

# **Wellcome Trust Monitor**

## ***Technical Report***

Elizabeth Clery, Sarah Butt and David Hussey

September 2011

Prepared for: the Wellcome Trust



# Contents

<b>Tables</b> .....	<b>4</b>
<b>Figures</b> .....	<b>4</b>
<b>1 Introduction</b> .....	<b>1</b>
1.1 Background and objectives .....	1
1.2 Archiving of data.....	1
1.3 Report on the findings .....	2
<b>2 The samples</b> .....	<b>3</b>
2.1 Overview.....	3
2.2 Drawing the core sample.....	3
2.3 Focused enumeration.....	4
2.4 Historical database.....	4
<b>3 Questionnaire development and piloting</b> .....	<b>5</b>
3.1 Scope of development work .....	5
3.2 Cognitive pilot.....	5
Respondents .....	5
Briefing and de-briefing .....	6
Outcomes.....	6
Post-pilot modifications .....	6
3.3 CAPI dress rehearsal pilot.....	7
Respondents .....	7
Briefing and debriefing .....	7
Outcomes.....	8
Post-pilot modifications .....	8
3.4 Questionnaire development.....	9
<b>4 Data collection</b> .....	<b>10</b>
4.1 Advance letter.....	10
4.2 Briefings.....	10
4.3 Scheduling of interviews .....	10
4.4 Quality control.....	11
4.5 Fieldwork progress .....	11
4.6 Thank you letter and voucher.....	12
4.7 Interview length .....	12
<b>5 Response</b> .....	<b>13</b>
5.1 Adult response rate .....	13
5.2 Young person response rate .....	13
5.3 Outcomes .....	15
5.4 Further information on response .....	17
Efforts made to maximise response.....	17
Further information on refusals .....	18
5.5 Characteristics of the achieved sample.....	18
<b>6 Weighting</b> .....	<b>23</b>
6.1 Overview.....	23
6.2 Adult weight .....	23
Dwelling unit selection weight .....	23
Adult selection weight .....	23
Combined selection weight .....	23
Calibration to the population .....	23
Scaling the weights .....	24
6.3 Young person weight.....	24
Dwelling unit selection weight .....	24
Young person selection weight .....	25
Calibration to the population .....	25
Scaling the weights .....	26
<b>7 Sampling errors</b> .....	<b>27</b>

<b>8 Data processing and management .....</b>	<b>30</b>
8.1 Editing.....	30
8.2 Coding .....	30
Other specify questions.....	30
Open questions .....	30
Occupation coding .....	31
8.3 Early programming error.....	31
8.4 Using the dataset for analysis .....	32
Selecting adult or young person respondents for analysis .....	32
Defining “science” subjects .....	32
Derived variables .....	32
Missing codes .....	32
Significance testing .....	33
<b>References .....</b>	<b>34</b>
<b>Appendix A Pilot Documents.....</b>	
<b>Appendix B Fieldwork Documents .....</b>	
<b>Appendix C Editing and coding documents .....</b>	
<b>Appendix D Questionnaire and showcards .....</b>	
<b>Appendix E Derived variables .....</b>	

## Tables

Table 5-1	Fieldwork outcomes for adult sample .....	13
Table 5-2	Fieldwork outcomes for young people (core addresses).....	14
Table 5-3	Fieldwork outcomes for young people (FE addresses) .....	14
Table 5-4	Full breakdown of core address outcomes .....	15
Table 5-5	Full breakdown of outcome codes for eligible young people (core addresses) .....	16
Table 5-6	Full breakdown of outcome codes for eligible young people (FE addresses).....	17
Table 5-7	Characteristics of adults respondents (aged 18+) .....	20
Table 5-8	Characteristics of young person respondents (14-18).....	20
Table 5-9	Comparison of Wellcome Trust Monitor and HSE respondents, by education and social class .....	21
Table 5-10	Comparison of Wellcome Trust Monitor and NTS respondents, by education and social class .....	22
Table 6-1	UK adults (18 and over), by region .....	24
Table 6-2	UK adults (18 and over), by age and sex .....	24
Table 6-3	UK 14-18 year olds, by region .....	25
Table 6-4	UK 14-18 year olds, by age and sex.....	26
Table 7-1	Complex standard errors and confidence intervals of selected variables for adults .....	28
Table 7-2	Complex standard errors and confidence intervals of selected variables for young people.....	29

## Figures

Figure 4-1	Summary of fieldwork progress (core addresses) .....	12
------------	--	----

# 1 Introduction

This report describes the research methods used in the Wellcome Trust Monitor survey, carried out in 2009. This was a baseline survey, designed with repetition in mind. The aim is to repeat the survey every three years, in order to measure continuity and change in public attitudes, awareness and understanding in this area.

## 1.1 Background and objectives

In 2008, the Wellcome Trust commissioned NatCen to carry out a survey of attitudes to medical research in the United Kingdom. The main aim of the survey was to explore public attitudes towards and knowledge and awareness of medical research, in addition to people's levels of interest in and engagement with this topic. There was a particular interest in exploring the views of young people, both on these issues and on the topic of science education and careers. For this reason, the survey involved a boost sample of young people aged between 14 and 18. As the major non-governmental funder of biomedical science in the UK, the Wellcome Trust has a critical interest in understanding public attitudes to biomedical science and technology, and in fostering greater citizen understanding and engagement with the scientific research that it conducts. The survey was designed to expand on the existing knowledge base in this area, to enable its more effective use in informing and influencing policy-making and public funding decisions.

In addition to these objectives, it was envisaged that the Wellcome Trust Monitor would have benefits, both for its funder and for other organisations working in the areas of science and medical research - most importantly in influencing strategy in relation to public engagement and education in the areas of medical research and science in the UK. It was also intended that the findings would have a broader impact on science policy and would provide an impartial and objective source of evidence on public attitudes that could inform and fuel future debates - amongst government policy-makers, in the media, amongst practitioners and in a range of other spheres.

The survey covered a range of general and specific issues in relation to medical research, including the public's awareness of, support for and participation in this area, with more focused sections examining attitudes to genetics research and, for the young people, attitudes to science education and careers. One key aim of the survey was to establish how far and in what ways the awareness, attitudes and experiences of different sections of the public vary and, for this reason, these data were complemented by a range of socio-economic, demographic and more general attitudinal information, to allow analysis by respondent characteristics.

The remainder of this report focuses on the questionnaire development process, fieldwork and data processing procedures. Chapter 2 describes the sample design. Development work on the survey and the data collection process used are outlined in chapters 3 and 4. Response rates are described in detail in chapter 5 whilst chapter 6 reports on the derivation of weights to be used in the analysis. Chapter 7 looks at sampling errors and finally, chapter 8 describes the procedures for the editing and coding of the data.

## 1.2 Archiving of data

A data set with complete documentation will be deposited with the Data Archive at the University of Essex in winter 2009-2010.

### 1.3 Report on the findings

A substantive report based on the survey findings will be published by the Wellcome Trust in 2010 (Butt *et al*, 2010).

## 2 The samples

### 2.1 Overview

The Wellcome Trust Monitor survey comprised two samples – one of adults aged 18 years and over, and one of young people aged between 14 and 18 years. Sampling of both populations was undertaken at designated “core” addresses, while focused enumeration (FE) was employed to obtain additional young people.

The sample for the Wellcome Trust Monitor survey covered England, Wales, Scotland (south of the Caledonian Canal) and Northern Ireland. The core sample was drawn from the Postcode Address File (PAF). At each sampled core address, the interviewer screened for Dwelling Units (DUs) containing at least one person aged 18 years or over. DUs not containing anyone in that age range were not eligible for the survey. If there was more than one dwelling unit at the sampled address, one dwelling unit was randomly selected. At responding DUs interviewers selected one individual aged 18 years or over at random to complete the adult interview. If an interview was achieved with this individual, the interviewer then selected one young person where available, aged between 14 and 18 years, at random, to complete the young person interview<sup>1</sup>. The core samples were designed to be representative of the general adult population aged 18 years and over and the population of young people aged between 14 and 18 years respectively, living in private households in the UK.

### 2.2 Drawing the core sample

The sample of 2,650 core addresses was drawn from the ‘small user’ Postcode Address File (PAF)<sup>2</sup>, a list of all addresses (delivery points) in the United Kingdom that receive less than 25 items of mail per day.

The sample was drawn in two stages: at the first stage the Primary Sampling Units (PSUs) were selected, at the second stage addresses were selected within the sampled PSUs. Each PSU was defined as a postcode sector or group of sectors. Postcode sectors containing fewer than 1,000 addresses were grouped with neighbouring sectors, this was to ensure selected addresses were not too close to one another. The grouped sectors were treated as a single PSU.

The sample file was sorted prior to sample selection. The stratifiers used were Government Office Region (GOR), the proportion of the population with qualifications at A-level and above and the proportion of the population in owner-occupied households. The latter two stratifiers were based on data from the 2001 Census. It was envisaged that they would strongly correlate with attitudes and experiences in relation to medical research, as they are both linked with levels of education and income, known to be associated with attitudes to such ‘academic’ subject areas.

The first stratifier was region; the PSUs were first sorted into 13 regions (nine GORs, plus Scotland, Wales and Northern Ireland<sup>3</sup>). Postcode sectors that spanned regional boundaries were

---

<sup>1</sup> An 18 year old at a core address was initially classified as an adult. If s/he was not selected as the adult respondent and an interview was obtained with another adult, this individual became eligible for the young person sample. The consequent under-representation of young people aged 18 was addressed through the weighting strategy.

<sup>2</sup> The version of the PAF used was Royal Mail postcode update 45.

allocated to the region containing the most addresses. Within each of the 13 regions, the PSUs were then listed in increasing order of the proportion of the population with qualifications at A-level and above. Cut-off points were drawn to create three equal sized bands (in terms of addresses). Within each of the 39 bands, the PSUs were listed in increasing order of the proportion of the population in owner-occupied accommodation.

Once the sampling frame had been stratified, 106 PSUs were selected with probability proportional to the number of addresses within them<sup>4</sup>. Twenty-five addresses were then selected systematically from each sampled PSU, giving a total of 2,650 core addresses.

## 2.3 Focused enumeration

To obtain a sample of young people aged between 14 and 18 years sufficient to enable detailed analysis, for each core address, either two or four focused enumeration (FE) addresses were selected, that were listed either directly before or after the core address on the PAF. In 50 per cent of cases, four FE addresses were designated and in the other 50 per cent two FE addresses were designated, meaning that a total of 7,950 FE addresses were specified in total. Interviewers were instructed to ask at each core address about the presence of young people aged between 14 and 18 years, at each of the associated FE addresses. Except when the core address confirmed that there were no young people of the required age at the FE address, or when the FE address was too far to be seen from the core address, interviewers were instructed to visit each FE address to further screen for the presence of young people aged between 14 and 18 years and, where possible, to select one at random to undertake the young person interview. When the informant at the core address was unsure about the presence of young people at the FE address, refused to provide information or where the core address was deadwood, interviewers were instructed to visit the associated FE addresses, in order to undertake screening directly.

## 2.4 Historical database

The sampling contractor routinely flags any addresses previously sampled for any NatCen general population surveys. These addresses are then excluded from subsequent surveys for a period of three years. This is to prevent respondents from being sampled too often. Any addresses flagged on the NatCen historical database were excluded before sampling addresses for both the core and FE samples. The selected addresses for the core and FE samples were both subsequently added to the NatCen historical database.

---

<sup>3</sup> London (as a Government Office Region) was further divided into two regions – Inner London and Outer London – meaning that we obtained a total of 13, rather than 12, regions.

<sup>4</sup> This was expanded by the Multiple Occupancy Indicator (MOI) in Scotland



## 3 Questionnaire development and piloting

### 3.1 Scope of development work

The development stages of the survey were conducted over a seven-month period from May 2008 to November 2008. As this was a new survey, the two main requirements of this work were to develop and test new questions, and to identify, where possible, relevant repeat questions for inclusion.

The programme of development work was based around two pilots. The first involved a cognitive pilot of a selection of the new and more challenging questionnaire material and the second consisted of a full dress rehearsal pilot using the CAPI (Computer Assisted Personal Interviewing) program and testing all of the survey procedures to be used in the main stage. The questions for the survey were primarily new questions designed specifically for this study, although the final questionnaires did include some repeat items, many of which had a number of minor amendments. The repeat items had previously been fielded on surveys such as the British Social Attitudes survey, the Eurobarometer and the National Science Foundation Survey. Details of the repeat questions, their origin and how and why they were amended are included in an appendix to the main survey report.

The development stage sought to produce two distinct but overlapping questionnaires – one for the adults and one for the young people. It was envisaged that the two survey populations would be asked some similar and some different sets of questions, with those unique to the adults focussing on issues that might be too challenging for the young people, and those specific to the young people examining their perceptions of science education and careers.

### 3.2 Cognitive pilot

A cognitive pilot took place in June 2008. The primary aim of the cognitive pilot was to test how well newly-developed questions worked, in terms of: the respondents' understanding of the terms or concepts used in the questions; whether questions had the same meaning for different groups of respondents; whether questions were clear; and whether questions were easy for respondents to answer.

#### Respondents

Four interviewers, in a mix of urban and rural areas of the country, were asked to carry out five interviews each. Attempts were made to recruit respondents from both low-income and high-income areas, to try and capture the range of knowledge and understanding of medical research and science that is likely to exist in the populations as a whole, which it was anticipated would strongly relate to socio-economic status. The interviewers used their knowledge of their local areas and approached residential addresses to find respondents.

In addition to this, the interviewers were each given individual quotas in order to recruit respondents with a variety of characteristics. Overall, the interviewers were asked to recruit 10 males and 10 females, eight young people aged between 14 and 18 years (four aged 14-16 and

four aged 17-18) and twelve adults aged 18 years and over (four aged 18-29, 30-59 and 60 and over respectively).

## **Briefing and de-briefing**

Interviewers attended a face-to-face briefing on 17<sup>th</sup> June 2008 where they were given background information about the purpose of the survey and were shown how to administer the questionnaire. Each interviewer received:

- 1 copy of the cognitive pilot instructions
- 1 copy of a quota sheet
- 5 copies of a screening questionnaire
- 3 copies of the adult questionnaire and probe sheet
- 2 copies of the young person questionnaire and probe sheet
- 1 set of showcards
- 5 copies of a general probe sheet
- 5 copies of the National Centre's leaflet for respondents, which were to be left with each respondent

All relevant documents are presented in Appendix A of this report. Interviewers were asked to make full notes as they conducted each interview, noting down any general problems and responses to the specified probes. In addition, interviewers were asked to complete a general probe sheet, recording respondents' general reactions to the content and coverage of the questionnaire and any previous experience of the issues involved that may have influenced their responses. Interviewers were also asked to complete an electronic template for the adult and young person questionnaires, amalgamating all of their feedback, prior to the debrief.

The face-to-face debrief was held on 1<sup>st</sup> July 2008. The completed electronic template questionnaires formed the basis of the discussion that took place.

## **Outcomes**

The four interviewers achieved 20 interviews between them, 12 with adults and eight with young people, and met all of the required quotas, with the exception that 11 interviews were obtained with males and nine with females. Our tentative conclusion from the sample obtained was that the lower socio-economic groups may have been under-represented although, overall, we did obtain a good cross-section of respondents.

## **Post-pilot modifications**

Feedback was provided to the Wellcome Trust as part of the general questionnaire development process, with recommended revisions to particular sections of the questionnaire being submitted and discussed.

A number of issues with specific questions and topic areas arose as a result of the cognitive pilot and these were addressed in subsequent question development work, leading up to the dress rehearsal pilot. These included:

- The need to develop a showcard identifying the definition of medical research we wanted respondents to think about throughout the survey, when answering general questions about "medical research"

- The problems of including specific examples when asking about unfamiliar concepts, as respondents tended to answer questions with these examples primarily in mind
- The fact that the term “cure” was more widely understood by the public than the (more factually accurate) term “effective treatment”
- The need to re-assure respondents that they would not be asked to participate in medical research projects as a result of taking part in the survey

### 3.3 CAPI dress rehearsal pilot

A full dress rehearsal pilot took place in September 2008. Its primary aim was to test all aspects of the Wellcome Trust Monitor, with a view to maximising the quality and effectiveness of the first main-stage survey, due to go into the field in January 2009. The dress rehearsal pilot sought to test two distinct aspects of the study – the procedures associated with implementing the survey in the field, including sampling procedures and the process of selling the survey to potential respondents, and the content and length of the adult and young person survey questionnaires.

#### Respondents

Six interviewers were issued with 150 core addresses, each in a sample point (postcode sector), containing 25 core addresses. The six postcode sectors were selected randomly but analysis suggested that they represented a good cross-section of the UK population in terms of educational levels and levels of economic resources, measured by home ownership. For each core address, either two or four FE addresses were specified, replicating the ratio that would be used for the main-stage survey.

The dress rehearsal aimed to obtain 50 adult interviews. While a sample of 150 addresses is larger than is normally necessary to achieve this goal, a sample of this size was issued to compensate for the short fieldwork period (three weeks), which inevitably provided interviewers with fewer opportunities to deal with broken appointments, follow up initial non-contacts and so on.

#### Briefing and debriefing

Interviewers attended a face-to-face briefing on 3<sup>rd</sup> September 2008 where they were given background information about the purpose of the survey and were shown how to administer the sampling procedures and questionnaires and given suggestions on how to sell the survey on the doorstep. Each interviewer received:

- 25 labelled ARFs for the core addresses within their individual sample point
- 25 FE sheets, to use at core addresses for recording outcomes for the associated FE addresses
- 15 FE ARFs, to be used in situations where it emerged that the FE address needed to be contacted
- 25 copies of an advance letter, to be posted to each core address
- 15 copies of an explanatory letter for FE addresses
- 15 copies of an explanatory letter for young people (at core or FE addresses) who were asked to participate in the survey
- 1 set of showcards
- 1 stand-alone showcard, including the definition of medical research to be used in the survey
- 1 set of dress rehearsal project instructions
- 1 copy of a dress rehearsal feedback form
- 15 copies of the National Centre’s leaflet for respondents, which were to be left with each respondent

The dress rehearsal feedback form is contained in Appendix A of this report. As other documents did not change substantially between the dress rehearsal and the main stage, Appendix B can be consulted to locate comparable versions of these.

The face-to-face debrief was held on 1<sup>st</sup> July 2008. Interviewers were asked to complete an electronic template for the adult and young person questionnaires, recording all of their feedback, prior to the de-brief. These formed the basis of the discussion at the debrief.

## Outcomes

Interviewers attempted to make contact with 139 of the 150 core addresses issued to them (although 10 of these addresses turned out to be deadwood. Adult interviews were achieved at 44 of the core addresses. In addition, interviews took place with a young person at eight of these core addresses.

Interviewers attempted to make contact with 63 FE addresses (they were asked to do this when it could not be definitively established at the core address that there were no young people aged between 14 and 18 years at the FE address). In 38 cases, it emerged that there was no eligible respondent at the FE address. Overall, seven interviews were obtained with young people at FE addresses, contributing to a total of 59 interviews obtained overall (44 with adults and 15 with young people).

The characteristics of the achieved adult and young person samples suggest that these represented a good cross-section of the population as a whole in terms of demographic characteristics, and in relation to some of the characteristics likely to be associated with attitudes towards science and medical research. Further details are provided in Appendix A.

## Post-pilot modifications

A number of issues were highlighted as a result of the dress rehearsal which were addressed, in consultation with the Wellcome Trust, prior to the main stage. These included the following:

- It was felt that the survey name may have been off-putting to some potential respondents, who had little interest in the topics of science and medicine, or felt that they would need to know more about them in order to respond. The word “science” was regarded as unnecessary and potentially off-putting for the adult survey. However, young people thought it was important that the word “science” was included in the title of their survey (due to the fact that it is the main focus of their questions). It was therefore recommended that two different field names for the survey were used – with only the adult survey name not containing the word “science”.
- Interviewers felt that having a survey-specific leaflet to use with respondents on the doorstep would have been helpful. Further consideration of this issue was undertaken. Ultimately, to avoid responses in relation to knowledge and perceptions of medical research being influenced by the content of such a leaflet, the decision was taken not to develop such a document for this study.
- The young person interview lasted 38 minutes on average while the average length of the adult interview was 59 minutes, indicating that approximately 14 minutes of material from the latter needed to be cut for the main stage. A reduced adult survey length was achieved by simplifying some of the longer question stems and placing additional information on showcards and as interviewer instructions, and by reducing the numbers of questions

asked by prioritising those which were most important to the funder and which would yield the most analytically useful data.

A full report on the findings of the dress rehearsal pilot can be found in Appendix A.

### **3.4 Questionnaire development**

In addition to piloting the questionnaire, the program was also tested by the Research and Operations teams. Checks were made to confirm the accuracy and sense of questionnaire wording and response options, as well as the accuracy of showcard references, and to ensure that adults and young people, within core and FE addresses, were routed to the appropriate question sets. Scenarios were tested to ensure that the survey routing was correct and that respondents would not be asked inappropriate questions dependent on their circumstances. There were also checks for screen layout, spelling and the clarity of instructions to interviewers.

## 4 Data collection

Fieldwork was undertaken by interviewers drawn from NatCen's regular panel and was conducted using face-to-face computer assisted interviewing.

### 4.1 Advance letter

Interviewers were supplied with letters to send to all sampled core addresses two days before they intended to visit. This provided a way of an introduction to the survey and explained to respondents how their addresses had been selected and what their participation would involve. The letter also contained a contact number for a member of NatCen's Operations Department in case the respondent had any queries. Interviewers were also provided with advance letters to post through the letter boxes of FE addresses, once it became apparent that contact needed to be attempted, and to pass on to selected young person respondents, at both core and FE addresses. Copies of the three different types of advance letter can be found in Appendix B.

### 4.2 Briefings

Eight six-hour briefings were held between 4<sup>th</sup> and 12<sup>th</sup> December 2009. The briefings were conducted by researchers from NatCen.

The briefings covered the aims and background of the survey, procedures for starting work and selecting a respondent at the core address, procedures for screening and making contact with FE addresses, an overview of the two questionnaires and strategies for gaining respondents' cooperation. The sessions also included a practice run-through of both questionnaires. Interviewers were given a copy of the project instructions. They were also supplied with a document containing sources of further information on science and medical research that could be left with the respondent after the interview was completed (copies of each document can be found in Appendix B).

### 4.3 Scheduling of interviews

The standard guidelines issued to all of NatCen's face-to-face interviewers about the timing and the number of calls they should make to an address in the sample were followed by the interviewers on this study. These stipulate that a minimum of six calls must be made at each address before accepting a non-contact or refusal, with a maximum of nine calls, as it is envisaged that further effort beyond that point is unlikely to yield many more productive interviews. Interviewers recorded details of attempts to make contact with each address and, where selected, each respondent, on the relevant Address Record Form (ARF); two different versions were available – one for core addresses and one for FE addresses. In addition, as part of the core ARF, interviewers completed an FE information sheet, indicating the outcome, if any, of their screening at the core address for the presence of young people at the associated FE addresses. In circumstances where it emerged that a particular FE address needed to be visited, as insufficient information had been obtained at the associated core address or it was thought that there was a young person resident at the FE address, an FE ARF was opened. Copies of the two ARFs and the FE Information Sheet can be found in Appendix B.

Within core addresses, where there was also an eligible young person respondent, interviewers were asked to interview the selected adult respondent first. This was to avoid the situation where a

young person was interviewed at a core address and the selected adult respondent subsequently refused to be interviewed.

Progress on fieldwork was monitored using NatCen's computerised booking-in system.

#### **4.4 Quality control**

The time, date and outcome of all calls were recorded by interviewers and checks were made by field management. Non-contacts were not accepted unless the pattern, as well as the number of calls conformed to the basic requirements that normally at least one call must be made at a weekend, and one on a weekday evening and at least six calls must be made in total.

#### **4.5 Fieldwork progress**

Once contact with an address had been made, the final outcome relating to that case was transmitted to NatCen's Brentwood office by the interviewers via telephone modem. With this information, fieldwork progress could be updated on a daily basis.

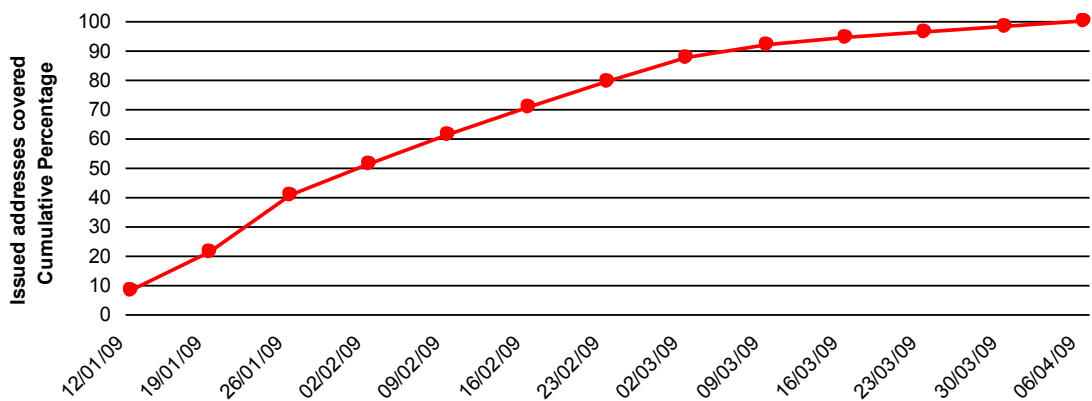
Information on fieldwork progress was reported on a fortnightly basis to the Wellcome Trust.

Using this information, researchers were able to identify potential problems with fieldwork. This data influenced decisions about re-issuing unproductive cases, as detailed in Chapter 5, and was used to inform quality control.

Fieldwork lasted for a total of 13 weeks. Figure 4-1 plots the progress made by interviewers across this period in covering core addresses. Progress on FE addresses was less easy to track because we did not have a target number of addresses that we were aiming to contact (the number requiring contact being dependent on the result of the screening undertaken at core addresses). However, our Operations team was in regular contact with project managers across the different fieldwork areas and so was able to ensure that work on FE addresses was progressing at an appropriate rate.

**Figure 4-1 Summary of fieldwork progress (core addresses)**

Base: All issued core addresses



## 4.6 Thank you letter and voucher

A letter was sent to all respondents who took part in the survey, thanking them for their cooperation. In addition to this, all respondents who completed the questionnaire were sent a £10 gift voucher as a token of appreciation for their participation. A copy of the letter can be found in Appendix B.

## 4.7 Interview length

The adult and young person interviews took an average of 53 minutes and 39 minutes respectively to complete.



## 5 Response

This chapter looks at the fieldwork outcomes for the survey. We start by presenting separate response rates for adults and young people, then give a full breakdown of individual outcomes for the sample. The last section of the chapter offers some thoughts and considerations about the achieved outcomes and response rates.

### 5.1 Adult response rate

Table 5-1 below presents a breakdown of the fieldwork outcomes for adults in our sample. Response is calculated as a range from a lower limit where all unknown eligibility cases (for example, address inaccessible, or unknown whether address is residential) are assumed to be eligible and therefore included in the unproductive outcomes, to an upper limit where all these cases are assumed to be ineligible (and are therefore excluded from the response calculation).

In total, we achieved 1,179 productive interviews with adult respondents aged 18 years and over. The main reason for unproductive outcomes was refusal – 40 per cent of eligible addresses were unproductive for this reason. Non-contacts accounted for four per cent of the eligible addresses, with a further five per cent covered by other unproductive outcomes, such as being away or ill during fieldwork.

**Table 5-1 Fieldwork outcomes for adult sample**

	Number	% of issued sample	Lower limit of response rate (%)	Upper limit of response rate (%)
<b>Addresses issued</b>	2650	100.0		
Definitely out of scope	251	9.5		
<b>Upper limit of eligible cases</b>	2399	90.5	100.0	
Uncertain eligibility	56	2.1	2.3	
<b>Lower limit of eligible cases</b>	2343	88.4	97.7	100.0
<b>Interview achieved</b>	1179		49.1	50.3
Interview not achieved	1164		48.5	49.7
Refused <sup>1</sup>	940		39.2	40.1
Not contacted <sup>2</sup>	97		4.0	4.1
Other non-response <sup>3</sup>	127		5.3	5.4

<sup>1</sup> 'Refused' comprises refusals before selection of an individual at the address, refusals to the office, refusal by the selected person, 'proxy' refusals (on behalf of the selected respondent) and broken appointments after which the selected person could not be re-contacted

<sup>2</sup> 'Non-contacted' comprises households where no one was contacted and those where the selected person could not be contacted

<sup>3</sup> 'Other non-response' includes people who were ill or away during the entire fieldwork period, otherwise physically or mentally incapable, or who had language difficulties.

### 5.2 Young person response rate

Table 5-2 and Table 5-3 below show a breakdown of the fieldwork outcomes for the young people (aged 14-18) in our sample. Separate figures are presented for core and FE addresses. It is not possible to calculate meaningful response rates for FE addresses in the same way as for core addresses. This is because we cannot be certain of the status of those addresses which were not

visited by the interviewer. For instance, informants at core addresses may have wrongly stated that there were no young people available at FE addresses (when in fact there were eligible respondents present). We therefore focus on those FE addresses where young people were identified as being eligible to take part and calculate a “response rate” based on the proportion of these eligible young people who agreed to take part.

A total of 374 young person interviews were obtained, 121 at core addresses and 240 via focused enumeration. At core addresses, 86 per cent of those young people who were eligible agreed to be interviewed. At FE addresses, 64 per cent of the young people identified agreed to be interviewed. In both cases, the proportion of eligible young people agreeing to take part was higher than anticipated<sup>5</sup>

**Table 5-2 Fieldwork outcomes for young people (core addresses)**

	Number	% of issued sample	% of eligible young people
<b>Addresses issued</b>	2650	100.0	
Address out of scope <sup>1</sup>	251	9.5	
14-18 identified as resident at address	239	9.0	
14-18 year old not eligible for interview as no productive adult interview <sup>2</sup>	99	3.7	
<b>14-18 year old eligible for interview</b>	140	5.3	100.0
<b>Interview achieved</b>	121		86.4
Interview not achieved	19		13.6
Refused	18		12.9
Not contacted	1		0.7
Other non-response	0		0.0

<sup>1</sup> Addresses identified as deadwood i.e. unoccupied and/or non-residential

<sup>2</sup> We only attempted to interview 14 to 18 year olds at core addresses if a productive interview was obtained with an adult aged 18+.

**Table 5-3 Fieldwork outcomes for young people (FE addresses)**

	Number	% of issued sample	% of eligible young people
<b>Addresses issued</b>	7950	100.0	
<b>Addresses identified for direct screening<sup>1</sup></b>	1621	20.4	
Address out of scope <sup>1</sup>	17		
No eligible respondent 14-18	1162		
Unknown whether eligible respondent 14-18	48		
<b>Eligible respondent (14-18) identified</b>	394	5.0	100.0
<b>Interview achieved</b>	253		64.2
Interview not achieved	141		35.8
Refused	127		32.3
Not contacted	5		1.3
Other non-response	9		2.0

<sup>1</sup> Interviewers did not follow up at FE address if a) they were able to identify the address as being deadwood b) it was too far away from the core address c) the core address confirmed that there were no eligible respondents living at the FE address.

<sup>5</sup> We set different response targets for young people depending on whether they were to be found at core (65%) or FE (50%) addresses. It was felt that the young people identified at core addresses would be more likely to participate given that (in order for the young person to be eligible) an adult at the address would have already agreed to undertake the adult interview.

## 5.3 Outcomes

The full set of outcomes for core addresses is provided in Table 5-4 below. We also provide further details of the unproductive outcome codes for those young people identified as being eligible for interview at either core or FE addresses (Table 5-5 and Table 5-6).

**Table 5-4 Full breakdown of core address outcomes**

	Number	% of issued sample	Lower limit of response rate (%)	Upper limit of response rate (%)
<b>Issued sample</b>	<b>2650</b>	<b>100.0</b>		
<b>Ineligible</b>	<b>251</b>	<b>9.5</b>		
Not yet built / under construction	5	0.2		
Demolished / derelict	20	0.8		
Vacant / empty housing unit	160	6.0		
Non-residential address	31	1.2		
Address occupied, no residents	20	0.8		
Communal establishment - no private dwellings	4	0.2		
No eligible respondent 18+	4	0.2		
Other ineligible	7	0.3		
<b>Potentially eligible</b>	<b>2399</b>	<b>90.5</b>	<b>100.0</b>	
<b>Unknown eligibility</b>	<b>56</b>	<b>2.1</b>	<b>2.3</b>	
Inaccessible	5	0.2	0.2	
Unable to locate address	3	0.1	0.1	
Unknown if address residential due to non-contact	3	0.1	0.1	
6Residential-Don't know if eligible person(s)- no contact	7	0.3	0.3	
Other unknown eligibility (due to non-contact)	8	0.3	0.3	
Information refused about whether address residential	2	0.1	0.1	
Contact but could not confirm resident HH	4	0.2	0.2	
Information refused about whether resident(s) are eligible	14	0.5	0.6	
Unable to confirm eligibility of -lack of knowledge	4	0.2	0.2	
Unable to confirm eligibility - language problems	2	0.1	0.1	
Other unknown eligibility	4	0.2	0.2	
<b>Definitely eligible sample</b>	<b>2343</b>	<b>88.4</b>	<b>97.7</b>	<b>100</b>
<b>Non-contact</b>	<b>97</b>		<b>4.0</b>	<b>4.1</b>
No contact with anyone at the address	35		1.5	1.5
Contact made at address. but not at selected DU	3		*	*
Contact made but not with responsible adult	5		*	*
Contact made but not with selected respondent	54		2.3	2.3
<b>Refusal</b>	<b>940</b>		<b>39.2</b>	<b>40.1</b>
Office refusal	27		1.1	1.2
Information refused about number of DUs at address	5		0.2	0.2
Information about number of eligible respondents refused	101		4.2	4.3
Refusal by selected respondent	661		27.6	28.2
Refusal by proxy	64		2.7	2.7
Refusal during the interview	2		0.1	0.1
Broken appointment - no re-contact	80		3.3	3.4

**Table 5-4 (Continued) Full breakdown of core address outcomes**

	Number	% of issued sample	Lower limit of response rate (%)	Upper limit of response rate (%)
<b>Other non-productive</b>	<b>127</b>		<b>5.3</b>	<b>5.8</b>
Ill at home during field period	30		1.3	1.3
Away / in hospital throughout field period	29		1.2	1.2
Respondent physically or mentally unable / incompetent	41		1.7	1.7
No information as all household members mentally/physically incompetent	2		0.1	0.1
Household language barrier	2		0.1	0.1
Language barrier with target respondent	12		0.5	0.5
Lost interview	1		*	*
Other non-response	10		0.4	0.4
<b>Productive</b>	<b>1179</b>		<b>49.1</b>	<b>50.3</b>
Fully productive adult interview (no eligible young person)	1037		43.2	44.3
Fully productive adult interview + productive young person interview	121		5.0	5.2
Fully Productive adult interview but unproductive young person interview	19		0.8	8.1
Partially productive adult interview	2		0.1	0.1

**Table 5-5 Full breakdown of outcome codes for eligible young people (core addresses)**

	Number	% of eligible young people
<b>Core addresses where eligible respondent (14-18) identified</b>	<b>140</b>	<b>100.0</b>
<b>Fully productive</b>	<b>121</b>	<b>86.4</b>
<b>Non-contact</b>	<b>1</b>	<b>0.7</b>
No contact with parent or guardian to obtain consent	1	0.7
<b>Refusal</b>	<b>18</b>	<b>12.9</b>
Refusal by selected respondent	8	5.7
Refusal by parent or guardian	8	5.7
Broken appointment – no Re-contact	2	1.4

**Table 5-6 Full breakdown of outcome codes for eligible young people (FE addresses)**

	Number	% of eligible young people
<b>FE addresses where eligible respondent (14-18) identified</b>	<b>394</b>	<b>100.0</b>
<b>Fully productive</b>	<b>253</b>	<b>64.2</b>
<b>Non-contact</b>	<b>5</b>	<b>1.3</b>
No contact with selected young person	5	1.3
<b>Refusal</b>	<b>128</b>	<b>32.5</b>
Information about number of eligible respondents refused	2	0.5
Refusal by selected respondent	52	13.2
Refusal by parent or guardian	59	15.0
Refusal during the interview	1	*
Broken appointment – no re-contact	14	3.6
<b>Other non-productive</b>	<b>8</b>	<b>2.0</b>
Away / in hospital throughout field period	3	0.8
Respondent physically or mentally unable / incompetent	3	0.8
Other unproductive	2	0.5

## 5.4 Further information on response

The response rate among adult aged 18 and over was lower than anticipated. Based on the response rates obtained on other attitudinal surveys such as the annual British Social Attitudes survey, we were aiming for an adult response rate of 55 per cent. However, our achieved response rate for adults was 49 per cent. Possible reasons why response was lower than anticipated, and the measures taken to try and maximise response, are discussed in more detail below.

At both core and FE addresses the proportion of eligible young people agreeing to take part was higher than anticipated. However, despite the higher than expected response rate among young people, we achieved slightly fewer young person interviews than anticipated. The reason for this was that focused enumeration identified a smaller number of eligible 14 to 18 year olds than was expected.

### Efforts made to maximise response

Efforts were made to maximise response rates before the survey was launched. Many of these focused on the way in which the study was presented to potential respondents. For example, we deliberately chose a survey name which avoided the term “medical research”. Following feedback from the dress rehearsal pilot, which found that mentioning “science” was attractive to young people but off-putting to adults, we used different survey field names for adults and young people. All respondents (adults and young people) were sent a £10 conditional incentive along with a thank you letter after taking part, and were informed about this in their respective advance letters in order to encourage response.

During fieldwork we implemented an extensive reissue strategy with selected non-productive cases being reissued, often to a different interviewer, for a second attempt. Reissues concentrated on core addresses in an attempt to raise the adult response rate. Given the high response rate among eligible young people, we did not reissue any FE addresses. Overall 781 core addresses were reissued, resulting in 116 additional interviews. This represents a conversion rate of 15 per cent.

## Further information on refusals

The main reason we were unable to obtain productive interviews with respondents identified as eligible was because of refusals, either by the selected respondent themselves or a proxy (see Section 5.3 above). We collected some further information from selected respondents about their reasons for refusing as part of the ARF completion. The majority of the reasons given were not specific to this particular survey and included reasons such as “never takes part in surveys”, “can’t be bothered” and “inconvenient time”. However, a significant minority (13 per cent of those giving reasons) said that they refused to take part because the subject matter was not interesting or relevant to them. Some people were also put off because they or someone they knew had recently had health problems.

Feedback from interviewers further suggests that, despite our best efforts, potential respondents were sometimes intimidated by the prospect of having to answer questions about medical research, a topic about which they felt they had little knowledge.

Although respondents may have been reluctant to take part, there is little evidence that they were put off once the interview was started. We had a very small number of partial interviews (two partial productives and three partial unproductives) indicating that only a few respondents were led to abandon the interview part way through. Two-thirds (68%) of adult respondents and three-quarters (77%) of 14-18 year old respondents agreed that the Wellcome Trust could contact them again in the event of future research, suggesting that the overall experiences of being interviewed was positive.<sup>6</sup>

Among eligible respondents at FE addresses, there was a relatively high proportion of refusals by a parent/guardian; 46 per cent of all refusals at FE addresses came from a parent or guardian rather than the selected young person. This is despite the fact that interviewers were briefed to try and talk to the young person directly wherever possible.

When the Wellcome Trust Monitor is repeated, the positive feedback provided by those respondents who took part in the first wave could perhaps be used as a way of selling the survey to potential respondents as an interesting and enjoyable experience; for instance, on other surveys, quotations from respondents to previous waves have been included on survey-related materials such as the advance letter. It might also be worth revisiting the idea of a survey-specific leaflet and considering whether such a document could be produced without limiting the interviewer’s ability to obtain uncontaminated information from respondents about their perceptions and knowledge of medical research.

## 5.5 Characteristics of the achieved sample

The relatively low response rate among adult respondents raises a concern that those who did agree to take part in the Wellcome Trust Monitor may be different from those who did not take part

---

<sup>6</sup> The proportion of adults agreeing to be re-contacted on this study was, however, lower than on British Social Attitudes. The proportion of BSA respondents agreeing to be re-contacted in 2007 was 80%.

and that this may introduce bias into any results reported. It is possible, for example, that those who agreed to take part may have had higher than average levels of education and be more interested in and engaged with the subject of medical research compared with the general population. To try and identify whether this was in fact the case we conducted some analysis of our respondents as compared against both the general population and respondents to other surveys. The results of this analysis, presented below, suggest that the achieved sample on the Wellcome Trust Monitor is comparable with that obtained by other random probability sample surveys. We are, therefore, reassured that the data collected in this study are broadly representative of the views of the general population.

Table 5-7 and 5.8 show the unweighted distributions of our achieved adult and young person samples with respect to some key characteristics, namely age, sex, and geographical region. We have compared these distributions with the latest (mid-2007) figures for the population as a whole.

From Table 5.7 it appears that our sample of adult respondents over represents women and (as is often the case) under represents younger respondents and those living in London. The degree of over/under representation is fairly typical of general population surveys.

Table 5.8 shows that our sample of young people is more evenly split between young men and young women but over-represents younger respondents, particularly 14 year olds, at the expense of 18 year olds (this is partly explained by the fact that some 18 year olds were selected for the adult sample). As will be discussed in Chapter 6, we have been able to correct for under/over representation on the basis of sex, age or region via the weighting procedures.

**Table 5-7 Characteristics of adults respondents (aged 18+)**

	Number of survey respondents	% of achieved adult sample	% of adult population
<b>Sex</b>			
Men	472	40.0	48.5
Women	707	60.0	51.5
<b>Age</b>			
18-29	157	13.3	20.3
30-39	197	16.7	17.6
40-49	207	17.6	18.7
50-59	193	16.4	15.6
60-69	185	15.7	12.9
70+	240	20.4	14.8
<b>Region</b>			
North East	47	4.0	4.2
North West	141	12.0	11.2
Yorkshire and the Humber	105	8.9	8.5
East Midlands	80	6.8	7.2
West Midlands	111	9.4	8.7
East	109	9.2	9.3
London	114	9.7	12.4
South East	150	12.7	13.6
South West	113	9.6	8.6
Wales	69	5.9	4.9
Scotland	104	8.8	8.6
Northern Ireland	36	3.1	2.8

**Table 5-8 Characteristics of young person respondents (14-18)**

	Number of survey respondents	% of achieved young person sample	% of 14 to 18 population
<b>Sex</b>			
Men	183	48.9	51.5
Women	191	51.1	48.5
<b>Age</b>			
14	93	24.9	19.2
15	84	22.5	20.0
16	82	21.9	20.4
17	77	20.6	20.2
18	38	10.2	20.2
<b>Region</b>			
North	103	27.5	24.7
Midlands	52	13.9	16.6
South	112	29.9	31.4
London	36	9.6	10.8
Wales	25	6.7	5.1
Scotland	29	7.8	8.2
Northern Ireland	17	4.5	3.2

We also wished to investigate the extent to which our achieved sample of adult respondents is representative of the general population in terms of other characteristics considered to be important in influencing attitudes to medical research, namely education, social background and self-rated health. We have therefore compared the characteristics of respondents to the Wellcome Trust Monitor with the characteristics of respondents to two other general population surveys: the



National Travel Survey (NTS) and the Health Survey for England (HSE). Both of these surveys are well-established and obtain comparatively high response rates. The response rate for NTS 2008 was 60 per cent whilst the household response rate for HSE 2007 was 66 per cent (see Craig and Shelton, 2008; Anderson et al., forthcoming).

The comparisons presented in Table 5-9 and Table 5-10 indicate that the profile of adult respondents to the Wellcome Trust Monitor was broadly similar to respondents to these two surveys. Notably, the proportion of respondents obtaining higher education qualifications was similar across all three surveys.<sup>7</sup> This is contrary to expectations as, if anything, it was anticipated that the scientific subject matter covered by the Wellcome Trust Monitor might have deterred less educated respondents and led to a bias in favour of more educated respondents.

Table 5-9 also suggests that, compared with HSE, the Wellcome Trust Monitor may slightly over-represent those in good health. Interviewer feedback from our pilot studies suggests it is possible that some people with recent experience of medical treatment were put-off answering a survey about medical research. However, it is also possible, given its focus on health conditions and the provision of a nurse visit, that HSE over represents people with poor health and specific health concerns.

**Table 5-9 Comparison of Wellcome Trust Monitor and HSE respondents, by education and social class**

*Base: all adult respondents 18+ resident in England*

	Wellcome Trust Monitor		HSE	
	% (weighted)	% (unweighted) <sup>8</sup>	% (weighted)	% (unweighted)
<b>Highest educational qualification achieved</b>				
Higher education qualification	30	29	31	30
A Level or equivalent	14	14	14	13
GCSE or equivalent	21	20	22	21
CSE or equivalent	10	10	5	5
No qualification	24	26	27	29
<b>Social Class (5 Category NSSEC)<sup>9</sup></b>				
Managerial and professional	36	36	33	33
Intermediate	10	10	13	13
Small employers and own account workers	8	7	9	9
Lower supervisory and technical	7	6	9	8
Semi-routine and routine manual	31	32	32	32
<b>Self-rated health</b>				
Very good	37	36	34	32
Fairly good	43	44	41	41
Fair	15	15	19	20
Bad	3	3	5	6
Very bad	2	2	2	2
<i>Minimum base:</i>	<i>987</i>	<i>1725</i>	<i>6682</i>	<i>6698</i>

<sup>7</sup> "Highest educational qualification obtained" is a derived variable; the three surveys all asked slightly different questions in order to obtain the information from which this variable was derived.

<sup>8</sup> The Wellcome Trust Monitor selected just one adult aged 18 years and over per address whereas both HSE and NTS interviewed all eligible adults in the household. To ensure comparability across the surveys, the Wellcome Trust Monitor "unweighted" figures are based on data weighted by an individual selection weight (see Chapter 6 for details of weighting).

<sup>9</sup> Percentages do not sum to 100% as the base includes those for whom no information is available (respondents who have never had a job for example).

**Table 5-10 Comparison of Wellcome Trust Monitor and NTS respondents, by education and social class**

Base: all adult respondents 18+ resident in Great Britain

	Wellcome Trust Monitor		NTS	
	% (weighted)	% (unweighted)	% (weighted)	% (unweighted)
<b>Highest educational qualification achieved</b>				
Degree level qualification	20	18	22	21
Other qualification below degree level	56	55	58	58
No qualification	24	26	20	21
<b>Social Class (5 Category NSSEC)</b>				
Managerial and professional	35	35	34	33
Intermediate	10	10	12	12
Small employers and own account workers	8	8	8	8
Lower supervisory and technical	7	7	9	9
Semi-routine and routine manual	32	33	32	33
Base:	1146	2037	16612	16459

# 6 Weighting

## 6.1 Overview

The dataset contains two weight variables. The weight "WtAd" is the adult weight and should be used when analysing data for adult respondents, aged 18 years and over. The weight "WtYP" is the young person weight and should be used when analysing data for young person respondents aged 14 to 18 years. Analysis should always be conducted separately for adults and young people; the weights are not set up to allow analysis for all respondents together.

The two weights are discussed in turn below.

## 6.2 Adult weight

The adult weight WtAd:

- Adjusts for differential selection probabilities resulting from the selection of one dwelling unit per address and one adult per dwelling unit.
- Adjusts for differential non-response by region and, separately, by age and sex thereby making the sample representative of the population on these variables.

The weights were created in a series of steps detailed below, starting with the selection weights.

### Dwelling unit selection weight

One dwelling unit (DU) was selected at each address. Dwelling units at addresses comprising more than one DU therefore had a lower chance of selection than those at addresses comprising a single DU. To correct for this, a dwelling unit selection weight was created. This was equal to the number of DU's found at the address. The weight was trimmed at 3 to avoid a small number of very high weights. These would inflate the standard errors and reduce the precision of the survey estimates, causing the weighted sample to be less efficient.

### Adult selection weight

One adult aged 18 or over was interviewed at each selected dwelling unit; adults living in DUs with one or more other adults therefore had a lower chance of selection than those in DUs containing only one adult. To correct for this, an adult selection weight was created. This was equal to the number of adults in the DU. The weight was trimmed at 4.

### Combined selection weight

The dwelling unit selection weight and the adult selection weight were combined (multiplied together) to create one selection weight for each adult in the sample.

### Calibration to the population

The next step was to take the weighted sample and to 'calibrate' the totals in each region (GOR), and each of twelve age/sex categories, to population totals derived from the latest (mid-2007) population estimates for the UK. Calibration adjusts a set of input weights to sum to the totals specified in each category. This step adjusts for differential non-response by region and (separately) by age and sex.

After calibration, the total numbers in the weighted sample equated to those in the UK population as shown in Table 6-1 and Table 6-2 below.

**Table 6-1 UK adults (18 and over), by region**

Region	Number of adults 18+	% of adult population
North East	2,032,197	4.2
North West	5,365,559	11.2
Yorkshire and the Humber	4,064,484	8.5
East Midlands	3,464,506	7.2
West Midlands	4,181,848	8.7
East	4,434,413	9.3
London	5,929,173	12.4
South East	6,503,973	13.6
South West	4,121,091	8.6
Wales	2,343,014	4.9
Scotland	4,096,793	8.6
Northern Ireland	1,327,281	2.8
Total	47,864,332	100.0

**Table 6-2 UK adults (18 and over), by age and sex**

Age group	Men		Women	
	Number of adults 18+	% of population	Number of adults 18+	% of population
18-29	4,953,979	10.4	4,778,574	10.0
30-39	4,189,763	8.8	4,237,511	8.9
40-49	4,430,899	9.3	4,533,361	9.5
50-59	3,687,295	7.7	3,790,694	7.9
60-69	2,997,125	6.3	3,183,298	6.7
70+	2,936,250	6.1	4,145,583	8.7
Total	23,195,311	48.5	24,669,021	51.5

## Scaling the weights

The final step was to re-scale the weights so that the weighted total for the whole sample was equal to the unweighted total (1179); this results in weights with an average of 1. As part of this process, some extreme weights were trimmed to be equal to the next highest weight (approx 3.67).

### 6.3 Young person weight

The young person weight WtYP:

- Adjusts for differential selection probabilities resulting from the selection of one young person aged 14-18 in each selected dwelling unit
- Adjusts for differential non-response by region and, separately, by age and sex thereby making the sample representative of the population on these variables.

The weights were created in a series of steps detailed below, starting with the selection weights.

#### Dwelling unit selection weight

All young people interviewed were found in addresses comprising a single dwelling unit. There was therefore no need for a dwelling unit selection weight.

## Young person selection weight

At both core and FE addresses, one young person aged 14-18 was interviewed. Those young people living with other 14-18 year olds therefore had a lower chance of selection than those living at addresses containing only themselves and one or more adults aged 19 or over. To correct for this, a young person selection weight was created.

The calculation of these weights had to take into account the fact that, at a core address, one young person aged 14-18 was picked *after* the selection of one adult aged 18 or over. Prior to selection therefore, an 18 year had a chance of being picked either for the adult sample or for the young person sample. The weights for 18 year olds were therefore calculated differently from the weights for those aged 14-17. In both cases, the calculation took into account the relative probabilities of the address having been picked either as a core address or as a FE address, and the ages of other household members.

## Calibration to the population

The next step was to take the weighted sample and to 'calibrate' the totals in each of seven regions (based on GOR), and each of ten age/sex categories, to population totals derived from the latest (mid-2007) population estimates for the UK. Calibration adjusts a set of input weights to sum to the totals specified in each category. This step adjusts for differential non-response by region and (separately) by age and sex.

Some regions were collapsed into the following groups due to small numbers:

- North = North East + North West + Yorkshire & Humber
- Midlands = East Midlands + West Midlands
- South = East of England + South East + South West

After calibration, the total numbers in the weighted sample equated to those in the UK population as shown in Table 6-3 and Table 6-4 below.

**Table 6-3 UK 14-18 year olds, by region**

Region	Number of 14-18 year olds	% of 14-18 population
North	977,023	24.7
Midlands	654,510	16.6
South	1,239,023	31.4
London	427,989	10.8
Wales	199,850	5.1
Scotland	323,963	8.2
Northern Ireland	127,416	3.2
Total	3,949,774	100

**Table 6-4 UK 14-18 year olds, by age and sex**

Age group	Young men		Young women	
	Number of 14-18 year olds	% of 14-18 population	Number of 14-18 year olds	% of 14-18 population
14	390,132	9.9	369,852	9.4
15	404,902	10.3	383,424	9.7
16	415,336	10.5	389,480	9.9
17	410,939	10.4	385,903	9.8
18	412,009	10.4	387,797	9.8
Total	2,033,318	51.5	1,916,456	48.5

## Scaling the weights

The final step was to re-scale the weights so that the weighted total for the whole sample was equal to the unweighted total (374); this results in weights with an average of 1.

## 7 Sampling errors

No sample precisely reflects the characteristics of the population it represents, because of both sampling and non-sampling errors. If a sample were designed as a random sample (if every individual had an equal and independent chance of inclusion in the sample), then we could calculate the sampling error of any percentage,  $p$ , using the formula:

$$s.e. (p) = \sqrt{\frac{p(100 - p)}{n}}$$

where  $n$  is the number of respondents on which the percentage is based. Once the sampling error had been calculated, it would be a straightforward exercise to calculate a confidence interval for the true population percentage. For example, a 95 per cent confidence interval would be given by the formula:

$$p \pm 1.96 \times s.e. (p)$$

Clearly, for a simple random sample (srs), the sampling error depends only on the values of  $p$  and  $n$ . However, simple random sampling is almost never used in practice because of its inefficiency in terms of time and cost.

As noted above, the Wellcome Trust Monitor sample, like that drawn for most large-scale surveys, was clustered according to a stratified multi-stage design into 106 postcode sectors (or combinations of sectors). With a complex design like this, the sampling error of a percentage giving a particular response is not simply a function of the number of respondents in the sample and the size of the percentage; it also depends on how that percentage response is spread within and between sample points.

The complex design may be assessed relative to simple random sampling by calculating a range of design factors (DEFTs) associated with it, where:

$$\text{DEFT} = \frac{\text{Standard deviation of estimator with complex design, sample size } n}{\text{Standard deviation of estimator with srs design, sample size } n}$$

and represents the multiplying factor to be applied to the simple random sampling error to produce its complex equivalent. A design factor of one means that the complex sample has achieved the same precision as a simple random sample of the same size. A design factor greater than one means the complex sample is less precise than its simple random sample equivalent. If the DEFT for a particular characteristic is known, a 95 per cent confidence interval for a percentage may be calculated using the formula:

$$p \pm 1.96 \times \text{complex sampling error } (p) \\ = p \pm 1.96 \times \text{DEFT} \times \sqrt{\frac{p(100 - p)}{n}}$$

Calculations of sampling errors and design effects were made using the statistical analysis package SPSS.

**Table 7-1 Complex standard errors and confidence intervals of selected variables for adults**

Variable	Estimate description	Estimate	Standard error	Design effect	Unweighted count	95% confidence interval	
						Upper	Lower
Interest in medical research	% very interested	34	1.63	1.19	418	30.5	37.0
Understanding of the term DAN	% at least some understanding of the term or better	90	1.17	1.35	1,037	87.6	92.2
Whether has tried to find out any information on medical research in the past year	% have tried to find information	39	1.84	1.30	437	35.8	43.1
Perceived amount of information received about medical research	% too little or much too little information	47	1.95	1.34	547	43.3	51.0
Importance of undertaking medical research to test methods of identifying, preventing and treating illnesses and diseases	% very important	84	1.47	1.38	981	80.9	86.8
Whether medical research will lead to an improvement in quality of life in the UK over the next 20 years	% definitely will	41	1.88	1.31	462	36.9	44.4
Optimism about medical advances as a result of genetics research	% at least somewhat optimistic	84	1.30	1.22	975	81.4	86.6
Level of trust in medical research charities to provide accurate information about medical research	% complete trust	12	1.08	1.16	138	9.7	14.0
Willingness to take part in a medical research project testing a new drug or treatment	% at least fairly willing	30	1.44	1.08	360	27.3	33.1
Importance of teaching science up to the age of 16 in school	% very important	79	1.53	1.28	921	75.6	81.7



**Table 7-2 Complex standard errors and confidence intervals of selected variables for young people**

Variable	Estimate description	Estimate	Standard error	Design effect	Unweighted count	95% confidence interval	
						Upper	Lower
Interest in medical research	% very interested	22	2.34	1.08	86	18.1	27.4
Understanding of the term DAN	% at least some understanding of the term or better	93	1.50	1.17	353	89.7	95.8
Whether has tried to find out any information on medical research in the past year	% have tried to find information	51	2.81	1.08	201	45.7	56.8
Perceived amount of information received about medical research	% too little or much too little information	51	3.32	1.28	182	44.9	58.1
Importance of undertaking medical research to test methods of identifying, preventing and treating illnesses and diseases	% very important	77	2.35	1.08	280	72.1	81.5
Whether medical research will lead to an improvement in quality of life in the UK over the next 20 years	% definitely will	40	3.64	1.43	144	33.1	47.5
Optimism about medical advances as a result of genetics research	% at least somewhat optimistic	71	2.71	1.16	266	65.5	76.4
Level of interest in science lessons at school	% at least fairly interested	81	2.05	1.02	301	76.9	85.1
Young people's level of interest in science	% at least fairly interested	69	2.74	1.15	258	63.6	74.5
Whether science is a good area of employment for young people to go into	% agree	81	2.75	1.35	304	74.7	85.8

# 8 Data processing and management

## 8.1 Editing

A number of checks were included in the CAPI programme and carried out by the interviewer when prompted during the interview – for example, if a respondent gave an answer that appeared not to be compatible with an answer provided to a previous question (such as a respondent in their 20s indicating that they had adult children).

Some post-interviewing editing was done by researchers to remove minor inconsistencies between certain factual variables. The specific editing that was done was:

- In the engagement module, to ensure that the answer given at PInfo (whether the respondent could remember the last piece of information they saw/heard/read about medical research *and could provide details*) was consistent with whether the respondent had actually been able to give details at the open-code question that followed. As the first piece of information represented the interviewer’s judgement, the respondent’s actual ability to provide further details was given priority.
- In the current/future study plans module, to ensure that young people who said they were planning to study science in the future did in fact mention at least one science subject among the subjects they were planning to study.

Given that most of the questions asked as part of this study related to the respondents’ own attitudes and it is perfectly possible that one individual may hold a variety of inconsistent attitudes, these were not subject to editing and any inconsistencies in the respondents’ answers remain as given during the interview.

## 8.2 Coding

Post-interview coding was undertaken by members of NatCen’s coder panel using an adapted version of the CAPI program. Coders were briefed by researchers and provided with full instructions (Appendix C).

### Other specify questions

For “other – please specify” questions, coders were asked to check the “other” answers to see whether any could be back-coded into any of the pre-existing codes. Researchers also considered whether any additional codes needed to be added to the code frame, based on the data received from the first 500 interviews (see Appendix C).

### Open questions

The adult interview contained eight open-code questions and the young person interview contained nine open-code questions. These open-code questions were mainly designed to measure respondents’ awareness and knowledge of medical research. Open-code questions were used in order to gain a more accurate picture of what the respondent did or did not know about medical research without giving them any prompts. They also enabled us to obtain a picture of the sorts of language and terms the public use when talking about medical research. Finally, using open-code questions allowed us to collect detailed information about the precise nature of any contact respondents had had with information about medical research, without pre-defining what the nature of that information might be.

Based on the data received from the first 500 interviews, researchers developed code frames for all open-code questions. The code frames were deliberately designed to be very detailed as the Wellcome Trust had a particular interest in the specific words or phrases which respondents used to talk about medical research. All code frames were seen and approved by the Wellcome Trust. All open-code questions had a residual code “98” available where any vague or irrelevant answers could be coded.

Several steps were taken to ensure that verbatim answers to these open-code questions were coded consistently. Coders completed a short coding exercise during the coder briefing, with researchers checking and discussing all answers provided (see Appendix C). The first batch of 50 interviews coded by each interviewer was fully checked by the Operations team. Researchers checked all of the answers coded as “other” to see whether any could be coded to more specific codes or whether additional codes were necessary. Several additional codes were added at this stage (see Appendix C).

In several instances, large numbers of the verbatim answers provided to the open-code questions remain in the “other” category. This is because many respondents tended to provide highly specific answers, identifying a range of different elements, many of which were not identified by more than a small number of respondents across the survey as a whole (and thus not justifying a newly created code).

## Occupation coding

The adult respondents’ job details were coded to the Standard Industrial and Standard Occupational classifications – SIC (2007) and SOC (2000). Industry was classified to a 2-digit level and Occupation to a 4 digit-level.

Where parents’ job details were collected as part of the young person interview, this was done using a simplified set of questions which allowed researchers to code parents’ occupation to the 5 category NS-SEC classification. At core addresses where one of the young person’s parents had been interviewed as the adult respondent, their NS-SEC classification was carried over from the adult interview.

## 8.3 Early programming error

Checks on an early dataset a few weeks into fieldwork revealed a couple of minor routing errors in the CAPI program being used in field. These were quickly rectified and a new version of the program issued to all interviewers on 22<sup>nd</sup> January 2009.

A small number of early cases suffered missing data as a result of these program errors.

- Parents’ demographic module at the start of the young person interview: 41 young person interviews contain missing data regarding their parents’ education or occupation.
- Information sources module: 96 adult respondents who said that they had “very little” or “no trust” in their family or friends to provide information about medical research were not routed to the follow-up question asking them why this was the case.

Where data is missing as a result of program errors, variables have been coded as -3.

## 8.4 Using the dataset for analysis

This next section contains some useful information to bear in mind when using the study dataset for analysis.

### Selecting adult or young person respondents for analysis

The dataset is a combined dataset including all adult respondents (1,179) and all 14-18 year old respondents (374). Analysis should always be conducted separately for adults and young people; the dataset is not set up to allow combined analysis of the two samples as, essentially, they represent two distinct 'populations'.

The variable "adyp" should be used in order to identify whether the respondent is an adult (adyp=1) or young person (adyp=2). Respondents cannot be identified on the basis of age given that 18 year olds at core addresses may have been selected either as the adult respondent or as the young person respondent. The weight variable "WtAd" should be used when analysing data for adult respondents. The weight variable "WtYP" should be used when analysing data for young people.

### Defining "science" subjects

The boundaries of the subject area of "science" can be defined in a number of different ways, ranging from limiting this definition to covering only the pure science subjects (such as biology, chemistry and physics), to including practical subjects based on the application of scientific theory (such as electronics and engineering), to encapsulating all subjects with some scientific theory or content (such as mathematics and psychology). For the purpose of this survey, when asked about "science" qualifications, the young people were presented with as inclusive a list as possible. However, for the analysis undertaken in this report, the second definition of science outlined above was adhered to.<sup>10</sup>

It should be noted that, where adults and young people were asked questions about "science" in general, not pertaining to qualifications, no definition of this term was provided, meaning that the answers provided would have related to personal and subjective definitions that could have reflected any of the three definitions of 'science' suggested above

### Derived variables

A list of the main derived variables used for analysis can be found in Appendix E. All derived variables have a variable label which starts "DV:"

### Missing codes

The following codes for missing data have been used throughout the data set.

-1 = "Not applicable – question not on route"

-2 = "Not applicable – question not part of adult/young person interview"

-3 = "Data missing – program error"

---

<sup>10</sup> Subjects included in our definition of science: physics, chemistry, biology, applied science, environmental science, double award science, geology, electronics, science in society, medicine, nursing, dentistry, veterinary science, engineering, biochemistry. Subjects specifically excluded were: maths, psychology, computer science and information technology (IT).

Don't know and refusal codes are treated as non-missing values and have been coded as follows:  
8 = Refusal  
9 = Don't know

## **Significance testing**

All significance testing for the main substantive report was carried out using the Complex Surveys feature in SPSS. This ensures that significance testing is based on standard errors which take into account design effects as a result of clustering and stratification at the sampling stage.

Two variables on the main data set ("Strata" and "PSU") should be used to specify design effects if using the Complex Surveys feature in SPSS (or a similar package).

# References

Butt, S. , Clery, E., Abeywardana, A, and Phillips, M. (2010) Wellcome Trust Monitor: Survey Report Wellcome Trust: London

Anderson, T., et al. (forthcoming) National Travel Survey 2008 Technical Report NatCen: London

Craig, R. and Shelton, N. (2008) Health Survey for England 2007 Volume 2 Methodology and Documentation NHIC: London