain; Latin America – Colombia; Austrian; Basque; Canadian; French [2]; Latin American; Polish [2]

K2 A1+A2 [12]; A2+A6(White African) [1]; A1+A6(Australian); A1+A6(Polish); A1+A6(South African); A1+A6 (St. Helena)

K3 A1 (Irish)+A3 (Greek) [1]; A1+A2+A6(Jewish) [1]; A1+A3 [4]; A1+A5 [6]; A2+A5 [1]; A1(Irish)+A3(Gk Cypriot) [1]; A1(Scottish)+A5(Italian) [1]; A1+A6 (Middle Eastern) [1]; A6(East Europe)+C4(Iraqi) [1]; A6(not known)+D2 [1]; A6(Portuguese)+D2 [1]

K4 A1+A2+C4(Iran)+E4 [1]; A1+A5+C1 [2]; A1+A5+D1+E1 [1]; A1+A6(English)+D1+D3(Jamaican) [1]; A1+A6(French)+D2 [1]; A1+C1+D1 [3]; A1+C3+D1 [1]; A1+D1 [20]; A1+D2+E4 [1]; A1+D3(Black British) [1]; A3+D1+D3(American) [1]; A5+E4 [1]; A6(African)+D2 [1]; A6(American)+D1 [1]; A6 (French)+d!+D2 [1]; A6(not known)+D2 [1]; A6(Portuguese)+C4(Bombay)+D1 [1]; A1+C1[1]; A1+C4(Sri Lankan)[1]; A1+D2 [4]; A1+D1+D2+E1 [1]; A2+D2[1]

K5 C1+C2 [1]; C1+D1 [5]; D1+D3(Black American) [1]; D1+D2 [2]; D1+D3[1]; D2+D3[1]; D2+D3(British); D2+E1 [1]; C1+D1 [2]; C1+D1+E4 [1]; C2+D1 [1]; C4(Filipino)+E1+E4 [1]; D1+D3(American) [1]; D1+D3 (Black British) [1]

K6 Caribbean; Mauritian; Sri Lanka; Anglo-Asian – white English father, Sri Lankan mother

K7 D3 [1]; D3 (Black British) [1]; Brazilian [1]; Guyanese [1]; Jamaican [1]

Note: In one case the original form was missing & only the retest available (omitted from the table).

Leicester

L1 A1+A3 [2]; A1+A3+A4 [1]; A1+A4+A5(American/East European) [1]; A1+A5 [18]; A1+A5 (Italian ringed) [2]; A1+A5+A6(Canadian) [1]; A2+A5 [1]; A4(Yemen)+A5 [1]

L2 A1+A5+D1 [1]; A1+A6+E3 [1]; A1+C1 [4]; A1+C1+D1 [1]; A1+C1+E3 [1]; A1+C4(Sri Lanka) [1]; A1+D1 [9]; A1+D2 [2]; A1+D3 [4]; A1+E4 [1]; A5+C1 [1]; A6(Morocco)+D3(Morocco) [1]' A6 (German+Asian)+C4(German) [1]; A6+C1+D3+E5 [1]

L3 C1+E3 (Malaysian) [1]; C1+E4 [1]; D1+D2+D3 [1]; C4(Malaysia)+E3 [1]; C4(Malaysian Chinese)+E1+E3 [1]; E1+E3 [1].

There is some evidence in the Exeter data that a few people from the Middle East (an area occasionally referred to as 'West Asia', for example, in Canada) and SE Asian ticked 'Any other Asian background' but nearly always combined this with another more specific category. Examples of this are: 'Any other Asian background (Syrian)'; 'Any other Asian background' & 'Malaysian, Vietnamese or Filipino' both ticked; 'Any other Asian background (Japanese)', 'Any other White background', and 'Japanese' all ticked. Similarly, a respondent had ticked both 'Any other White background (Persian)' and 'North African, Arab or Iranian' categories. One mixed white and Burmese respondent correctly utilised the 'Any other Asian background' option. With respect to specificity of categorisation, the

question appears to have performed very well. Borderline categorisation included the use of D3 for 'Maori' and 'Egypt'.

The Birmingham responses to question A were straightforward and raised no ambiguities or misinterpretations. There was only one response with missing data, constituting a gross error rate of 1.1%.

Question A also performed well in the Kings dataset. The question was especially successful at capturing mixed ethnicity, the responses including a total of 102 mixes (33 mixed white, around half at risk; 50 mixed white & minority ethnic group [included 20 mixed White English… and Black Caribbean]; and 19 mixed minority ethnic group only). Again, the question was successful at capturing Mediterranean ethnicities, 21 respondents ticking A5 ('Italian, Maltese, or other Mediterranean'), 9 singly and 12 in combination with another category. Of the 26 respondents who ticked 'White, Any other background', the descriptions 'Spain', 'South or Latin American' [3]/'Latin American-Colombia'/'Ecuador', and 'Middle Eastern' were possibly at risk but as descriptions given in open response their risk status could be ascertained. There were no examples of ambiguous assignments. Of the three respondents who ticked C4, the two who wrote in Caribbean and Mauritian, respectively, are likely to be cases of complex ethnicity, that is, Indian-Caribbean and Indian-Mauritian rather than misassignments.

The only examples of technical misuse of a category, all of which were explicit in the free text descriptions and none consequential, were: a respondent who ticked A1, A2, C4(Iran)+E4, again someone of Middle East origins who identified as 'Asian' but also ticked Middle East, this removing any ambiguity (for problems with the 'Asian' category, including its use by Middle Eastern and East/SE Asian groups, see Aspinall 2003); a further example of this is a respondent who ticked A6(East Europe) & C4(Iraqi) (the free text description removing ambiguity, as before); a respondent ticking C4 (Filipino) & E1 & E4 is yet a further example of this misuse, again made explicit by the free text descriptor; finally, a respondent used the C4 category to describe mixed race: 'Anglo-Asian – white English father and Sri Lankan mother'.

In the Leicester data a similar use of the residual 'Any other Asian background' (to that found in the Exeter data) occurs in a very small number of responses. Of the 15 respondents utilising this category, most were Indian subcontinent: Sri Lankan [n=4], and Hindu, Indian Hindu, Indian, Sikh, and Nepalese [n=6 in all]; others were Afghanistan [1] and 'mixed' [1].

However, 3 respondents who were Malaysian or Malaysian Chinese used this category to write in their origins in free text but ticked the relevant categories (E1 and E3) as well. Similar examples of multi-ticking to declare a single origin might be 'A6(Morocco) and D3(Morocco)' and 'D2 and D3(Somali)'. None of these examples compromise risk status ascertainment. Moreover, there were no other misinterpretations that were consequential for risk status ascertainment. On the three occasions where the term 'South African' was used (in the 'White-Any other background' category), the race status of the person was clear. One respondent identified as 'Egypt' in this same (open response) category but, clearly, the description itself permitted risk status ascertainment (and this was also the case with a respondent identifying as 'Macedonian'). Another woman wrote in 'Morocco' in this category and also ticked the 'Any other black background' category and wrote in 'Morocco'. Of the 3 respondents identifying as 'Somali' in the Black or Black British question set, all wrote in this term. Thus, the Leicester data strongly supports question A in terms of sensitivity. Moreover, a total of 35 women identified as 'Italian, Maltese, or other Mediterranean' in single or mixed responses.

Conclusion

Question A performed very successfully in all four sites. There was missing data for only 12 respondents (that is, people recruited to question A who did not tick a box) out of a total of 2147 who answered question A, a rate of 0.6% that is tolerable. Moreover, the misinterpretation rate (n=2, 0.09%) related to women who ticked a box for their partner, not a misinterpretation of the ethnic question as such but one of *subject* status.

Question B

Part one

Question B contains within it an evaluative criterion as respondents were given the option in part one of three response options (yes, no, don't know) to the question: 'Do you or any of your ancestors, as far back as you can recall, have ethnic/family origins from areas of the world outside of the United Kingdom or Republic of Ireland?'.

Table 8: Response options to question B (part 1)

	Birmingham	**Exeter**	**Kings**	**Leicester**
Yes	48	73	306	277
No	16	358	179	709
Don't know	7	16	26	38
Ethnicity information refused	0	0	0	1
Blank	0	4	13	10
Total	71	451	524	1035
Gross error rate (don't knows + blanks)/total	7 (9.9%)	20 (4.43%)	39 (7.44%)	49 (4.7%)

In Birmingham 48 respondents (67.6%) stated that they or their ancestors had ethnic/family origins from areas of the world outside of the United Kingdom or Republic of Ireland and 16 (22.5%) that they did not. 7 respondents (almost a tenth) did not know (table 8).

In Exeter 73 respondents to question B (16.2%) declared that they or their ancestors had ethnic/family origins from areas of the world outside of the United Kingdom or Republic of Ireland and 358 (79.4%) that they did not (79.3%) [although one of these appears to be equivocal: 'E.'s mother is adopted? GI babe']. Of the 16 who stated that they didn't know 3 gave reasons (in annotated comments on the schedule): 'My sister has pigmentation in her skin with a querie of where her father originates from. Eastern?'; 'Adopted'; and 'Maternal grandfather unknown'. It is not known how many of the 'don't knows' didn't have complete knowledge of their ancestry; the cognitive research suggests that the number was probably small, although one of the respondents gave this as a reason in the annotations. The total gross error rate of 4.4% in *part one* of the question is of concern.

Amongst Kings respondents, 306 respondents (58.4%) stated that they or their ancestors had ethnic/family origins from areas of the world outside of the United Kingdom or Republic of Ireland and 179 (34.2%) that they did not. 26 respondents (4.96%) did not know, including one who stated she was adopted at birth. Part one of the question in a further 13 schedules was left blank.

Finally, in Leicester, 277 respondents (26.8%) stated that they or their ancestors had ethnic/family origins from areas of the world outside of the United Kingdom or Republic of Ireland and 709 (68.5%) that they did not. 38 respondents did not know, including one who stated 'mother was adopted out at birth, biological family not known'. Part one of the question in a further ten schedules was left blank.

Conclusion

On part one of question B a gross error rate of 5.53% was incurred amongst the 2081 women recruited into the trial who answered this question. Most of this error (4.18%) was attributable to 'Don't knows' but 1.30% to missing answers. This is a substantial error rate – close to 6% - bearing in mind the kind of error ranges that Zeuner *et al.* (1999) were modelling. This error rate, of course, excludes errors that arose in part 2 of the question.

Part two

The responses to question B, given in the five free text boxes, have been mapped back to the same ethnic categorisation used to report the findings of question A. This has been undertaken using the full coding frames for the 2001 Census developed by the Office for National Statistics and the Department of Health and the author's personal knowledge of ethnic categorisation. However, this does not mean that there is an exact conceptual equivalence between the two sets of data. Question A relates to ethnic family origins located within broad pan-ethnic (and, to some extent, colour based) groupings used in the 2001 Census. Question B relates primarily to countries of family origin (reported without reference to the broad-based groupings used to structure question A). For example, someone reporting family origins in South Africa could belong to any of several ethnic groups found in that country. Further, multi-ticking on question A would identify (1) persons of conventionally defined mixed race (e.g. those who ticked English, Scottish, Welsh, or Irish and Caribbean) and (2) persons of disparate mixed ethnic (rather than racial) origins, e.g. those ticking English, Scottish, Welsh, or Irish and Italian, Maltese, or other Mediterranean. Mixed descriptions on question B cannot easily be read as either conventionally defined 'mixed race' or disparate mixed origins. Such descriptions may represent either `multiple ancestral origins or merely complex ethnicity (as where, for example, a respondent writes in 'Caribbean' and 'American' or 'Trinidad' and 'India'. It is frequently difficult to infer that such descriptions relate to mixed 'population groups'.

Table 9. Responses to question B ascertaining ethnic/family origins

	Birmingham	Exeter	Kings London	Leicester
WHITE				
English/Scottish/Welsh/Irish	0	3 (all Irish)	2 (Ireland)	4
Other N European	0	28	29	29
Greek/Greek Cypriot & Turkish/Turkish Cypriot†	0	2	4	3
Italian, Maltese, or other Mediterranean†	0	5	9	19
Any other White background	2	3	14	6
Mixed white – not at risk	0	6	18	9
Mixed white – at risk†	0	1[E1]	3	3
MIXED INVOLVING AT LEAST ONE MINORITY ETHNIC GROUP (MEG)				
Mixed MEG + White†	7[B1]	4[E2]	31[K1]	11
Mixed MEG only†	4[B2]	2[E3]	36[K2]	15
ASIAN OR ASIAN BRITISH				
Indian or African-Indian†	5	3	12	80
Pakistani†	12		2	4
Bangladeshi†	6		1	7
Any other Asian background†	0	3	3	5
BLACK OR BLACK BRITISH				
Caribbean†	8	4	26	15
African†	3	1	102	36
Any other Black background†	0			1
CHINESE AND OTHER				
Chinese†	0	2	1	4
Japanese†	0			0
Malaysian, Vietnamese, or Filipino†	0		2	7
N African, Arab, or Iranian†	1		5	6
Other origins†	0	1	3	1
MISINTERPRETATIONS	0	2		2
MISSING DATA (TO 'YES')	0	2	3	17
TOTAL	48	73	306	284

Note: † = origins at risk of haemoglobinopathies

Note: Responses for question B have been mapped to the above categories based on the woman's responses. For example, if a woman wrote only 'Central America (Belize) mother', the response would be assigned to Black Caribbean (the assumption would not be made that the woman had a white father and that she is mixed Black Caribbean & White as no other information on her ancestry was supplied). It is therefore a classification of ethnic family origins by country of origin.

Birmingham

[B1] 'Jamaica+Jamaica+ Spanish'; 'Jamaica + Malta'; 'Jamaica+America+Canada'; 'Jamaica+Canada'; 'Jamaica+ Cuba+ Mexico+America'; 'Jamaica+USA'; & 'Angola+Portugal'.

[B2] 'Pakistan + India'; 'Jamaica + Africa'; 'India + Jamaica'; & 'Barbados+St Vincent+Trinidad+Africa'.

Exeter

[E1] 'Father is from Italy (North), Mother is from German descent'.

[E2]'India and America' (n=2). 'Great grandfather South African, Great Grandfather French' (n=1); and 'I was born in Uganda, East Africa, my father is a German Jewish Refugee, and all his family were German Jewish' (n=1)

[E3] 'Zimbabwe, South Africa, and Malaysia'; 'Egypt, South Africa, and India'.

Mixed white at risk Eastern Europe+Jewish[1]

Kings

K1 Africa+France+Jamaica+India [1]; Barbados+Canada [1]; Cameroon+Germany [1]; England+Wales+Barbados [1]; Guyana (South America)+ Dutch [1]; India+Netherlands [1]; Ireland+Jordan [1]; Ireland+Wales+China+England [1]; Jamaica + America [1]; Jamaica+Ireland [1]; Jamaica+Ireland+Scottish+Jewish [1]; Jamaica+USA+Africa+Trinidad+Barbados+Australia [1]; Mali+Ivory Coast+France [1]; Martinique+Congo+France [1]; Norway+Jamaica [1]; Portugal+Angola [1]; Scotland+Brazil+Portugal [1]; South America –Guyana+Scotland [1]; Zambia – father+Switzerland-paternal grand mother+Germany – paternal grand father [1]; France & Germany & Poland&Russia&Middle East somewhere [1]; Africa, Nigeria & Turkey [1]; Iran+Germany+Russia [1]; Jamaica, India, Chinese, African, Scottish [1]; Jamaica, Scotland [1]; Jamaica+America [2]; Jamaica+China+France [1]; Jamaica+Cuba+America [1]; Jamaica+Cuba+Ireland [1]; 'Jamaica+England+America+Canada+Bahammas [1]; Nigeria+London+America [1]; Portugal&Brasil&France&England&Canada [1]; South Africa – black South African + Scotland – white Scottish [1]

K2 Antigua+Jamaica [1]; Barbados+Jamaica [1]; Grandmother-Jamaica+Great grandmother – Portugal+Dad's mother half-Chinese [1]; Grandparents Jamaica (Mum's side)+grandparents Barbados (dad's side)+ Great grandmother Jamaica (mum's side) [1]; India+Africa [1]; India+Kenya+Burma [1]; India+Pakistan [1]; Jamaica+St Kitts+Dominica [1]; Jamaica, W Indies+India+Hong Kong+Africa [1]; Jamaica+Barbados [1]; Jamaica+Barbados+Cuba [1]; Jamaica+China [1]; Jamaica+Cuba+Morocco [1]; Jamaica+Grenada [1]; Jamaica+St Kitts [1]; Jamaican+St Lucian [1]; Kingstown St Vincent+Cape Town South Africa [1]; Nigeria+Jamaica+Cuba [1]; Sierra Leone+Nigeria+Caribbean [1]; Uganda+Sudan [1]; Zambia+Mozambigue [1]; Zimbabwe+South Africa+Zambia [1]; American (African American), African+Native American [1]; India+Africa [1]; Jamaica+Cuba [1]; Jamaica+Guyana [1]; Jamaica+Guyana [2]; Malaysia, China (1); Rwanda+Uganda [1]; Senegal+Saudi Arabia+Gambia [1]; Sierra Leone/Africa+Lebanon/Middle East [1]; Tanzania+Kenya [1]; Zimbabwe+Botswana [1].

Although limited to only 48 cases, Question B in Birmingham performed satisfactorily (table 9). There were no misinterpretations, missing data, or consequential ambiguous descriptions. Overall, just two respondents (4.2%) referred to particular ancestry (parents & grandparents) while another two completed the five boxes to suggest that they were interpreting the question as ancestry ('Iraq + Iraq + Iraq + Iraq' and 'Jamaica + Jamaica + Spanish'). Some of the descriptions contain ambiguities but none are consequential, e.g. 'Zimbabwe-Africa+South Africa' could be a mixed race/heritage identification if the South African heritage was white; similarly, the 'America' response in 'Jamaica, Cuba, Mexico, America' could be white or black; and 'Jamaica+Africa' could be a mixed heritage in two black groups or ancestry in Africa. 'India + Jamaica' is more problematic as it could imply complex ethnicity – that is, Indian Caribbean origins - or a mixed heritage of Indian and Jamaican ethnic origins, both having somewhat different risk statuses.

Question B in Exeter also performed reasonably satisfactory. 20 respondents (27.4%) answered this question with specific reference to ancestors (e.g. 'great grandmother Spanish'; 'maternal granny's side Russian blood'; 'father is German', 'grandfather American'; 'French great grandmother'). 3 respondents wrote in 'Republic of Ireland' although such responses should have been screened out in part 1 of the question. Two respondents gave the country of

family origin of their husbands only (e.g. 'partner from Venezuela, South America'). There were two ambiguous descriptions ('Great grandfather South African, Great Grandfather French' and 'I was born in Uganda, East Africa, my father is a German Jewish Refugee, and all his family were German Jewish'). It is clearly not possible to infer ethnicity from a 'South African' country of origin and this is a recurring problem in responses to question B. With respect to the second example, the respondent could have been the daughter of an expatriate German Jew living in East Africa). There were also two respondents who answered 'yes' to part one but did not write in a description. Thus, the gross error rate for this question was 8.2% (12.3% including the 'Republic or Ireland' responses). There were also several other examples that suggest uncertainty with respect to risk status, for example, 'Spain, Argentina'; 'Italy or Spain'; 'Zimbabwe, South Africa, and Malaysia'; and 'Egypt, South Africa, and India', although these are all likely to be 'at risk'.

Amongst the Kings data, a total of 307 persons gave free-text responses to Question B, including 3 blanks. Many of these responses were straightforward, for example, 12 women wrote in the word 'India', 34 the word 'Jamaica', and 31 the word 'Nigeria', and 8 the word 'Sierra Leone'. A small number identified as just 'Africa' and a few Africa and a country name (e.g., Africa [Nigeria], Africa [Zambia]) but most Africans gave their country. There were many complex answers involving several countries of origin. As in the other sites, some respondents (a total of 11) answered this question with respect to particular ancestors ('Grandad Morocco. Nan Morrocain'; 'Grandmother German'; 'Grandmother – Jamaica+Great Grandmother – Portugal + Dad's mother half-Chinese'; 'Grandparents Jamaica (Mum's side)+grandparents Barbados (dad's side)+Great grandmother Jamaica (mum's side)'; Mum – Jamaica+Dad Jamaican+Gran Jamaica'; on my father's side I know they are Dutch but I don't know how far back'; and 'We all stem back from the USA to my great, great Grandparents. Their parents were from Ireland & Scotland'; 'Zambia – father+Switzerland – paternal grand mother + Germany – paternal grand father'; 'Grandparent (West Africa) + Immediate Parent (West Africa); Grandparents St Lucia; 'Mother – South Africa (Zimbabwe) all from mother's side + father – Africa (Malawi)'; and 'Myself & all my family (mother, father & grandparents) are from Sweden'). As with question B in other sites, many of the descriptions given are ambiguous and may be indicative of mixed origins or complex ethnicity, e.g., 'Trinidad + India' and 'India + Africa'.

The most extensive data for question B came from Leicester: 284 responses by respondents who had answered 'Yes' to part 1. However in 15 cases the five free text boxes had been left blank and in two others 'unsure' had been written – 17 in all or 6.0% of all responses. Further, 2 respondents had given a response for their husband/partner (husband's grandparents' ethnic

origin & partner's ethnic origin). In three cases the midwife had substantially assumed the task of ascription ('Client didn't read question but didn't mind me filling in her ethnic status'; 'He was black but I don't know were he came from'; and 'It (the question) was too wordy for client therefore I simplified it').

A total of 21 respondents (7.4%) answered the question with reference to specific family members/ancestors: parents, grandparents, and great grandparents. These responses have been mapped to the categories in table 9. There were a large number of descriptions open to varying interpretations but that were not consequential as sufficient information was given to identify them as at risk, for example, 'Africa – me, India – parents, Pakistan – grandparents'; 'Africa & Scotland', 'Asia and Africa', etc. With respect to many of these descriptions it is not possible to be specific about the nature of the risk (especially descriptions suggesting East African Indian origins – such as 'Africa – Malawi & India' – but which could also indicate mixed family origins). However, there were others that could be consequential, including those that depended on interpretation of the term 'South Africa': 'Canada, South Africa, & Australia'; 'England & South Africa & Germany'; 'South Africa & France'; 'South Africa, Malta, & Russia'; and 'MGM Austria'. Although the South Africa black population has a relatively low incidence of clinically relevant haemoglobinopathies, high geographical mobility related to labour market conditions and the consequences of war and poverty amongst black African countries qualifies South Africa as an "at risk" country. Perhaps a total of 5 of the Leicester question B responses could be interpreted as 'categorical' consequential misinterpretations

Conclusion

Question B performed reasonably well in all 4 sites. However, there is a greater potential for error in this question as it is in two parts. Part one serves the purpose of "screening out" those who do not have any known ancestry from areas of the world outside of the United Kingdom or Republic of Ireland. There is in the set of response options to the first part a 'don't know' category. Secondly, errors may occur in part B where the respondent is invited to write in the countries outside the UK and Republic of Ireland from whom her known ancestors came (figs. 5 & 6). Errors arising in part one of the question have already been reported.

Errors in part two comprised missing data and misinterpretations (figs. 5 & 6). There were a total of 22 missing responses for respondents who had ticked 'yes' in the screening question out of a total of 709 responses (or 3.1%, over five times the rate for question A). Moreover, misinterpretations numbered 4 (or 0.68%, over seven times the rate in Question A). Further,

there were consequential categorical ambiguities in Question B (for example, five in the Leicester data).

Fig. 5:

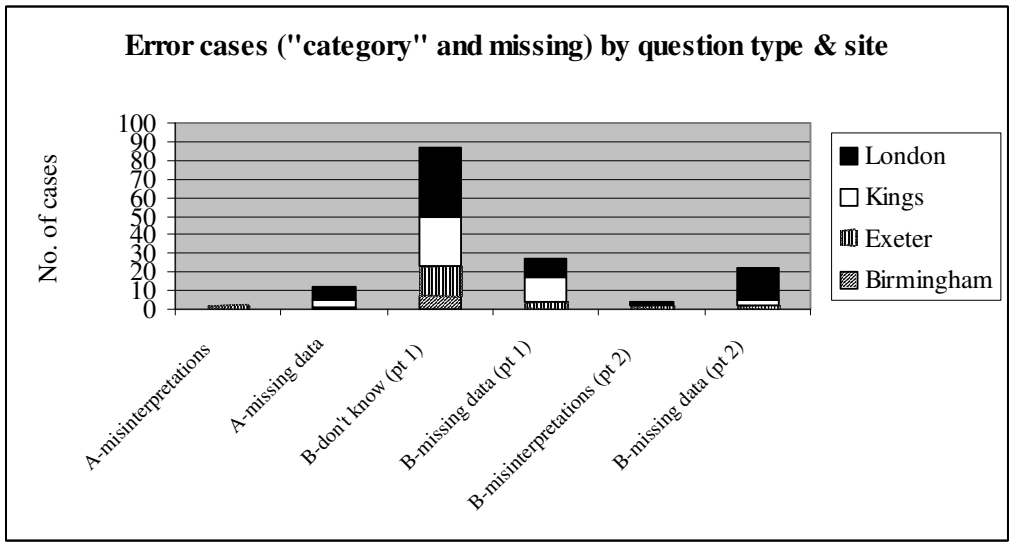
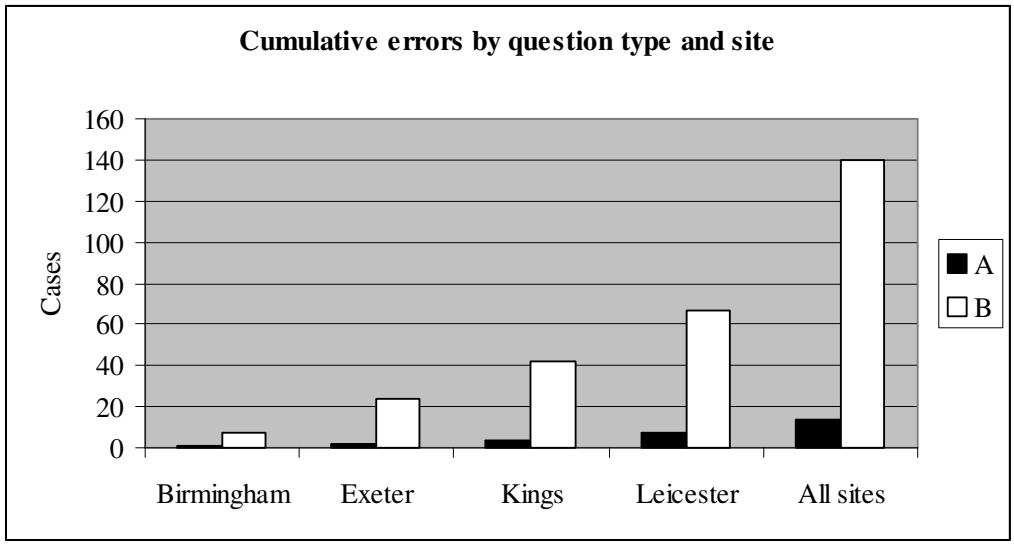


Fig. 6:



Problems of misinterpretation in Question B arise from a number of factors. Most importantly, part B of the question asks for names of the *countries* of origin of known ancestors for those whose ancestry is not confined to the United Kingdom or the Republic of Ireland. Although the question's conceptual base is ethnic/family origins, question B can only access such origins in part 2 by requesting the countries. However, there is not a simple read-across from ethnic origins (those that are of interest in terms of risk status ascertainment for the haemoglobinopathies) and countries of origin. Some countries have many ethnic groups, some of which may be at high risk, low risk, or negligible risk). In answering part B of the question the respondent will be motivated primarily to list *countries* of ancestral origin without putting those countries through a 'sieve' to assess whether they can be read in terms of risk status. Perhaps the most problematic of these is South Africa whose population is around 76% black, 13% white, 8% coloured, and 3% Indian. According to the 2000 US Census around 75% of the population of the United States in white, 12.3% Black or African-American, and 3.5% Asian. Similarly, the Canadian population is multicultural, comprising (according to the 1996 Census), a total visibility minority population of 11%, 3% of the population being Chinese, over 2% South Asian, and 2% Black. Someone answering 'USA', 'America', 'Canada', or 'South Africa' tells us nothing about their *ethnic* origin. They could be White, or Black, or have some other ethnic origin. This difficulty was not identified by professional staff participating in the study, except in the case of two observations in Exeter: (i) For a woman writing in 'Zimbabwe' in part 2 of Question B – 'Is this lady Black African or White African'. (ii) 'This style of question does not gather if a person is Black Australian or White Australian or Black African or White African'.

Another problem is the possibility that ancestral origins in the former colonies of European countries may be indicative of a white rather than black or other minority ethnic group origin. Some examples in the data (for example, responses comprising an ancestry that includes Germany and one or other of the former German East African colonies raises concerns about *ethnic* vs. *country* origins). Clearly, it is possible that some of these ambiguities would be resolved on face to face contact with the woman booking but there are problems of trying to infer even broadly-based ethnic origins by observation and such efforts are subject to substantial error in mixed heritage populations.

A third problem with interpretation of question B responses is that of staged migration. Someone identifying their *countries* of ancestral origin as Kenya and India tells us nothing of their *ethnic* origins except by inference. Such persons are likely to be East African Indians, born in the Indian subcontinent but migrants to East Africa. However, a listing of such ancestral countries of origin could also be descriptive of biological mixing, a person of dual

heritage. Having to infer race/ethnicity from countries of ancestral origin was a recurrent problem in trying to map back these responses to the classification of *ethnic* origins given in table 9. Inevitably, some of these judgements are probabilistic and, consequently, subject to error.

Finally, one cannot infer *mixed* ethnic origins from data for question B as the question only asks for countries of ancestral origin outside of the United Kingdom or Republic of Ireland. We may know, for example, that a person has ancestry in Nigeria but it is not possible to say whether that person is mixed race (that is, White British and Black-African).

In conclusion, this question can probably provide a reasonably accurate means of identifying the 'White Northern European' (or White British and Irish) and 'all other origins split' but is much less satisfactory at identifying *specific* risks associated with defined ethnic origins (as reflected in the various trait prevalences for the different haemoglobinopathies which have been presented by ethnic group (Davies *et al*., 2000)).

5.5. The congruence of language and ethnic family origins

Both proformas contained 3 questions on language: 'What was the client's preferred language for this booking-in interview?' [options of Arabic, Bengali, Chinese, English, French, Gujarati, Portuguese, Punjabi, Somali, Turkish, Urdu, Any other language (free text), Don't know, & Could not establish]; If the client's preferred language was not English, how was the interview conducted? [options of Professional interpreter, Other health worker, Other health professional, Relative, Bilingual Midwife, Language Line, Other (free text); and 'What problems, if any, do you think arose in cases where a client did not speak English as a first language?' (free text). The information on language has been used to establish concordance or discordance with ethnic/family origins. Clearly, someone ticking, say, A1 (English, Scottish, Welsh, or Irish) on question 1 *or* answering 'No' to part one on question B (that is, the respondent has no known ancestors with ethnic/family origins from areas of the world outside of the UK or Republic of Ireland) but who then also ticks, say, 'Bengali', has given an incongruent response (that is, a categorical error). Congruence could be established using the client's preferred language for the booking-in interview, the manner in which the interview was conducted if the client's preferred language was not English, and the recording of any problems that arose in cases where a client did not speak English as a first language.

Table 10. Ethnic Origin and Language Congruence.

Language congruence	***Birmingham***	***Exeter***	***Kings***	***Leicester***
Strong *A* *B*	 **84** **62**	 **443** **421**	 **500** **479**	 **1036** **925**
Equivocal/Incongruent *A* *B*	 **0** **2†**	 **2*** **6****	 **2#** **9~**	 **1+** **30"**
Language missing *A* *B*	 **2** **0**	 **36** **23**	 **51** **37**	 **30** **31**
	150	**931**	**1078**	**2053**
Gross errors (excl missing from numerator & denominator)	**1.35%** **A=0%** **B=3.13%**	**0.92%** **A=0.42%** **B=1.41%**	**1.11%** **A=0.40%** **B=1.84%**	**1.56%** **A=0.10%** **B=3.14%**

Note: A cases are those with a valid code. B cases are those with a valid entry (that is, 'No' or 'Yes' with accompanying free text, but excluding 'don't know').

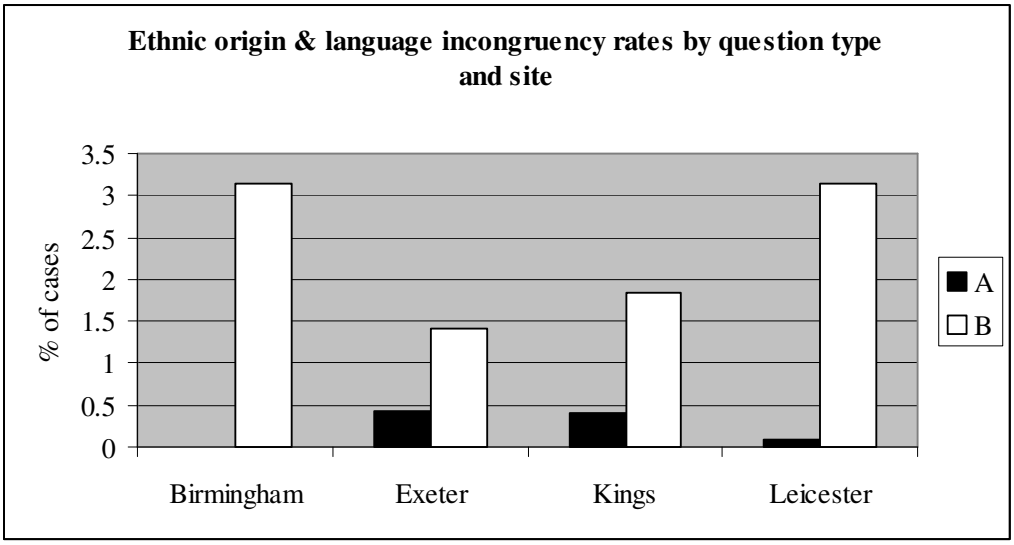
Language codes: Client's preferred language for this booking-in interview: 1 Arabic; 2 Bengali; 3 Chinese; 4 English; 5 French; 6 Gujarati; 7 Portuguese; 8 Punjabi; 9 Somali; 10 Turkish; 11 Urdu; 12 Any other language [free text field].

† Cases: (i) Part 1 B=No; language=Bengali; interview conducted by 4(husband). (ii) Part 1 B=No; language=Bengali & Punjabi; interview conducted by 2 (other health worker); annotated note: ' patient born in India, husband born in UK. Ancestors are from India'.
*Cases: (i) A1 + comment: 'Chinese was her first language – her understanding was good but I found some difficulty understanding her accent'; (ii) A1 + 4 (English) but interview conducted by interpreter (husband) & comment 'her husband was a very good interpreter'.
**Cases: (i) B – No + 4 & comment 'speaks fluent English & answer with discussion with partner'; (ii) B – No + 4 & comment ' Good English spoken. N. did not understand the question on previous page. Should I have corrected her? No translations to Thai in folder'. (iii) B – No + blank & comment 'Good English' (interview conducted without translator). (iv) B – No + 4 & comment 'Very good English spoken' (interview conducted without translator). (v) B – No + 12 (Any other language: Thai) (interview conducted without translator). (vi) B – No + 4 & comment 'K had difficulty understanding the ethnicity qn. Due to poor reading skills - K is British, half British ancestors'.
Cases: (i) A1 + 3 (Chinese). (ii) A1(English circled)+3 (Chinese)
~ Cases: (i) [Equivocal] B – Yes (Jamaica, St Kitts & Dominica) +3(?); (ii) B – No + 4 (+Tegern) & comment 'Client finding the question difficult to understand'; (iii) B – Yes ('Asian') + 3(Chinese); (iv) B – No + Don't know + 3 (Chinese); (v) B – Yes ('Africa') + 3(Chinese); (vi) B – No + 1(Arabic); (vii) B – No + 7 (Portuguese) & Brasilian; (viii) B – Yes ('Democratic Republic of Congo') + 6 (Gujarati); (ix) B – Yes ('Ghana') + 3 (Chinese).
+ Cases: A1+3(Chinese).
" Cases: (i) B4 – No + Tamil; (ii) B – No + (4+9 [Somali]); (iii) B – No + Sudan (interview conducted without translation); (iv) B – No + Indi (relative interpreter); (v) B - No + (blank preferred language) & interview conducted 'other'; B – No & comment 'explanation through interpreter; agrees through interpreter'; (vi) B – No & comment ' speaks very little English. Possibly not understood impact of results positive'; (vii) B – No & comment 'P's husband does not know anything about sickle cell/thalassaemia – patient needs an interpreter'; (ix) B – No & comment 'Reasonable level of English…words to long and complicated'; (x) B – No + 6 (French); (xi) B – No + 4 & comment 'Husband understood shortened version of explanation. Question read out to them'; (xii) B – No + 4 & comment 'Explained through interpreter'; (xiii) B – No + 4 & comment 'Explained through

interpreters'; (xiv) B – No & comment 'Client's husband interpreter. Booking took longer due to husband used as interpreter, lady very anxious re blood tests'; (xv) B – No + 4 & comment 'Fiance interpreted and appeared fairly supportive & to explain information in detail'; (xvi) B – No + 5 (French); (xvii) B – No + 4 & comment 'Friend interpreted. Better that explanation is shorter now. Translated explanation for interpreter slightly'; (xviii) B – No & comment 'Managed to speak & understand English'; (xix) B – No & comment 'bilingual mother – adequate English'; (xx) B – No + 5 (French); (xxi) B – No + 4 + relative interpreter; (xxii) B – No + 4 + relative interpreter & comment 'unsure how much was understood'; (xxiii) B – No + 4 & comment 'R speaks English…no problems'; (xxiv) B – No + 4 & comment 'no problems – speaks good English'; (xxv) B – No + 5 (French); B – No +4 & interpreter ('husband helped'); (xxvi) B – No + 4 + professional interpreter & comment 'very difficult for professional interpreter to explain the research questionnaire – but P just said yes to taking part in the end because they said it may help'; (xxvii) B – No + 4 + professional interpreter & comment 'the client would not be fully informed of all consequences'; (xxviii) B – No + Sri Lankan Tamil; & (xxix) B – No + 4 & comment 'Understands English more than speaking it. Understands written English – able to read information letter'.

Conclusion

Fig. 7:



Language data was missing on 119 cases of Question A and 91 cases of Question B (the identical language question set was asked on both schedules) (table 10). There was evidence of a lack of congruence (strong or equivocal) between assignment on the ethnic questions and language in a total of 52 cases (fig. 7). Ethnicity/language incongruence was much higher on Question B (2.43%, 47/1934) than Question A (0.24%, 5/2063). Language incongruence on Question B was highest in Birmingham and Leicester (at 3.1%), both centres with significant numbers of respondents of South Asian ethnic origin. Part 1 of Question B may have been subject to misinterpretation (as invoking, say, conceptualisations of 'Britishness' or nationality). It is notable that language incongruence on Question B was much lower in Kings (indeed, the third lowest across the 4 sites) whose minority ethnic group population is primarily Afro-Caribbean.

5. 6. Test-Retest on the ethnicity question

There are a variety of ways in which a trial of an ethnicity question (that is used as a criterion of selectivity for risk status ascertainment) can incorporate measures of the utility of that question. In this particular clinical context, there are a number of essential measures. Firstly, in a conventional setting of a test for a particular disease, one of the most important statistics is *sensitivity* [that is, the ability of the test to identify the disease when it really is present (as assessed by a gold standard test] – that is, the proportion positive of those who have the disease] ; *specificity* [that is, the ability of the test to identify the absence of the disease when the disease really is not present – that is, the proportion negative of those who do not have the disease] is also important but less so than sensitivity. These two statistics and others (such as the positive predictive value of the test) that can be derived from a matrix of results of the test vs. a gold standard give us a measure of the discriminant power of the test to detect cases of the disease. However, in this context, the use of such measures is problematic and gives an incomplete picture. This is because we are interested in how well the candidate ethnic questions identify risk status for carrying a haemoglobinopathy trait and not, specifically, how well the ethnicity questions capture those who have clinically relevant haemoglobinopathies as determined by laboratory testing (the actual trait carriers). Thus, measures of sensitivity and specificity cannot be calculated. The only statistic that can be derived is the proportion of trait carriers who were not identified as at risk by the ethnicity questions.

The trait carriers in any population of antenatal women booking will only be a very small proportion of women at risk because of their ethnicity (that is, the standard error will be high).

The sample statistic based on analysis of carrier status data will therefore be imprecise. Within the parameters of trait prevalence, carriers are, effectively, only a sample of all those at risk and the majority of those at risk will not be trait carriers. What we also need, therefore, is a measure of how well the candidate ethnic questions predict true risk status (that is, what the woman's real ethnic/family origins are, based on a gold standard test, such as a retest interview or validation survey of ethnic/family origins. Some further measure is needed of how well the ethnicity question performs for those cases where the woman's ethnic family origins define her as at risk but for which the laboratory tests did not identify a clinically relevant haemoglobinopathy (as any of these cases *potentially* could have been trait carriers). Clearly, the issue that is important is whether the ethnic family origins given by the woman accurately reflect those origins or were given in error.

In cross-sectional data this can be assessed by a variety of means (such as that of cognitive research used in this study). However, in quantitative data involving approaching 5000 cases cognitive research can only indicate the range of problems identified in a subset of women interviewed. The standard measure for assessing the stability of responses in Government social surveys is *test-retest reproducibility (validity),* an approach usually adopted in Census Validation Programmes (including that for the 1991 Census) (Aspinall 2001). The test-retest methodology enables net and gross error rates in question responses to be identified by comparing initial responses with those on retest. The numbers of incongruent findings are too small to justify gross error rates for different ethnic family origins.

Test-Retest Results

Table 11: Test-Retest Results

	Birmingham	***Exeter***	***Kings***	***Leicester***
1. ALL TEST-RETESTS:	36	158	83	291
Midwife's quality grading:				
1	31	144	56	276
2	0	5 (A=5;B=0)	2	7(A=2; B=5)[L1]
3	0	0	2	1 (B=1)
4	2 (B=2)	1 (B=1)	5	0
5	0	3 (A=1; B=2)	0	0
Missing	3(A=1; B=2)[B1]	5 (A=3; =2)[E1]	18(A=10; B= 8)	7(A=3; B=4)
2. RETESTS INVOLVING CLINICALLLY RELEVANT HAEMOGLOBINOPATHIES:	8 (excl. one case of 'MCH<25 with ethnicity unknown)	1 respondent only: a known sickle cell trait	46	38
Midwife's quality grading:				
1	8	1	33	33
2	0	0	1	2
3	0	0	1	0
4	0	0	3	0
5	0	0	0	0
Missing	0	0	8	3
Consequential errors question A	0	0	3	0
Consequential errors question B	4	4	16	1

Note: The content re-interviewer was invited to grade the retest against original test result for the ethnicity question on a scale of 1-5:
1: Very Good – A single box (or multiple boxes for mixed origins) was ticked in the original interview and the re-test has given me no reason to come to any other conclusion.
2: Good – Person has attempted to show their true origin but did not feel the question catered for them. Answer given did not affect allocation to at-risk/not at-risk status.
3: Fair – Person gave answer that was inconsistent or was at variance with interviewer's understanding of the 'true' answer. Answer given did **not** affect allocation to at-risk/not at-risk status.
4: Poor – Person gave answer that was inconsistent or was at variance with interviewer's understanding of the 'true' answer. Answer given **did** affect allocation to at-risk/not at-risk status.
5: Very poor – Information was missing on original form. Person acknowledged incorrect answer was on original (e.g. form-filler mistake). Answer given **did** affect allocation to at-risk/not at-risk status.

[B1] All 3 responses were the same on test-retest (ie. Grade 1).
[E1] All 5 responses were the same on test-retest (ie. Grade 1).

On test-retest in Birmingham, question B performed much more poorly than question A with evidence of significant errors or uncertainty on 4 responses (table 11). Of the two instances graded '4' on question B, one woman had answered 'no' to question B on test and retest. The midwife had added: 'Patient Black Caribbean, obviously not answered correctly' and 'family all British' (which may explain the patient's misinterpretation). In the other instance, where the woman had answered 'don't know' to both test and retest, the midwife had added; 'This patient was black Caribbean and obviously didn't understand the question…different answer not given [on retest] but did not answer appropriately. Did not know where ancestors came

from. Ancestors may have come from Caribbean but not sure'. In another response to question B, the woman had answered 'don't know' to both test and retest with the comment 'mother died – have no other relatives so don't know' (it being likely that she would have assigned on question A as it is explicitly focused on ethnic origin). In a further response a woman entered 'no' to question B on both test and retest but worryingly added: 'Can only answer what known. Same answer'. It is clear that some women are interpreting this question's request for information on *ancestors* in an exact way and failing to give a response if ancestry is unknown (even though other clues to that person's ethnic origins may be evident). In the case of the 8 clinically relevant haemoglobinopathies (22% of the test-retests), all the responses were concordant (and there were no 'don't knows' on question B).

The Exeter data yielded a total of 158 test-retest results. Again, the results for question B were less favourable than for question A. One of the responses to question B was 'Don't know (maternal grandfather unknown) on test and 'Don't know, doesn't know about maternal Grandmother' on retest (graded 1 by the midwife). Of those assigned 2-5 the cases varied in terms of consequence. In a 5-graded B response, the woman declared 'Great grandmother possibly Jewish' on test and on retest 'B: No', the midwife adding 'Thought grandmother may have been Jewish...has since asked around family & found she is not'. One woman more seriously answered 'no' to question B but on retest gave 'Yes, India'. The researcher's comment on the initial assignment was: 'Based on this woman's name (which sounds Asian) I contacted the community midwife to ask about this woman. The midwife confirmed that she "looks Indian" but has answered "No" to question B! This could prove interesting at Retest'. Indeed, as the re-tester recorded: 'When questioned at re-test this lady initially answered "No" again but when I re-phrased the question she immediately understood & changed it to "Yes" as she is of Indian origin. She appeared to have made the assumption that only women of known ethnic origin were being questioned & then asking about any other relations!'. Moreover, the schedule recorded that no specific sickle cell or thalassaemia screening had been requested, the midwife allocating a grade 4. A respondent to question B reported 'France' in the test and retest, the midwife noting that she: 'Initially forgot about French connection! This is on her Grandmother's side, a long way back. Corrected her answer when reminded about original answer' (graded 2). A woman gave the response 'A1+C4(Burmese - husband's father half Burmese)' on the test but A1 only on the retest, the midwife adding: 'Did not fully understand the question. Thought it related to her and her husband & their combined effect on the fetus. Clearly an articulate & intelligent woman, this confusion is not related to literacy skills' (graded 5, with a request to make the question clearer). In a further equivocal response to question B, a woman reported: 'I was born in Uganda, East Africa and my father is a German Jewish Refugee and all his family were German Jewish'. In the retest

she stated simply: ‘Yes, European-German’ (graded 1 by the midwife, although involving uncertainty with respect to birth in Uganda in part one of the question). Finally, a woman responded to part one of question B affirmatively but only with a question mark as free text, the researcher's note recording ‘ambiguous answer..will retest’. On retest ‘Yes’ was affirmed and the midwife recorded: ‘Grandad's family - not sure where from. However, say's that (grandfather) mother was "half cast". No mention of this earlier. Would affect risk category’ (graded 5). This woman was not tested for clinically relevant haemoglobinopathies.

The lower grade cases for question A were inconsequential. They included: A6 (Persian)+C4 on test and A6 (Persian) and E4 on retest, the midwife adding: ‘Client found difficulty with which boxes to tick in Q1 as she felt 'Persian' was an Asian origin, rather than 'Chinese & other' ’ (graded 2); another woman gave A5+E4 on test but E4 on retest, the midwife adding: ‘She ticked for Mediterranean last time thinking of the Mediterranean being also North Africa/Arab. She was happy to only have one tick on retest’ (graded 2); in a similar case (graded 1 by the midwife without comment), the respondent gave on A6+C4 (Japanese) +E2 on test and A6+E2 on retest (again, claiming an ‘Asian’ origin); and in an almost identical case a woman gave C4+E3 on test and E3 (C4 crossed out) on retest, the midwife adding: ‘Too hasty - hadn't got to Malaysian which she feels is MOST appropriate & thought "other Asian" may be the category for her. Almost did the same on retest but stated Malaysian when asked to give which "Other Asian" was her country of origin’ (graded 2). Another question A case involving Asian ethnic origins records ‘C4 (Syrian)’ in the test but ‘E5 (Syrian)’ on retest, on retest the midwife recording: ‘The husband answered the question.. when asked about the difference in answers he did not really see it as different. He was happy with Asian-Syrian and Arab or other – Syrian’ (this case being given a grade 2). Another case of inconsequential misassignment (within ‘Northern European’ origins) was a respondent who gave A1 on both test and retest, the researcher noting: ‘This lady's name sounded foreign so I looked her up on the computer and saw that her place of birth is Germany yet she has said that her ethnic family origin is White - Eng, Scottish, Welsh Irish. I will see what she says at retest’ (graded 1).

In Kings there were 16 errors on question B. The respondent answered ‘Yes – grandparents St Lucia’ on the test but ‘No’ on retest, adding: ‘…but I do not know what you are trying to ask. You have made this form/question more complicated than it need be’ (ungraded). In the next case the respondent answered ‘Yes – Jamaica’ on test and ‘No’ on retest (again ungraded). In another, the respondent answered ‘No’ in the test but ‘Yes – Jamaica’ on the retest (graded 4, the midwife on test adding ‘The client clearly did not understand the question). The next case was ambiguous, the respondent answering ‘No’ & ‘Don’t know’ in part 1 and ’Yes’ & ‘Don’t

know' on retest. In the test the respondent ticked category 3 (Chinese) on the language question but 'Ga' (a major language of Ghana) on retest (a response graded 4). In a retest where the original form was missing, the woman completed both questions A & B, giving D2 (African) on A and 'No' to part 1 of question B (incompatible responses) (ungraded). In the next case the respondent gave 'No' to part 1 on both the test and retest but this was graded 4 (indicating an observation at variance with the response). In a further case the woman gave 'Yes - Jamaican' on test and 'No' on retest (ungraded); the re-tester recorded: 'Original question was answered in my opinion correctly…the reinterview question was answered in variance with my understanding of the "true" answer and this would have affected allocation to risk category'. The next case was a 'Yes – Zaire' response on test and simply 'Yes' on retest (ungraded). This was followed by a recording of 'Don't know' on test and retest (not queried by the re-tester and ungraded). Another case recorded 'Yes – Nigeria' on test and 'No' on retest (ungraded but with the midwife recording 'Client is African'). In a further response, the woman stated 'Don't know' on test but 'Yes, my mum & mum's parents are from Jamaica and my dad's parents are from Jamaica' on retest, the midwife adding: 'This client is of Caribbean descent' and the re-tester: 'client stated that she thought the question was asking if she knew if any of her ancestors had sickle cell disease' (graded 1). In the next case the woman recorded 'Yes – Barbados & Jamaica' on test and 'Barbados, Jamaica, & Syrina' (the last an unknown location) on retest, the re-tester recording: 'Identification of haemoglobinopathy caused client to question ethnicity & revealed more detail than previously aware of' (graded 2). The next case involved a response of 'No' on test and 'Yes, Jamaica, Dominica' on retest (graded 4). Another (less consequential) involved a 'Barbados & Canada' response on test and 'Barbados & Guyana' on retest, (graded 3, the re-interviewer adding 'Didn't really understand what ancestry/ethnicity was'. Another respondent gave 'Yes – Jamaica' on test and 'No' on retest (ungraded). Finally, a woman gave 'Don't know' on test and 'Yes – St Vincent & Jamaica' on retest (graded 4). Clearly, the vast majority of these errors are consequential for risk status.

There were only five errors on Question A. A respondent gave A1 + D2 on test and D2 on retest (graded 1), not consequential with respect to risk status. Another respondent gave A1 + C1 +D1 on test and A6 (Black British), the respondent adding 'do not understand the question') (graded 2). A third respondent gave C1 + D1 on test and E4 on retest, comments on the form stating: 'Error completing form where boxes checked by midwife' & 'Client stated that she did not complete the question herself at first interview. Did not affect allocation to risk category, but did affect her response at the second interview' (ungraded). Another respondent gave C2 on test and A1+D1 on retest, the re-interviewer adding: 'Not sure that client fully understood. Did not have original form at time of second interview. Client did not

know what form she had previously used. Answer at second interview (self-completed by client) is at variance to her answer at recruitment. This initial answer appears to be more correct than her answer at reinterview' (ungraded). Another respondent gave C1+D1 at test and A1+C2+D1 at retest, an inconsequential change (graded 1). Finally, a woman gave D2 at test and A1+D2 at retest, the midwife adding 'Clearly did not understand the question' (graded 3 but an inconsequential change with respect to risk status).

In Leicester 291 retests were recorded. Midwives' quality grading recorded 149 Question A and 127 Question B grade 1 results. However, of the poorer grades (2/3), there were 2 Question A and 6 question B results. However, two grade 2s for Question A and two for Question B were assigned by the counsellor on retest as she had no original test schedules to link with. The only serious error on the retests in Leicester were one woman who answered 'yes' on part one of Question B and wrote in Leicester but, on the retest, gave her correct ethnic/family origins (Antigua, Dominica, and St Vincent). Another question B respondent gave 'don't know' responses at test and retest. A third woman who answered question B reported 'Jamaica' at the test and 'Jamaica & St Kitts & Nevis' at retest (noting 'I forgot about one side of my family).

5.7. What happens in a selective screening programme: the case of Exeter

Three of the four sites in the trial of the ethnicity questions employed universal screening (requiring all women booking to be tested for trait carrier status, although – in practice – not all women are tested): Birmingham, Kings, and Leicester (the last introducing universal screening for the purposes of the trial). However, one of the sites, Exeter, characteristic of a low density of minority ethnic groups, continued to follow a practice of selective antenatal screening. The findings from the trial in this site enable us to evaluate this practice from one important perspective: Were women at risk because of their ethnic/family origins tested for clinically relevant haemoglobinopathies. However, unlike the sites in which universal screening took place, statistics are not available on whether women who were identified as carriers by laboratory tests identified as at risk by their responses to the ethnicity questions (as the latter were the determinant of testing).

Table 12. Ethnic origin and laboratory testing in a selective antenatal programme: Question A

Ethnic Origin	***No information on testing recorded***	***No sickle cell / thalassaemia test***	***Other***	***Tested***	***Result[4]***
White: English, Scottish, Welsh, or Irish (A1)	25	360[1]		0	
Other N European (A2)	0	3		0	
A1+A2	2	8[1a]		0	
A1+A2+A3†		1		0	
A1+A3†		1		0	
A1+A2+E5 (N Am Indian)†		1		0	
A1+A5†		8		0	
A1+A6		2			
A2+A5†		1		0	
A5†	1	3		1	negative
A1+A5+A6†		1		0	
A6		4		0	
A1+C1†		1	2[2]	1	negative
A1+C4 (Burmese)†				1	negative
A1+D1†		1			
A1+D3(Egypt)†		1			
A1+E4†		1			
A1+E5(Aboriginal)†				1	+ve (11)
A1+E5 (Jewish)†(?)		1			
A5+E4†		1[2a]			
A6+E4†			1		
A6+C4(Japan)+E2†(?)		1			
C1 (Gujarati speaker)†		1			
C1 (African Indian)†				1	+ve (11)
C4 (Syrian)				1	negative
C4+E3†		1			
E3†		3			
E4†				1	+ve (11)

Notes:
[1] Included one woman who was adopted ('Patient is adopted & does not know about her background. Booking was observed. Woman was offered screening in view of above but declined').
[1a] Schedule records for one patient: 'R has ancestors who were Czech/Jewish which was not elicited by the questions, though it came up in discussion afterwards. I'll ask for screening on her next sample at 28/40'.
[2] Will take blood at next test [1]; unable to check woman & unable to identify midwife [1]
[2a] 'Discussed with lab – they think she should have sickle screen. I will leave a message for her community midwife to this effect' [1]
[3] Already been tested & negative
[4] 11 = MCH>25<27
893 cases

This data shows that around 25-27 women who were at risk because of their ethnic family origins were not tested, including all but one of the 15 who ticked 'Italian, Maltese, or other

Mediterranean' (singly or in combination) and all three who ticked 'Malaysian, Vietnamese, or Filipino'. Clearly, accurate capture of ethnic/family origins at risk is, alone, unsatisfactory if midwives do not know which ethnic origin groups *are* at risk.

Table 13. Ethnic origin and laboratory testing in a selective antenatal programme: Question B

Ethnic Origin	***No information on testing recorded***	***No sickle cell / thalassaemia test***	***Other***	***Tested***	***Result***
'No' to pt. 1	23	335[1]		1[2]	negative
'Don't know' to pt. 1	2	14		0	
Blank (ethnic question) pt. 2		5			
All North European or other White background not at risk		40		0	
China†				1	negative
Spain, Argentina† (?)		1			
India†		1			
Gt grandmother possibly Jewish†				1	negative
Cyprus†		1			
Grandfather Burmese, Dad half Burmese†				1	negative
Father is from Italy (N.) & mother is from German descent†				1	negative
India & America†		1			
Zimbabwe†		1			
Italy†		2[3]			
Myamar (Burma) †		1			
Great grandfather S Africa, Great grandfather French † (?)		1			
Cuba†		1			
Grandfather Sicilian†		1			
Hong Kong & China†		1			
Egypt & S Africa & India†		1			
'half British ancestors'† (?)		1			
Jamaica†		1		1	SC trait
Grandad Jamaica†		1[4]			
India†			1 (done in May)	1	negative
Born in Uganda, father's family German Jewish† (?)		1			
Sri Lanka†		1			
India+America†				1	+ve (11)
Central America (Belize) mother†		1			
Greece†				1	negative
Italy or Spain†		1			
Slovakian/Croatian†		1			
Zambia+S Africa+Malaysia				1	negative

Notes:

[1] Included (i) a case where researcher recorded: 'Based on this woman's name (which sounds Asian) I contacted community midwife to ask about this woman. The midwife confirmed she "looks Indian" but

has answered "No" to question B! This could prove interesting at retest'. Despite the uncertainty, no specific SC/Thal screening was requested. (ii) One patient whose mother was adopted.

[2] This patient gave 'No' in part one but gave language as 'Thai' and was tested.

[3] For one of these cases there is the annotation: 'Thal screen requested of community midwife 8/5/03. Letter to community midwife reminded her of risk groups & need for test. 13/6/03 FBC results back – within normal limits but no specific SC/thal screen requested!!'.

[4] Annotated with comment: 'This questionnaire done in January 03. Not put forward for screening. D/W lab – for sickle dex –community midwife informed around 9/1/03/ 29/1/03 still no sickle dex results. 26/1/03 sickle dex req. but wrong blood bottle sent. Requested another sample from community midwife. 8/5/03 letter to community midwife requesting sickle cell screening. 12/5/03. FBC taken but no request for sickle cell test. 30/6. No FBC results more recent than 12/5. Still no sickle cell results as not tested'.

On question B around 20 women whose origins suggested they were at risk did not get tested. Moreover, a further 19 women whose ethnicity was not recorded (blank, unknown) were not tested.

Conclusion

This evidence suggests that some 45 women or around 5% of those recruited into the study were at risk but not tested.

5.8. Did the ethnicity questions identify those with clinically relevant haemoglobinopathies: The results of the universal screening programme (Birmingham, Kings, and Leicester)

All women are supposed to be screened in a universal screening programme, although there were failures in all three sites. However, for women for were tested, it is possible to derive statistics of the efficacy of the ethnicity questions in identify those women laboratory tests indicated that they were carriers of a clinically significant haemoglobinopathy (these women would only be a small subset of all women at risk because of their ethnicity status.

In Birmingham a total of 21 women who participated in the trial had clinically relevant haemoglobinopathies (table 14). In all but one case ethnic/family origins were concordant. The discordant case was a response to question B where the woman had given the answer 'Don't know' to part one of the question. Additionally, the preferred language of the interview gave no clue to her ethnic/family origins as it was 'English'. The schedule contained the annotation: 'This couple are from Guyana and have returned for a visit so were not available for re-tests'.

Table 14. Ethnic origin and laboratory testing in Birmingham: Clinically relevant haemoglobinopathies

Trait (n=21)	***Question***	***Concordant ethnic origin***	***Discordant or missing ethnic origin***
S (=9)	A=6 B=3	6 2	0 1 (ans. 'Don't know to pt. 1)
C (=3)	A=2 B=1	2 1	0 0
D (=2)	A=2	2	0
E + MCH>25 <27 (=1)	A=1	1	0
B-Thal (=5)	A=3 B=2	3 2	0 0
B-Thal (?on course of iron & will recheck) (=1)	B=1	1	0
Gross categorical errors			**A=0** **B=1**

Note: The data excludes one case of trait C where the woman declined to participate in the trial but the midwife completed information on clinically relevant haemoglobinopathies.

Amongst the King's data there were 102 recorded clinically relevant haemoglobinopathies (table 15). There were 11 categorical errors, all but one relating to question B. These related to responses by the women of 'No', 'Don't know' or blank to part one of the question.

Table 15. Ethnic origin and laboratory testing in Kings: Clinically relevant haemoglobinopathies

Trait (n=102)	***Question***	***Concordant ethnic origin***	***Discordant or missing ethnic origin***
S (=77)	A=45 B=32	44 25	1^{K1} 6^{K2}
C (=10)	A=4 B=6	4 4	0 2^{K3} (ans. 'Don't know' to Qn. B, pt. 1^{1})
D(=1)	A=1	1	0
E(=1)	B=1	0	1^{K4}
S (co-existent alpha thal trait) (=1)	B=1	1	0
C + MCH>25<27 (=2)	A=1 B=1	1 1	0 0
S + MCH >25 <27 (=1)	A=2	2	0
S + MCH<25 with other ethnic/family origins	B=1	1	0
B Thalassaemia (=1)	B=1	1	0
Homozygous or Compound Heterozygous Conditions: Sickle Cell Disease (=3)	A=3 B=1	2 1	0 0
Other Compound Heterozygous (include haemoglobin E/β^0 –thalassaemia (=2)	A=2	2	0
Gross categorical errors			**A=1** **B=10**

Notes:
[K1] This case was 'A1+A3' (possibly concordant as A3=Greek or Greek Cypriot).
[K2] These cases were: 'B – Don't know'; 'B – No'; 'B – Don't know'; 'B – Don't know'; 'B – No'; 'B – blank'.
[K3] These cases were: 'B – Don't know' [The schedule is annotated with the comment (on retest): 'This client is of Caribbean descent']; 'B – No + Don't know'.
[K4] 'B' (blank)

Table 16. Ethnic origin and laboratory testing in Leicester: Clinically relevant haemoglobinopathies

Trait (n=51)	***Question***	***Concordant ethnic origin***	***Discordant or missing ethnic origin***
S (=18)	A=10 B=8	10 7	0 1(data discordant for one[2])
S + MCH <25 with other ethnic/family origins (=3)	A=2 B=1	2 1	0 0
C (=2)	B=2	2	0
C + MCH>25<27 (=1)	B=1	1	0
D (=3)	A=2 B=1	2 1	0
E (=2)	A=1 B=1	1 1	0 0
E + ?α^0Thal? MCH<25 with ethnic/family origins Chinese/SE Asian/Mediterranean (=1)	B=1	1	0
B-Thal (=15)	A=10 B=5	8 5	2 (data missing for 2[1])
B-Thal + MCH<25 with ethnicity unknown (=1)	B=1	0	1 (No to pt. 1 & language=English[3])
Other haemoglobinopathies (Lepore, $\delta\beta$ Thal, HPFH) (=1)	A=1	1	0
?α^0Thal? MCH<25 with ethnic/family origins Chinese/SE Asian/Mediterranean (=1)	A=1	1	0
Other Compound Heterozygous (including haemoglobin E/β^0-thalassaemia) (=1)	A=1	1	0
Clinically relevant haemoglobinopthy (answer=yes) but no details of traits (=2)	A=2	2	0
Gross categorical errors			**A=0** **B=2**

Notes:

[1] Neither of these errors were categorical errors (that is related to the ethnic question) but operational. In the first case the retest result recorded that the midwife had failed to ask the ethnicity question at booking; the second case was one of seven retests undertaken by counsellors that could not be linked to the booking in interview schedule, so, by definition, ethnic results at the booking in interview were unavailable.
[2] The woman had answered 'Yes' to question B & written in 'Leicester'.
[3] The woman answered 'No' to part one of question B and her preferred language (English) gave no clue as to her ethnic/family origins.

In Leicester universal testing identified 51 cases of clinically relevant haemoglobinopathies. All but four were detected by the ethnic/family origins questions. There was one response to Question B (S trait) where the woman simply identified Yes to B and wrote in 'Leicester' in one of the free text fields. In a second question B case the woman gave a 'No' response to part 1 of question B (and her preferred language for the interview [English] gave no clue to her ethnic/family origins). In two cases there were responses to question A where the ethnicity field was blank (one case where the ethnicity question was not asked & a second where data was only available for the retest & with no record linkage via the schedule number (where, by definition, the test ethnic data at booking was unavailable).

Conclusion

The 3 sites in which universal screening took place – Birmingham, Kings, and Leicester – reported a total of 174 clinically relevant haemoglobinopathies. Of these there were 14 incongruent responses, all but one relating to Question B or 8.0% (18.1% of all Question B responses [n=69] in which clinically relevant haemoglobinopathies occurred). This high rate of incongruent responses on Question B falls well outside the tolerance limits suggested by Zeuner *et al*. (1999), regardless of all the other cumulative source of error.

5.9. How did the ethnicity questions identify those with abnormal MCH (mean corpuscular haemoglobin) results, possibly indicating thalassaemia traits: other haemoglobinopathies (MCH<25 with other ethnic/family origins; MCH<25 with ethnicity unknown; & MCH>25<27): The results of the universal screening programme (Birmingham, Kings, and Leicester)

Antenatal screening for haemoglobinopathy disorders usually comprises a stepwise process of carrier testing of expectant mothers – and, if positive, carrier testing of partners – to identify if the pregnancies are at risk of an affected fetus. A cascade sequence of sequential laboratory tests is needed to establish whether the woman is a carrier. These start with measurement of mean corpuscular haemoglobin (MCH) and characterisation of structural haemoglobin (Hb) variants via Hb-electrophoresis/HPLC which are interpreted together. An MCH of ≥27 pg is regarded as normal and indicative of no thalassaemia trait. A low MCH (<27 pg) may indicate a thalassaemia trait. As part of routine obstetric care, laboratory testing is then undertaken in these low MCH cases to quantify HbA_2 and HbF, the main indicators for thalassaemia traits. A low MCH with normal HbA_2 constitutes an uncertain result. If the woman is of Chinese,

south-east Asian, or eastern Mediterranean ethnic origin with considerably reduced MCH (<26 pg), a possible diagnosis is α^0 thalassaemia trait. In women with other ethnic origins the likely explanation is insignificant α^+ thalassaemia trait/homozygous state of iron deficiency (or both). Thus, in the screening algorithm, the finding of a low MCH in women is critical, whether or not HbA_2 has been found to be elevated, and requires reference to the woman's ethnic origin. The trial schedule recorded MCH in three categories: MCH<25 with other ethnic/family origins; MCH<25 with ethnicity unknown, & MCH>25<27.

Question A

Table 17

	Birmingham			*Kings, London*			*Leicester*		
WHITE	MCH<25 with other origins	MCH<25, Ethnicity unknown	MCH >25 <27	MCH<25 with other origins	MCH<25, Ethnicity unknown	MCH >25 <27	MCH<25 with other origins	MCH<25, Ethnicity unknown	MCH >25 <27
English/Scottish/Welsh/Irish						1			17
Other N European									
Greek/Greek Cypriot & Turkish/Turkish Cypriot†						1			
Italian, Maltese, or other Mediterranean†									
Any other White background					1				1
Mixed white – not at risk							1		
Mixed white – at risk†									
MIXED INVOLVING AT LEAST ONE MINORITY ETHNIC GROUP (MEG)									
Mixed MEG + White†				1					
Mixed MEG only†									
ASIAN OR ASIAN BRITISH									
Indian or African-Indian†			1	2			5		19
Pakistani†							1		3
Bangladeshi†						1			
Any other Asian background†									1
BLACK OR BLACK BRITISH									
Caribbean†				1	2	2			
African†				5		6			4
Any other Black background†									
CHINESE AND OTHER									
Chinese†									
Japanese†									
Malaysian, Vietnamese, or Filipino†						1			
N African, Arab, or Iranian†									
Other origins†									
MISSING DATA									
TOTAL			***1***	***9***	***3***	***12***	***7***		***45***

Question B

Table 18

	Birmingham			*Kings, London*			*Leicester*		
WHITE	MCH<25 with other origins	MCH<25, Ethnicity unknown	MCH >25 <27	MCH<25 with other origins	MCH<25, Ethnicity unknown	MCH >25 <27	MCH<25 with other origins	MCH<25, Ethnicity unknown	MCH >25 <27
English/Scottish/Welsh/Irish									
Other N European									
Greek/Greek Cypriot & Turkish/Turkish Cypriot†									
Italian, Maltese, or other Mediterranean†							1		
Any other White background									
Mixed white – not at risk									
Mixed white – at risk†									
MIXED INVOLVING AT LEAST ONE MINORITY ETHNIC GROUP (MEG)									
Mixed MEG + White†									
Mixed MEG only†									
ASIAN OR ASIAN BRITISH									
Indian or African-Indian†				1		3	3		9
Pakistani†						1			
Bangladeshi†			1						
Any other Asian background†						1			2
BLACK OR BLACK BRITISH									
Caribbean†				1	2	5	1		2
African†				3		8	1		5
Any other Black background†									
CHINESE AND OTHER									
Chinese†									
Japanese†									
Malaysian, Vietnamese, or Filipino†									
N African, Arab, or Iranian†									1
Other origins†									
B – No (pt. 1)				3		2	4	1	17
B – Don't Know					1	1			1
B – Yes (blank)									1
TOTAL			***1***	***8***	***3***	***21***	***10***	***1***	***38***

The findings are consistent across both questions: a high number of cases of MCH >25<27 in the White-English/Scottish/Welsh/Irish group on question A and the equivalent group on question B ('B – No') and amongst women of Indian ethnic origins and somewhat smaller numbers across the different categories in women of Black-African, Black-Caribbean, or other black origins.

6. Conclusion and Recommendation

Two candidate evidence-based ethnic/family origin questions were tested in this formal trial to assess their efficacy in ascertaining risk status in selective antenatal screening for sickle cell and thalassaemia. One of these questions (Question A) is a classification question similar in structure to the 2001 Census England and Wales question but with extended categorisation to capture all appropriate risk groups and a "tick all that apply" method (as opposed to categories) to capture mixed heritage; the other (Question B) is an open response ancestry or ethnic/family origins question, similar in type to those used in US and Canadian Censuses, but comprising an initial "screening" question to identify those with ancestors from areas of the world outside of the UK or Republic of Ireland followed by free text provision to write in countries of ethnic/family origin.

The conceptual base of both questions – ethnic/family origins – used in large scale social including government surveys (such as the PSI Fourth National Survey and the 1999 Health Survey for England) was understood by women in a cognitive research exercise. However, some women did not understand the meaning of the term 'ancestors', a finding common to that reported in cognitive research for the 2001 Census Development Programme, raising doubts about its utility as a concept for a screening question.

A total of 4,775 women were formally invited to participate (that is, their responses to invitation were recorded) in this trial from 4 centres: Birmingham (n=240), Exeter (n=1059), Kings, London (1197), and Leicester UA & County (2279). These sites represented areas of varying ethnic density, low in Exeter and high in King's, London, Birmingham and Leicester UA. Refusal rates varied from 7.9% (Leicester) to 28.7% (Birmingham) on question A and 7.3% (Leicester) to 38.5% (Birmingham) on question B. Total achieved valid responses numbered 4232.

The mean time taken to ask the ethnic/family origin question in the four centres varied from 2.2 minutes in Birmingham to 4.8 minutes in Leicester (median range, 1.1-4.0). The time taken to ask question B was longer than that for question A, although the differences as measured by the median were significant in only two of the four sites (Birmingham and Exeter). Moreover, in two of the sites (Birmingham and Kings) there was much greater variability in the times recorded for question B than question A. The reasons for these differences are unclear but the lack of context in question B in contradistinction to the detailed categorisation provided in question A and possibility some confusion

concerning the conceptual base of question B could have contributed to some of the differences.

Question A performed very successfully in all four sites. There was missing data for only 12 respondents (that is, people recruited to question A who did not tick a box) out of a total of 2147 who answered question A, a rate of 0.6% that is tolerable. Moreover, the misinterpretation rate (n=2, 0.1%) related to women who ticked a box for their partner, not a misinterpretation of the ethnic question as such but one of subject status.

On part one of question B a gross error rate of 5.53% was incurred amongst the 2081 women recruited into the trial who answered this question. Most of this error (4.18%) was attributable to 'Don't knows' but 1.30% to missing answers. This is a substantial error rate – almost 5% - bearing in mind the kind of error ranges that Zeuner *et al.* (1999) were modelling. This error rate, of course, excludes errors that arose in part 2 of the question.

Errors in part two of question B comprised missing data and misinterpretations. There were a total of 22 missing responses for respondents who had ticked 'yes' in the screening question out of a total of 709 responses (or 3.1%, over five times the rate for question A). Moreover, misinterpretations numbered 4 (or 0.68%, over seven times the rate in Question A). Further, there were consequential categorical ambiguities in Question B (for example, five in the Leicester data).

Language data was missing on 119 cases of Question A and 91 of Question B (the identical language question set was asked on both schedules). There was evidence of a lack of congruence (strong or equivocal) between assignment on the ethnic questions and language in a total of 52 cases. Ethnicity/language incongruence was much higher on Question B (2.43%, 47/1934) than Question A (0.24%, 5/2063). Language incongruence on Question B was highest in Birmingham and Leicester (at 3.1%), both centres with significant numbers of respondents of South Asian ethnic origin. Part 1 of Question B may have been subject to misinterpretation (as invoking, say, conceptualisations of 'Britishness' or nationality). It is notable that language incongruence on Question B was much lower in Kings (indeed, the lowest across the 4 sites) whose minority ethnic group population is primarily Afro-Caribbean.

On the test-retest component of the trial – a measure of reproducibility or reliability of results and, consequently, the stability of the questions – the consequential error rate for question A was 3 and for question B 25, an extremal quotient of 8.3.

In Exeter, the only site utilising selective antenatal screening for sickle cell and thalassaemia, the evidence from the trial suggests that some 45 women or around 5% of those recruited into the study were at risk but not tested.

The 3 sites in which universal screening took place – Birmingham, Kings, and Leicester – reported a total of 174 clinically relevant haemoglobinopathies. Of these there were 14 incongruent responses, all but one relating to Question B or 8.0% (18.8% of all Question B responses [n=69] in which clinically relevant haemoglobinopathies occurred). This high rate of incongruent responses on Question B falls well outside the tolerance limits suggested by Zeuner *et al*. (1999), regardless of all the other cumulative source of errors.

Findings for MCH values suggestive of thalassaemia and ethnicity are consistent across both questions: a high number of cases of MCH >25<27 in the White-English/Scottish/Welsh/Irish group on question A and the equivalent group on question B ('B – No') and amongst women of Indian ethnic origins and somewhat smaller numbers across the different categories in women of Black-African, Black-Caribbean, or other black origins.

On all the measures used in the trial to assess validity and reliability (including test-retest reproducibility), Question A outperformed Question B. Cumulative gross errors on Question B substantially exceeded those on Question A.

It is recommended that Question A be adopted as the candidate question for use in selective antenatal screening for sickle cell and thalassaemia.

References

Ahdieh, L., Hahn, R.A. (1996). Use of the Terms 'Race', 'Ethnicity', and 'National Origins': A review of articles in the American Journal of Public Health, 1980-1989. *Ethnicity & Health* 1(1), 95-98.

Aspinall PJ. Operationalising the collection of ethnicity data in studies of the sociology of health and illness. *Sociology of Health and Illness* 2001; 23(6): 829-862.

Aspinall PJ. *Secondary Review of Existing Information in Relation to the Ethnic Question: Synthesis of the Literature*. Canterbury: CHSS, 2002 (March).

Aspinall PJ. Who is Asian? A category tat remains contested in population and health research. *Journal of Public Health Medicine* 2003; 25(2): 91-97.

Aspinall PJ, Dyson SM, & Anionwu EN. The feasibility of using ethnicity as a primary tool for antenatal selective screening for sickle cell disorders: pointers from the research evidence. *Social Science & Medicine* 2003; 56: 285-297.

Erens, B., Primatesta, P., & Prior, G. (Eds.) (2001). *The Health Survey for England 1999. Vol. 1: Findings. Vol. 2: Methodology and Documentation*. London: The Stationery Office.

Galactéros, F. (1999). Détection Néonatale de la Drépanocytose en France Métropolitaine. *Pathologie Biologie* 47(1), 13-18.

Gerber, E., & de la Puente, M. (1996). The development and cognitive testing of race and ethnic origin questions for the year 2000 decennial census. Paper presented at the Bureau of the Census 1996 Annual Research Conference, March 17-19, 1996 in Arlington, Virginia.

Jobe JB & Mingay DJ. Cognitive Laboratory Approach to Designing Questionnaires for Surveys of the Elderly. *Public Health Reports* 1990; 105: 518-524.

Modood, T., Berthoud, R., Lakey, J., Nazroo, J., Smith, P., Virdee, S., & Beishon, S. (1997). *Ethnic Minorities in Britain: Diversity and Disadvantage*. London: Policy Studies Institute.

Mortimer L & White A. Ethnic Group Question: Findings from focus group discussions. London: Office for National Statistics (Social Survey Division), 1996 (July).

Nazroo, J.Y. (1997). *The Health of Britain's Ethnic Minorities: Findings from a National Survey*. London: Policy Studies Institute.

ONS (1998). *Evaluation of the Main Objectives of the 1997 Census Test*. London: Office for National Statistics.

Rainford L. *2001 Census testing programme: Report on the ethnic group and religion question test carried out in March 1997*. London: Office for National Statistics (Social Survey Division), 1997 (August).

Willis GB, Royston PN, & Bercini D. The use of verbal report methods in the applied cognitive laboratory. *Applied Cognitive Psychology* 1991; 5: 251-267.

Appendices: 1. Questions

Question A

DO YOU HAVE ETHNIC/FAMILY ORIGINS THAT ARE…

Please tick one or more boxes to indicate these origins

A. WHITE

English, Scottish, Welsh, or Irish	□
Other North European	□
Greek or Greek Cypriot	□
Turkish or Turkish Cypriot	□
Italian, Maltese, or other Mediterranean	□
Any other White background (*please write in……...)*	□

B. MIXED ► ***Please tick all boxes in sections A, C, D and E (above & below) that apply to you***

C. ASIAN OR ASIAN BRITISH

Indian or African-Indian	□
Pakistani	□
Bangladeshi	□
Any other Asian background (*please write in…….)*	□

D. BLACK OR BLACK BRITISH

Caribbean	□
African	□
Any other Black background (*please write in…….)*	□

E. CHINESE AND OTHER

Chinese	□
Japanese	□
Malaysian, Vietnamese, or Filipino	□
North African, Arab, or Iranian	□
Any other (*please write in…….)*	□

Source: Aspinall (2002)

Question B

1. Do you or any of your known ancestors, as far back as you can recall, have ethnic/family origins from areas of the world outside of the United Kingdom or Republic of Ireland?

Please tick one box only

Yes	□
No	□
Don't know	□

3. If YES, then for you or for any of your known ancestors, as far back as you can recall, please write in all countries in the spaces below:

Source: Aspinall (2002)

2. Statistical power calculations

[Acknowledgement is due to Dr Colin Cryer, CHSS]

Sample sizes

Women presenting for booking would be randomised to two ethnic questions: XA (ethnic classification); XB (binary classification).

Power calculations were based on the following estimates of patient flows:

Based on a randomised design (using random number tables) approximately half the women would be randomised to question XA and half to question XB.

Power calculations

	R (trait carriers identified in population at risk)	NR	TOTAL (True Positives)
Question XA	R_1	(N_1-R_1)	N_1
Question XB	R_2	(N_1-R_1)	N_2

The calculation is based on true positives (N_1 & N_2) - that is - trait carriers and the difference between these true positives and the number of carriers identified in the at risk groups captured by the ethnic questions (R_1 & R_2). The measure used is sensitivity (true positives over [true positives and false negatives]). This is more powerful than specificity although that can be factored in (but would not make a substantial difference to the power calculations).

Sensitivity

Let us start with a base proposition that the best question (for argument's sake, Question A) will detect 97 out of every 100 trait cases (assuming no administrative failures or other extraneous sources of risk group misclassification). This was the level regarded by Zeuner *et al* (1999) as an acceptable level of risk group misclassification for a selective programme (i.e., 97%, the mid-level of misclassification in his three hypothetical outcomes). Thus, for question XA, $R_1/N_I = 97\%$. We then hypothesise that R_2= 95% or 94% or 93% or 92% or 91% or 90%.

Based on these propositions, the numbers needed would be as follows:

Parameters: Significance level .05, two-sided test, Allocation ratio 1.

Sample Size Table: Comparison of Proportions - Computation of Sample size

Power 80%

Proportion 1	**Proportion 2 .97**	**Subjects required to compare two groups**
.85	89	178
.86	100	200
.87	115	230
.88	134	**268**
.89	159	318
.90	194	**388**
.91	245	490
.92	326	652
.93	465	930
.94	749	1498
.95	1506	3012

If data collection were over 6 months in each centre, the numbers in the study would be adequate at 80% power to detect a 9 percentage point difference between question A and question B, enough to detect a difference of .89 vs. 97. Aggregation would appear to be legitimate because there is no reason to assume that the *sensitivity* of the questions would vary across areas with different densities of minority ethnic groups. If collection in SE London were extended over 9 months, the numbers in the study would be adequate at 80% power to detect a 7 percentage point difference between question A and question B, enough to detect a difference of .90 vs. 97.

Specificity

The parameters used here are:
Significance level 0.05
Two-sided test
Allocation ratio 1

Sample size in each of two groups = 4,500 (total = 9,000). Each cell gives the power.

	Proportion 2								
Proportion 1	**.991**	**.992**	**.993**	**.994**	**.995**	**.996**	**.997**	**.998**	**.999**
.99	.07	.17	.34	.56	.78	.92	.98	.9999	.9999
.991	-	.07	.18	.37	.62	.83	.95	.99	.9999
.992	-	-	.08	.20	.42	.69	.89	.98	.9999
.993	-	-	-	.08	.23	.48	.76	.94	.99
.994	-	-	-	-	.09	.27	.56	.85	.98
.995	-	-	-	-	-	.10	.32	.67	.93
.996	-	-	-	-	-	-	.12	.41	.81
.997	-	-	-	-	-	-	-	.15	.56
.998	-	-	-	-	-	-	-	-	.23

At sample size of 4,500 in each group, we should be able to detect at least a 0.05% difference in the specificity of the two questions.