

Wellcome Trust Monitor Wave 2

Research methodologies
and data management

May 2013



Contents

1. Introduction.....	1
Background and objectives	1
Archiving of data.....	2
Report on the findings.....	2
2. The samples	3
Overview	3
Drawing the “core” sample.....	3
The young person “boost” sample	4
Development of materials.....	6
Scope of development work.....	6
Cognitive pilot	6
CAPI field pilot.....	7
Questionnaire development	9
3. Data collection	10
Advance letter and leaflet	10
Briefings.....	10
Scheduling of interviews	10
Quality control	11
Fieldwork progress	11
Incentivisation.....	11
Interview length	13
4. Response	14
Adult response rate	14
Young person response rate	15
Outcomes	17
Further information on response.....	21
5. Weighting	23

Overview	23
Adult weight	23
Young person weight	25
6. Sampling errors	28
7. Data processing and management	31
Editing	31
Coding	31
Using the dataset for analysis.....	32
Appendices	34

1. Introduction

This report describes the research methods used in the second Wellcome Trust Monitor survey, a survey of the UK public conducted by Ipsos MORI on behalf of the Wellcome Trust, carried out in 2012. The Wellcome Trust Monitor is designed to measure the public's awareness, interests, knowledge and attitudes in relation to science, and, in particular, biomedical research. This wave of the survey was designed to provide comparability with findings from the baseline survey, carried out in 2009. The survey also aimed to build on the first wave, by refining the questionnaire and approach to maximise the level and quality of response from respondents.

Background and objectives

The first (baseline) Wellcome Trust Monitor was conducted in 2009 by the National Centre for Social Research (NatCen). The Wellcome Trust Monitor is repeated every three years in order to measure long-term trends in public attitudes.

Questions about a broad range of topics were asked of both adults and young people, while some questions were only asked of adults or young people. For example, adults were also asked about their involvement in medical research and their views on science governance, and young people were also asked about their attitudes towards and their experiences of science education as well as their perceptions of scientific careers.

For over 75 years the Wellcome Trust has worked to promote advances in the fields of animal and human health. Over this time the Trust has become the UK's largest charitable funder of biomedical research, aiming to improve health and wellbeing through new discoveries.

In addition to its support of scientific research, the Wellcome Trust has a long history of promoting public engagement with science and biomedical research. One of the Trust's key public engagement objectives is to listen to views and concerns raised by the public, and to track knowledge and attitudes about the social and ethical dimensions of biomedical research over time. In a modern economy, it is vital that current and future generations of workers and researchers are interested in and engaged with science and biomedical research. The objective is not to 'promote' scientific research to a sceptical public but to enable citizens to critically engage with and contribute to debates regarding the appropriate pace and direction of scientific and technological development, as well as make informed decisions where these developments affect their own lives.

The Wellcome Trust Monitor is an important study that will not only highlight interest in and attitudes towards biomedical science, but will also allow the Wellcome Trust and others to better direct their future public engagement work

A great deal has happened in biomedical research since the first Wellcome Trust Monitor in 2009, making it important to reassess the current state of public attitudes. For instance, in 2010 Dr Craig Venter created the world's first synthetic life form, while in 2012 the Encyclopedia of DNA Elements was published, building on the 12-year-old Human Genome Project

The way science and biomedical research are reported has also evolved. 2012 marked the tenth anniversary of the Science Media Centre (a charity that aims to provide accurate and evidence-based information about science and engineering through the media, set up following the House of Lords Science and Technology Select Committee's third report on Science and Society). The BBC has also expanded its science coverage since 2009 and

labelled 2010 as its “Year of Science”. Earlier controversies in the media’s coverage of medical research have developed, for example Andrew Wakefield was struck off the medical register by the General Medical Council in 2010 over his research which erroneously linked the MMR vaccine with autism. This second Monitor will provide evidence as to how this changing environment has affected the public’s attitudes to science and biomedical research.

Outside the scientific sphere, there have been broader social, political and economic changes in society since 2009. The ongoing economic weakness may have influenced the public’s perception of the relative importance of other issues. The new coalition government, elected in 2010, has also introduced a number of important changes to public services, with an emphasis on decentralisation and local choices.

This report focuses on the technical aspects of the second Wellcome Trust Monitor. Chapter 2 describes the sampling, with development work on the survey and the data collection process outlined in chapters 3 and 4. Response rates are described in detail in chapter 5, while chapter 6 reports on the weighting. Chapter 7 looks at sampling errors and, finally, chapter 8 describes the procedures for the editing and coding of the data.

Archiving of data

A data set with complete documentation will be deposited with the Data Archive (data-archive.ac.uk) at the University of Essex in May 2013. It is anticipated it will become available for scholars and other interested parties several weeks later.

Report on the findings

A substantive report based on the survey findings is published by the Wellcome Trust, available at: www.wellcome.ac.uk/monitor.

2. The samples

Overview

The survey comprises two samples: a representative sample of UK adults aged 18 or over living in private residential accommodation and a representative sample of young people in the UK aged 14 to 18 living in private residential accommodation. Sampling of both populations was undertaken at designated “core” addresses, while a hybrid of screening and focused enumeration (FE) was also employed to obtain additional interviews with young people.

The sample for the Wellcome Trust Monitor survey covered England, Wales, Scotland 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. If there was more than one eligible DU at the sampled address, one was randomly selected using a selection grid. At responding DUs interviewers used a selection grid to select one individual aged 18 years or over at random to complete the adult interview. At those DUs where an adult interview was achieved, and which also contained one or more individual aged between 14 and 18, the interviewer then randomly selected one young person to complete the young person interview¹. The “core” sample was designed to be representative of the general adult population aged 18 or over and the population of young people aged between 14 and 18 years respectively, living in private households in the UK.

Drawing the “core” sample

The Wellcome Trust Monitor uses a random probability sampling methodology. As is common in high-quality surveys of the general population, a multi-stage stratified sample was drawn to maximise precision while minimising cost.

The first stage of the sampling was to select the “clusters” (or Primary Sampling Units, PSUs) from which addresses for interviewers to visit would be sampled. Postcode sectors were chosen to be used as PSUs.

A list of all UK postcode sectors was drawn from the most up-to-date small-user Postcode Address File (PAF)², maintained by the Post Office. All sectors containing fewer than 1,000 delivery points were combined with adjacent sectors, so that each combined sector contained at least 1,000 delivery points.

Prior to selection the list of (combined) postcode sectors was stratified by Government Office Region, proportion of the population with qualifications at A level or better, and proportion of population living in owner-occupied dwellings based on 2001 census data. This was the same stratification scheme used in the first Wellcome Trust Monitor. Stratification can increase the precision of survey estimates if the variables used as stratifiers correlate with survey variables. Given the topics covered by the survey, educational achievement and tenure were considered appropriate choices.

¹ 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 in the sample of young people was addressed through the weighting strategy described in chapter 6. In summary we calculated a selection weight taking into account the adult selection, and calibrated the weighted sample to the known profile of young people.

² The version of the PAF used was Royal Mail postcode update from May 2012.

PSUs were selected with probability proportional to PAF delivery point count³ by applying the method of random start and fixed interval to cumulative PAF totals. Within each PSU, 25 addresses were randomly selected from the list of addresses in that PSU, sorted by postcode. Thus, each UK address had an equal probability of being sampled.

An initial sample of 2,625 “core” addresses was selected from 105 selected PSUs. During fieldwork a further nine PSUs and 225 addresses were selected using the same method⁴. Thus, overall, 2,850 “core” addresses were issued to interviewers.

At each “core” address interviewers, where necessary, randomly selected one dwelling unit, and approached those living there to take part. At each dwelling unit, interviewers attempted to identify and interview one adult aged 18 or over (where a household contained more than one adult, one was randomly selected). Where there were 18-year-olds in the dwelling unit, they were included in the adult selection. Any 18-year-olds not selected for an adult interview became eligible for the young person interview, along with all young people aged 14-17 living in the dwelling unit. Of the young people, one was randomly selected and an interview attempted with them.

Note that the survey covered individuals living in private residential accommodation only, and therefore those living in communal establishments, for example students living in university halls of residence, were not eligible. Some other surveys, for example the Labour Force Survey, include this type of student in their population definition and sample them via parents’ homes. However, in these cases student data are collected by proxy (through parents), an approach which can only work in a survey that collects relatively simple factual information. Although in principle it would be possible to sample students via parents’ addresses in this way and then to conduct follow up interviews (which could be via telephone), this was not considered a cost-effective or feasible option for the Wellcome Trust Monitor.

The young person “boost” sample

In the first Wellcome Trust Monitor, to obtain additional interviews with young people, a “boost” sample was also drawn. Focused enumeration was used to identify whether eligible young people were present at addresses near the “core” addresses.

Focused enumeration involves asking at each “core” address about the presence of the target population (in this case young people aged 14 to 18), at each of the focused enumeration addresses. Focused enumeration is used because it is less costly than full screening. This is because interviewers do not have to call at addresses which the householder at the core address confirm as containing no members of the target population. A shortcoming of focused enumeration, however, is that the target population is generally under-identified. This occurs where an interviewer undertakes proxy screening at a “core” address and is informed that no member of the target population lives at a given focused enumeration address, when actually one (or more) does. This under-identification can lead to biases in survey estimates. To overcome this problem, Ipsos MORI adopted a new methodological approach (face-to-face screening plus focused enumeration) which involved interviewers making two face-to-face visits to each boost address before reverting to screening by focused enumeration.

³ After expanding by MOI in Scotland but not elsewhere.

⁴ Additional addresses were selected to replace interviews where audio data had been lost as a result of problems experienced with the experimental audio capture methodology.

The purpose of this change was to enable interviewers to be more effective at identifying young people, reducing the risk of non-response, at the same time as retaining random probability sampling methods, the cost-efficiency of focused enumeration, and comparability with the first Wellcome Trust Monitor.

To calculate the number of “boost” addresses we would need to issue, we took into account assumptions about “deadwood” addresses, the proportion of addresses where we would be able to identify eligible young people, and response rates. We calculated we would need to issue 2.25 “boost” addresses for each core address. Hence at one in eight core addresses we selected alongside four “boost” addresses (these core addresses were designated “double boost” addresses) and at the remaining core addresses we selected two “boost” addresses (these core addresses were designated “single boost” addresses).

At each “boost” address, interviewers were required to make two calls to establish whether young people aged 14 to 18 lived in the household. In addition interviewers were instructed to ask at the “core” address or other neighbouring addresses whether young people aged 12 to 20 lived at each “boost” address. The wider age band was used to avoid false-negative responses by neighbours who do not know the precise age of the young people living at the selected addresses. Where interviewers established that young people were present, or where there was uncertainty, they attempted to make contact with the residents of the “boost” addresses. Where contact was made, interviewers randomly selected (in households containing more than one individual in the target age range) and attempted to interview one young person aged 14 to 18.

An initial sample of 5,906 “boost” addresses was selected. This equates to 2.25 “boost” addresses for each “core” address. During fieldwork a further 502 addresses were selected using the same method⁵. Thus, overall, 6,408 “boost” addresses were issued to interviewers.

The selection of the “boost” addresses was carried out in the following manner:

1. “Core” sample addresses were listed in the order they were selected.
2. A number between one and eight was randomly selected, and the corresponding address on the “core” sample list was designated a “double boost” “core” address. Thereafter every eighth address on the “core” sample list was also designated a “double boost” “core” address. All other “core” addresses remaining were designated “single boost” “core” addresses.
3. For “single boost” “core” addresses on PAF the address immediately preceding it on PAF and the address immediately following it on PAF were designated “boost” addresses.
4. For “double boost” “core” addresses on PAF the *two* addresses immediately preceding it on PAF and the *two* addresses immediately following it on PAF were designated “boost” addresses.

⁵ Additional addresses were selected to replace interviews where audio data had been lost as a result of problems experienced with the experimental audio capture methodology.

Development of materials

Scope of development work

The development stages of the survey were conducted over a three-month period from February 2012 to April 2012. While this was a tracking survey, many new questions were included to reflect topical issues in science and medical research. Furthermore, new features for the design of the survey were introduced in the second wave such as the split level of incentive payments (see chapter 4), the new face-to-face screening plus focused enumeration contact procedure (see chapter 2) and audio recording (see chapter 4). As such, the main requirements of this work were to test new questions, to identify any changes that needed to be made to existing questions, and to test the new methodologies.

The programme of development work was based around two pilots. The first involved a cognitive pilot of the new and a selection of the amended questionnaire material and the second consisted of a field pilot using the CAPI (Computer Assisted Personal Interviewing) program and the testing of many of the survey procedures to be used in the main stage. The questions for the pilot were primarily new questions designed specifically for this study, although the final questionnaire did include some repeat items, many of which had a number of minor amendments. Details of the repeat questions are included in Appendix D, while information on the cognitive testing is included in Appendix A.

Cognitive pilot

We undertook a thorough review of the questionnaire used in the first Wellcome Trust Monitor. New questions developed for the second Wellcome Trust Monitor were cognitively tested in an iterative process in February and March 2012. 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.

The cognitive pilot was split into four rounds, with findings discussed with the Wellcome Trust between each round, and the questionnaire being developed and re-tested from round-to-round in light of these discussions.

Respondents

The cognitive interviews were conducted by members of the research team, in Ipsos MORI's Borough office. Respondents were recruited by specialist recruiters within Ipsos MORI's field team from local areas. Attempts were made to recruit respondents from both low-income and high-income areas. It was anticipated that knowledge and understanding of medical research and science would vary by socio-economic status.

While strict quotas were not enforced for the cognitive pilot, recruiters were instructed to recruit respondents with a variety of characteristics, according to a recruitment specification created for each round of interviews. These were identified through the use of a screening questionnaire. In total, 35 interviews were conducted with adults and young people. These consisted of 17 adults and 18 young people, consisting of 19 women and 16 men. Adults ranged from 22 to 72 years of age, and young people ranged from 14 to 18 years of age.

In the first round of testing we carried out cognitive interviews with four adults on 23 February 2012 in London. In the second round cognitive interviews were carried out on 28 and 29 February with ten adults and four young people in London. In the third round cognitive interviews were carried out on 6 to 7 March with three adults and nine young people in London. The fourth and final round of cognitive interviewing took place on 8 March with five young people in London.

Cognitive pilot materials

Interviewers used the following materials to administer the cognitive interviews:

- cognitive pilot instructions
- cognitive testing questionnaire and prompts
- showcards
- Ipsos MORI's cognitive testing guidelines, detailing general prompts for the interviews
- copies of the advance leaflets for adults and young people, which were shown to respondents to test their impact and appropriateness

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.

Cognitive pilot modifications

Feedback was provided to the Wellcome Trust after each round of cognitive testing, 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 and re-tested throughout the cognitive pilot. These included the following:

- Questions around health-related findings in research proved too complex to deal with in the questionnaire. Some of the issues discussed required extensive deliberation, and respondents tended to lose sight of the issues in their totality, instead answering on the basis of general principles.
- The question which asked respondents about their support for different types of research was difficult for respondents to answer because they had different levels of knowledge as to what each type of research entails. In light of this, respondents simply listed those areas of which they possessed the greatest awareness.
- The section on vaccinations was reworked throughout the cognitive pilot as respondents found it difficult to think about the seriousness of illnesses in general terms (instead focusing on those groups who would be most vulnerable).

CAPI field pilot

A field pilot took place in April 2012. Its primary aim was to test important aspects of the Wellcome Trust Monitor, with a view to maximising the quality and effectiveness of the

main-stage survey, due to go into the field in May 2012. The field pilot sought to test a number of distinct aspects of the study – the practical administration of the questionnaire and survey in the field, the new method of audio-recording responses to certain questions, and the process of encouraging participation among potential respondents. In addition, this pilot helped confirm the content and length of the questionnaire for both adults and young people. The field pilot did not seek to test the contact or sampling procedures as these largely replicated those from the first wave of the Monitor.

Respondents

The CAPI field pilot used a quota sampling methodology and fifty respondents were interviewed between 3 and 15 April 2012. This meant that the survey could be tested amongst a broad spectrum of the general public. Four sample points (postcode sectors) were selected, an interviewer was assigned to each area, and a target was set of eight adults and five young people in each of the four areas. The four 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.

Briefing and debriefing

Interviewers attended a face-to-face briefing on 2 April 2012 where they were given background information about the purpose of the survey and were shown how to administer the questionnaire and given suggestions on how to encourage participation on the doorstep. Prior to the briefing, interviewers downloaded the questionnaire script on to their CAPI laptops. Each interviewer was also sent:

- 15 copies of an advance letter and explanatory leaflet, in envelopes;
- 10 copies of an explanatory letter for young people 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 field pilot project instructions;
- 15 copies of a field pilot feedback form;
- 1 quota sheet;
- 8 x £10 cash;
- 5 x £10 gift voucher;
- 1 paper copy of the survey questionnaire and
- 1 microphone.

All relevant documents are presented in Appendix A of this report.

The telephone debrief was held on 16 April 2012. Interviewers were asked to complete an electronic feedback form for each interview, recording all of their feedback, prior to the debrief. These formed the basis of the discussion at the debrief.

Outcomes

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 field pilot which were addressed, in consultation with the Wellcome Trust, prior to the main stage. These included the following:

- Providing the cash incentive or gift voucher on the spot was seen as a positive way of encouraging participation. Interviewers offered tips for the main stage of the survey to improve the administration and effectiveness of the incentive, such as instructing interviewers only to carry enough cash/vouchers for the number of interviews they expect to complete on any given day. In addition, it was felt that it would be helpful to have a leaflet to show young people where the gift vouchers could be used. This was introduced for the main stage of the survey.
- Interviewers expressed concerns about the audio recording process, as the CAPI script did not allow them to listen back to and check their recordings. In light of this, a test question was inserted into the CAPI script ahead of the main stage of interviewing. Interviewers were instructed to record respondents' answers to this question, replay those responses, and then make adjustments if necessary, before proceeding to the survey's audio questions. This was intended to ensure that the recording process was working correctly in all interviews.
- The interview length was found to be over the predicted duration for both adults and young people. Interviewers agreed that the average interview length was in the region of 50+ minutes for adults and 40+ minutes for young people. Aside from budgetary concerns, it was felt that younger respondents in particular struggled with the longer interview length. In light of this, some questions were removed for the survey, or omitted from this Wellcome Trust Monitor and ear-marked for inclusion in the third wave. A full report on the findings of the dress rehearsal pilot can be found in Appendix A.

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 "boost" addresses, were routed to the appropriate question sets. Dummy topline were run to ensure that the survey routing was correct and that respondents would be asked only questions appropriate to their knowledge and situation.

3. Data collection

Fieldwork was undertaken by Ipsos MORI interviewers with assistance from the Office for National Statistics' interviewers in the later stages. All fieldwork was conducted using face-to-face computer assisted interviewing.

Advance letter and leaflet

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. Letters were also issued for all "boost" addresses, which interviewers delivered by hand to addresses they called at.

This letter contained the contact details of Ipsos MORI's Operations Department as well as for the executive team, alongside links to the websites established for the young people and adult surveys.

In addition to the letters, leaflets were posted to all "core" addresses for both surveys (adults and young people), using the less formal and more attention-grabbing format to ensure that more people were aware of the survey and to help raise the response rate. As with the letter these were also given to interviewers to distribute to selected "boost" addresses.

Copies of the three different types of advance letter can be found in Appendix B.

Briefings

A total of five day-long briefings were held between 4 and 10 May 2012 in London, Bristol, Edinburgh and Leeds. A further two briefings were held later in the fieldwork period when new interviewers were added to the project. The briefings were conducted by researchers from Ipsos MORI.

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 "boost" addresses, the audio question recording process, an overview of the two versions of the questionnaire and strategies for gaining respondents' cooperation. The sessions also included a practice run-through of the questionnaire for adults and young people, including audio questions. Interviewers were given a copy of the project instructions.

Scheduling of interviews

Standard guidelines were issued to all interviewers about the timing and the number of calls they should make to an address in the sample. These stipulated that a minimum of six calls (three of which must be made at either a weekend or evening) must be made at each address over a three-week period before recording a non-contact or refusal. A maximum of nine calls was allowed, as it is envisaged that further effort beyond that point is unlikely to yield many more productive interviews.

Interviewers recorded details of every attempt to make contact with each address and, where selected, each respondent, on the relevant contact sheets; two different versions were available – one for “core” addresses and one for “boost” addresses. In addition, as part of the “core” contact sheet, interviewers completed a “boost” information sheet, indicating the outcome, if any, of their screening at the “core” address for the presence of young people at the associated “boost” addresses. In circumstances where it emerged that a particular “boost” address should 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 “boost” address, a “boost” contact sheet was used. Copies of the two contact sheets and the boost Information sheet can be found in Appendix B.

Consistent with the 2009 Monitor, young people were only interviewed at addresses where an adult interview was completed. At “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, and to prevent the young person from having to fill in the household grid section.

Fieldwork progress was monitored using Ipsos MORI’s computerised booking-in system.

Quality control

The time, date and outcome of all calls were recorded by interviewers and checks were made by field management.

Fieldwork progress

Interviewers updated their iProgress application with information from the paper contact sheets at the end of each interviewing day, and this information was transmitted back to Ipsos MORI’s Borough Road office over the internet. With this information, fieldwork progress could be updated on a daily basis.

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

Using this information, researchers were able to identify unproductive cases and points which could be reissued, as detailed in chapter 5.

Fieldwork lasted for a total of 22 weeks, from 21 May 2012 to 22 October 2012. This was longer than anticipated, owing in part to problems with the audio recording procedures and a lull during the Olympic and Jubilee period.

In light of the issue experienced with the quality of audio recording, fieldwork was halted for a few weeks, while the issue was identified and corrected.

Incentivisation

All respondents who completed the questionnaire were given an incentive immediately upon completion as a token of appreciation. For young people this incentive was a £10 LoveToShop voucher, which is accepted in many high street retailers. For adults the incentive was £10 in cash.

As part of an experiment into the effect of incentive payments on response rates, the cash/voucher incentive payment was raised to £15 in PSUs where response rates were predicted to be lower. A copy of the letter can be found in Appendix B.

To decide in which PSUs the incentive should be £10 and in which it should be £15, statistical modeling was undertaken using two years of field performance data from a large-scale random probability survey conducted in England and Wales in 2009/10. This survey used statistical wards as PSUs and these formed the units in the analysis. For each ward in the survey we included in the database: the first-issue co-operation rate, a wide range of Census 2001 variables, ACORN categorisation, and Index of Multiple Deprivation (IMD) quintiles (necessary because IMD is calculated differently in different countries of the UK). Wards with fewer than 15 interviews were excluded, reducing the initial sample of wards from 1,756 to 1,458. The co-operation rate was raised by the power of 1.33 before modelling in order to render its distribution closer to a normal one to create a “transformed co-operation rate”.

A linear regression model was estimated using “transformed co-operation rate” as the dependent variable and the remaining Census, ACORN, and IMD variables as potential independent ones. Only variables that were significantly correlated with “transformed co-operation rate” were considered for the model (p-value <0.05). A stepwise procedure was used with entry level of 0.1 and removal level of 0.2. Variables that were highly correlated with the last variable that entered in the model were excluded in order to reduce potential problems of multicollinearity. Once variables were selected, interaction and squared terms were also tested.

The correlation between the observed and predicted co-operation rates was 0.2827.

The equation developed in the process described above was applied to the selected PSUs for the Wellcome Trust Monitor. It should be noted that the Wellcome Trust Monitor PSUs were postcode-sector-based whereas the models were derived from a survey which used statistical wards as PSUs. Given that these two geographical units are of roughly equivalent size we do not think that this will have affected the analysis detrimentally.

We split the two incentive rates at the co-operation rate of 49.3 per cent (“transformed co-operation rate” of 58.7 per cent). PSUs at that level or below were allocated a £15 incentive, with the remaining being allocated a £10 incentive. The intention was to use a regression discontinuity design to test the effect of the increase in incentive on response propensity. However, initial analysis found that the correlation between the model predicted and the observed response rate was zero, so it was not possible to test whether the increase in incentive had any effect. The reason for the nil predictive power of the prediction equation is somewhat surprising. It could be because the prediction equation was based on a different survey topic than the Monitor, however, one should still expect some moderate positive correlation even taking this into account.

Audio recording

An innovation for the second Wellcome Trust Monitor was introduced for three questions which required respondents to say, in their own words, what came to mind when they were presented with a biomedical term. Instead of interviewers typing responses into the questionnaire program on their laptop, the answers were audio-recorded. The aim was to obtain fuller and richer data for these responses as well as minimising interviewer variability in typing of responses, to ensure that full and standardised responses were obtained. The questions were:

- What do you understand by the term DNA? (DNAMean)

This question was randomised using a split-ballot design so that for half the sample responses were recorded by the interviewer using a microphone and for half the

sample the responses were typed in to the interview program by the interviewer (as is standard practice for open-ended questions).

- What do you understand by the term human genome? (GenMean)
- What do you understand by the term GM, or genetic modification? (GMMean)

All respondents who were eligible for the relevant questions were asked to provide their consent to have their responses audio-recorded. Sixty-four per cent of eligible adults and sixty-eight per cent of eligible young people agreed to have their responses recorded.

During fieldwork, some issues were experienced with the process of recording and transmitting audio files. Consequently, a proportion of audio responses were lost before/during transmission to Ipsos MORI's main office. Ipsos MORI conducted a telephone survey to re-contact as many respondents as possible whose audio responses were lost.

Overall, the following responses were generated at the three selected audio questions:

Table 4.1 Audio question response by type	
<i>Base: All adults</i> <i>Wellcome Trust Monitor</i>	
Total number of responses by type	
Audio	2,168
<i>Of which lost</i>	407
Typed	1,896
Telephone audio (re-contact)	194

Interview length

The adult and young person interviews took an average of 50 minutes and 46 minutes respectively to complete.

4. 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 total achieved sample was 1,396 adults aged 18+ and 460 young people aged 14 to 18 (of whom 154 were interviewed as part of the “core” sample, and the remaining 306 as part of the “boost” sample). The response rate achieved among adults at the “core” addresses was 52.7 per cent⁶ (compared with 49.1 per cent in 2009). Among young people at the “core” addresses the response rate was 76.6 per cent⁷ (compared with 86.4 per cent in 2009). At the “boost” addresses, 67.3 per cent of young people identified agreed to take part (compared with 64.2 per cent in 2009).

Adult response rate

Table 5.1 shows a breakdown of the fieldwork outcomes for adults in our “core” sample. Because there are a small proportion of cases where we do not know if there was an eligible adult at the address, the “true” response rate falls within a range where all unknown eligibility cases (for example, address inaccessible, or unknown whether address is residential) are assumed to be eligible, to an upper limit where all these cases are assumed to be ineligible. The first Wellcome Trust Monitor found that the range of possible “true” response rates was only 1.2 percentage points. Thus, to permit easy comparison between the first and second Wellcome Trust Monitors, we have based our calculations on the (conservative) assumption that all unknown eligibility cases are eligible.

Table 5.1 Fieldwork outcomes for adult sample

	Wellcome Trust Monitor 1 (2009)		Wellcome Trust Monitor 2 (2012)	
	Number	%	Number	%
Addresses issued	2,650	100%	2,850	100%
Ineligible (out of scope)	251	9.5%	202	7.1%
Potentially eligible	2,399	90.5%	2,648	92.9%
<i>Of potentially eligible</i>				
Unknown eligibility	56	2.3%	104	3.9%
Definitely eligible	2,343	97.7%	2,544	96.1%
Interview achieved	1,179	49.1%	1,396	52.7%
Interview not achieved	1,164	50.9%	1,148	47.3%
<i>Of which:</i>				
<i>Non-contact</i>	97	4.0%	222	8.4%
<i>Refusal</i>	940	39.2%	811	30.6%
<i>Other unproductive</i>	127	5.3%	115	4.3%

In total, we achieved 1,396 productive interviews with adult respondents aged 18 years or over. The main reason for unproductive outcomes was refusal – 30.6 per cent of eligible addresses or addresses where eligibility was unknown were unproductive for this reason.

⁶ Response rates are calculated using response rate 1 defined by the American Association for Public Opinion Research Standard Definitions (2011) – see <http://bit.ly/ZWA0ST> for AAPOR’s Standard Definitions (2011)

⁷ NB – Young person response rate refers to core addresses at which a young person was confirmed as living.

Non-contacts accounted for 8.4 per cent of eligible addresses or addresses where eligibility was unknown, with a further 4.4 per cent covered by other unproductive outcomes, such as being away or ill during fieldwork.

Overall the adult response rate improved from 49.1 per cent to 52.7 per cent in 2012, mainly driven by the fall in the refusal rate from 39.2 per cent to 30.6 per cent. However, the non-contact rate worsened, rising from 4 per cent to 8.4 per cent in 2012.

Young person response rate

Tables 5.2 and 5.3 show a breakdown of the fieldwork outcomes for the young people (aged 14 to 18) in our sample. Separate figures are presented for “core” and “boost” addresses. We therefore focus on those “boost” addresses where young people were identified as being eligible to take part via our hybrid screening/focused enumeration procedures and calculate a “response rate” based on the proportion of these eligible young people who agreed to take part.

A total of 460 young person interviews were obtained, 154 at “core” addresses and 306 at “boost” addresses.

Table 5.2 shows that in 2012 a higher proportion of young people were identified at “core” addresses (7.1% compared with 5.3% in 2009). However, the co-operation rate fell from 86.4 per cent to 75.4 per cent, which was accounted for by an increase in the number of non-contact cases (from 0.7% to 4.4%) and other unproductive cases (from 0% to 8.9%)⁸.

Table 5.2 Fieldwork outcomes for young person “core” sample

	Wellcome Trust Monitor 1 (2009)		Wellcome Trust Monitor 2 (2012)	
	Number	%	Number	%
Addresses issued	2,650	100%	2,850	100%
Ineligible (out of scope)	251	9.5%	202	7.1%
Potentially eligible	2,399	90.5%	2,648	92.9%
<i>Of potentially eligible</i>				
14-to-18-year-old identified	140	5.3%	201	7.1%
Interview achieved	121	86.4%	154	75.4%
Interview not achieved	19	13.6%	47	24.6%
<i>Of which:</i>				
<i>Non-contact</i>	1	0.7%	9	4.4%
<i>Refusal</i>	18	12.9%	23	11.3%
<i>Other unproductive</i>	0	0.0%	15	8.9%

⁸ This outcome includes, for example, such cases as broken appointments where further contact is made but an interview cannot be rearranged.

Due to a change in the “boost” methodology in 2012 it is not possible to make direct comparisons with the “boost” methodology used in 2009. Hence below we present a comparison between our plan for the new methodology and the actual result. As can be seen, we were slightly less successful than we planned to be in identifying 14-to-18-year-olds, but managed a higher response rate than planned where we were able to identify a 14-to-18-year-old.

Table 5.3 Fieldwork outcomes for young person “boost” sample

	Planned		Actual	
	Number	%	Number	%
Addresses issued	6,408	100%	6,408	100%
Ineligible (out of scope)	5,799	90.5%	5,846	91.2%
Potentially eligible	609	9.5%	562	8.8%
Unknown eligibility	109		107	
14-to-18-year old identified	500	7.8%	455	7.1%
Interview achieved	325	65.0%	306	67.3%
Interview not achieved	175		149	32.7%
<i>Of which:</i>				
<i>Non-contact</i>			34	7.5%
<i>Refusal</i>			82	18.0%
<i>Other unproductive</i>			33	7.3%

Outcomes

The full set of outcomes for “core” addresses is provided in Table 5.4. We also provide further details of the unproductive outcome codes for those young people identified as being eligible for interview at either “core” or “boost” addresses (Table 5.5 and Table 5.6 respectively).

Table 5.4 Full fieldwork outcomes for adult “core” sample

	Wellcome Trust Monitor 1 (2009)	Wellcome Trust Monitor 2 (2012)	
	%	Number	%
Issued	100%	2,850	100%
Ineligible	9.5%	202	7.1%
Not yet built/under construction	0.2%	1	0.0%
Demolished/derelict	0.8%	7	0.2%
Vacant/empty housing unit	6.0%	124	4.4%
Non-residential address	1.2%	41	1.4%
Address occupied, no residents	0.8%	20	0.7%
Communal establishment no private dwellings	0.2%	5	0.2%
No eligible respondent 18+	0.2%	0	0.0%
Other ineligible	0.3%	4	0.1%
Potentially eligible	90.5%	2,648	92.9%
Unknown eligibility	2.3%	104	3.9%
Inaccessible	0.2%	7	0.3%
Unable to locate address	0.1%	5	0.2%
Unknown if address residential due to non-contact	0.1%	8	0.3%
Residential - don't know if eligible persons - no contact	0.3%	19	0.7%
Information refused about whether address residential	0.1%	1	0.0%
Contact but could not confirm resident household	0.2%	3	0.1%
Information refused about whether residents eligible	0.5%	43	1.6%
Unable to confirm eligibility - lack of knowledge	0.2%	7	0.3%
Unable to confirm eligibility - language problems	0.1%	3	0.1%
Other unknown eligibility	0.2%	8	0.3%
Issued but not attempted	0.0%	0	0.0%

Table 5.4 Full fieldwork outcomes for adult “core” sample (continued)

	Wellcome Trust Monitor 1 (2009)	Wellcome Trust Monitor 2 (2012)	
	%	Number	%
Definitely eligible sample	97.7%	2,544	96.1%
Non contact	4.0%	222	8.4%
No contact with anyone at address	1.5%	111	4.2%
Contact made but not with responsible adult	<0.5%	3	0.1%
Contact made but not with selected respondent	<0.5%	39	1.5%
Some contact with respondent, no interview	2.3%	69	2.6%
Refusal	39.2%	811	30.6%
Office refusal	1.1%	15	0.6%
Information refused about number of DUs	0.2%	22	0.8%
Information refused about number of eligible respondents	4.2%	93	3.5%
Refusal by selected respondent	27.6%	550	20.8%
Refusal by proxy	2.7%	85	3.2%
Refusal during interview	0.1%	0	0.0%
Broken appointment - no re-contact	3.3%	46	1.7%
Other non-productive	5.3%	115	4.4%
Ill at home during fieldwork period	1.3%	7	0.3%
Away in hospital throughout fieldwork period	1.2%	18	0.7%
Respondent physically/mentally incapable	1.7%	40	1.5%
Language barrier with selected respondent	0.5%	9	0.3%
Other non-response	0.6%	41	1.6%
Productive	49.1%	1,396	52.6%
Fully productive	49.0%	1,396	52.6%
Partially productive	0.1%	0	0.0%

Table 5.5 Full fieldwork outcomes for young people “core” sample

	Wellcome Trust Monitor 1 (2009)	Wellcome Trust Monitor 2 (2012)	
	%	Number	%
“Core” address where young person identified	5.3%	201	7.1%
Productive	86.4%	154	75.4%
Fully productive	86.4%	154	75.4%
Non contact	0.7%	9	4.4%
No contact with parent or guardian to obtain consent	0.7%	0	0.0%
No contact with young person	0.0%	9	4.4%
Refusal	12.9%	23	11.3%
Refusal by selected respondent	5.7%	7	3.4%
Refusal by parent or guardian	5.7%	14	6.9%
Refusal during interview	0.0%	0	0.0%
Broken appointment - no re-contact	1.4%	2	1.0%
Other non-productive	0.0%	15	8.9%
Ill at home during fieldwork period	0.0%	0	0.0%
Away in hospital throughout fieldwork period	0.0%	1	1.5%
Respondent physically/mentally incapable	0.0%	0	0.0%
Language barrier with selected respondent	0.0%	0	0.0%
Other non-response	0.0%	14	7.4%

Table 5.6 Full fieldwork outcomes for young people “boost” sample

	Wellcome Trust Monitor 2 (2012)	
	Number	%
Issued	6,408	100%
Ineligible	5,846	91.2%
Not yet built/under construction	6	0.1%
Demolished/derelict	22	0.3%
Vacant/empty housing unit	159	2.5%
Non-residential address	73	1.1%
Address occupied, no residents	17	0.3%
Communal establishment no private dwellings	10	0.2%
No eligible respondent aged 14-18	5,547	86.6%
Other ineligible	12	0.2%
Unknown eligibility	107	1.7%
Inaccessible	14	0.2%
Unable to locate address	25	0.4%
Unknown if address residential due to non-contact	5	0.1%
Residential - don't know if eligible persons - no contact	16	0.2%
Information refused about whether address residential	2	0.0%
Contact but could not confirm resident household	0	0.0%
Information refused about whether residents eligible	4	0.1%
Unable to confirm eligibility - lack of knowledge	0	0.0%
Unable to confirm eligibility - language problems	2	0.0%
Other unknown eligibility	0	0.0%
Issued but not attempted	0	0.0%
No contact with anyone at address	33	7.3%
Contact made but not with responsible adult	0	0.0%
Office refusal	5	1.1%
Information refused about number of DUs	1	0.2%

Table 5.6 Full fieldwork outcomes for young people “boost” sample (continued)

	Wellcome Trust Monitor 1 (2009)	Wellcome Trust Monitor 2 (2012)	
	%	Number	%
Definitely eligible sample	5.0%	455	7.1%
Non contact	1.3%	34	7.5%
Contact made but not with selected respondent	1.3%	22	4.8%
Some contact with respondent, no interview	-	10	2.2%
No parental permission because of no contact with parent	-	2	0.4%
Refusal	32.5%	82	18.0%
Information refused about number of eligible respondents	0.5%	5	1.1%
Refusal by selected respondent	13.2%	32	7.0%
Refusal by proxy	15.0%	37	8.1%
Refusal during interview	<0.5%	0	0.0%
Broken appointment - no re-contact	3.6%	8	1.8%
Other non-productive	2.0%	33	10.1%
Ill at home during fieldwork period	0.0%	0	0.0%
Away in hospital throughout fieldwork period	0.8%	2	0.4%
Respondent physically/mentally incapable	0.8%	7	1.5%
Language barrier with selected respondent	-	0	0.0%
Other non-response	0.5%	24	8.1%
Productive	64.2%	306	64.4%

Further information on response

The response rate among adults aged 18 or over was lower than anticipated (52.7% against our target of 55%). However, it was an improvement over the response rate for the first Wellcome Trust Monitor, which was 49.1 per cent.

The main issue encountered during fieldwork was problems with coverage reflecting the fact this was an extremely busy period for Ipsos MORI. Because of competing demands from existing fieldwork commitments, the field period lasted for 22 weeks compared with 13 weeks in the first Wellcome Trust Monitor. Despite this, the non-contact rate rose to 8.4 per cent from 4.0 per cent on the adult sample, although the refusal rate fell from 39.2 per cent to 30.6 per cent.

Efforts made to maximise 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.

In the first Wellcome Trust Monitor 781 “core” addresses were reissued once (29.5% of all “core” addresses). We reissued 1,269 “core” addresses (44.5% of all addresses). In addition a further 32 cases were reissued for a second time. Part of the reason so many cases were reissued was that a number of sampling points were worked by interviewers from the Office for National Statistics part of the way through fieldwork (having previously been issued to Ipsos MORI interviewers). To ensure they had assignments that were sufficiently large, sampling points were transferred to them as a whole, with the exception of addresses where an interview had already been achieved.

In the first Wellcome Trust Monitor no “boost” addresses were reissued. We reissued 331 “boost” addresses (5.2% of all “boost” addresses) and four “boost” addresses were reissued for a second time.

5. Weighting

Overview

The survey dataset has been weighted to ensure that it is representative of the two survey populations – adults aged 18 or over and young people aged 14 to 18 – according to certain known population distributions.

Unlike the first Wellcome Trust Monitor, three stages of weighting were applied. The data were weighted to account for differing probabilities of selection (to take into account differing numbers of dwelling units at a small number of addresses, and household composition). The next (new) step was (for the adult sample only) to create a non-response weight to adjust for the propensities of people in different areas to respond (using logistic regression modelling). The final step was to apply calibration weighting, meaning both samples were proportionally matched to the UK population with regard to age within gender, and to region.

The dataset contains one weight variable, which should be applied for all analysis. Analysis should always be conducted separately for adults and young people; as adults and young people are drawn from separate samples, the weight calculation is different depending on whether the respondent is in the adult or young person sample.

Adult weight

The weight for adult respondents:

- adjusts for differential selection probabilities resulting from the selection of one dwelling unit per address and one adult per dwelling unit;
- adjusts for a modelled propensity to respond in geo-demographically defined areas; and
- adjusts for differential non-response by region and, separately, by age and gender, thereby making the achieved sample representative of the population by these variables. The weights were created in a series of steps detailed below.

Dwelling unit selection weight

One dwelling unit was selected at each address and where there was more than one dwelling unit at an address the participating dwelling unit had a lower chance of selection than addresses where there was only one dwelling unit. To correct for unequal probabilities of selection, a dwelling unit selection weight was created. This was equal to the number of dwelling units found at the address. The weight was trimmed at four to avoid a small number of very high weights as these would inflate the standard errors and reduce the precision of the survey estimates.

Adult selection weight

One adult aged 18 or over was interviewed at each participating dwelling unit. Therefore adults living with others had a lower chance of selection than those living alone. To correct for this, an adult selection weight was created. This was equal to the number of adults in the dwelling unit. The weight was trimmed at four.

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.

Non-response modelling

A standard way of correcting for non-response is to model the probability of response in geo-demographically defined areas. We used a logistic regression model to estimate the probability that a selected address will yield a productive interview. The explanatory variables in the model were Government Office Region, ACORN category, the proportion of adults in the PSU with a high qualification and an Urban/Rural indicator.

When comparing the weights produced using non-response modelling and the weights produced using the same method for the first Wellcome Trust Monitor, the two weights had a correlation of 97.8 per cent, which indicates that the additional information from the nonresponse modelling has little effect on point estimates.

The non-response weight and the selection weight were multiplied together to obtain a “pre-calibration” weight.

Calibration to the population

The next step was to take the weighted sample and to “calibrate” the totals in each Government Office Region (GOR), and each of twelve age/gender categories, to population totals derived from the latest (mid-2010) 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 gender.

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.

Table 6.1 Regional profile of UK adults aged 18 or over

Region	Number of adults aged 18 or over	% of adult population
North East	2,080,000	4.2%
North West	5,458,000	11.1%
Yorkshire And The Humber	4,189,000	8.5%
East Midlands	3,548,000	7.2%
West Midlands	4,260,000	8.7%
East	4,597,000	9.4%
London	6,134,000	12.5%
South East	6,703,000	13.6%
South West	4,219,000	8.6%
Wales	2,380,000	4.8%
Scotland	4,183,000	8.5%
Northern Ireland	1,370,000	2.8%
United Kingdom	49,122,000	100%

Note: numbers are rounded to the nearest 1,000.

Table 6.2 Age and gender profile of UK adults aged 18 or over

	Men		Women	
Age group	Number of men aged 18 or over	% of adult population	Number of women aged 18 or over	% of adult population
18-29	5,228,400	10.6%	4,974,900	10.1%
30-39	4,044,300	8.2%	4,048,900	8.2%
40-49	4,543,300	9.2%	4,654,700	9.5%
50-59	3,723,200	7.6%	3,836,200	7.8%
60-69	3,252,100	6.6%	3,443,100	7.0%
70+	3,121,300	6.4%	4,251,600	8.7%
All 18+	23,912,600	48.7%	25,209,400	51.3%

Note: numbers are rounded to the nearest 100.

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 (1,396); this results in weights with an average of 1⁹.

Young person weight

⁹ Individual weights were multiplied by the unweighted base size divided by the sum of weights.

The weight for young people respondents:

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

The weights were created in a series of steps detailed below.

Dwelling unit selection weight

One dwelling unit was selected at each address and where there was more than one dwelling unit at an address the participating dwelling unit had a lower chance of selection than addresses where there was only one dwelling unit. To correct for unequal probabilities of selection, a dwelling unit selection weight was created. This was equal to the number of dwelling units found at the address. As the maximum number of dwelling units was two there was no trimming of weights.

Young person selection weight

At both “core” and “boost” dwelling units, one young person aged 14 to 18 was interviewed. Those young people living with other 14-to-18-year-olds therefore had a lower chance of selection than those living in dwelling units without any other 14-to-18-year-olds. 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 to 18 was picked *after* the selection of one adult aged 18 or over. The weights for young people had to take into account whether the address was a “core” or “boost” address, and the household composition.

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/gender categories, to population totals derived from the latest (mid-2010) 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 gender.

We collapsed regions into the following groups following the convention used in the first Wellcome Trust Monitor:

- 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.

Table 6.3 Regional profile of UK young people aged 14 to 18

Region	Number of young people aged 14 to 18	% of young people population
North	936,000	24.6%
Midlands	628,000	16.5%
South	1,202,000	31.6%
London	412,000	10.8%
Wales	192,000	5.1%
Scotland	313,000	8.2%
Northern Ireland	122,000	3.2%
United Kingdom	3,805,000	100%

Note: numbers are rounded to the nearest 1,000.

Table 6.4 Age and gender profile of UK young people aged 14 to 18

	Young men		Young women	
Age	Number of young men aged 14 to 18	% of young people population	Number of young women aged 14 to 18	% of young people population
14	374,000	9.8%	356,200	9.4%
15	378,500	9.9%	357,900	9.4%
16	389,400	10.2%	368,400	9.7%
17	398,300	10.5%	375,600	9.9%
18	415,600	10.9%	392,100	10.3%
All 14-18	1,955,800	51.4%	1,850,200	48.6%

Note: numbers are rounded to the nearest 100.

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 (460); this results in weights with a mean of 1.

6. 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 114 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 Factor	Unweighted base	95% confidence interval	
						Upper	Lower
Interest in medical research	% very interested	20	1.18	1.10	1,396	22.6	18.0
Understanding of the term DNA	% at least some understanding of the term	85	1.27	1.32	1,396	87.1	82.2
Whether has tried to find out any information on medical research in the past year	% have tried to find information	35	1.70	1.33	1,396	38.6	31.9
Whether medical research will lead to an improvement in quality of life in the UK over the next 20 years	% definitely will	51	2.12	1.59	1,396	55.0	46.7
Optimism about medical advances as a result of genetics research	% at least somewhat optimistic	80	1.14	1.07	1,396	82.7	78.2
Level of trust in medical research charities to provide accurate information about medical research	% complete trust	14	1.31	1.43	1,396	16.3	11.1
Willingness to take part in a medical research project testing a new drug or treatment	% at least fairly willing	60	1.91	1.46	1,396	63.9	56.4

Table 7.2 Complex standard errors and confidence intervals of selected variables for young people

Variable	Estimate description	Estimate	Standard error	Design Factor	Unweighted base	95% confidence interval	
						Upper	Lower
Interest in medical research	% very interested	8	1.34	1.05	460	10.9	5.6
Understanding of the term DNA	% at least some understanding of the term	90	1.36	0.98	460	92.9	87.5
Whether has tried to find out any information on medical research in the past year	% have tried to find information	37	3.06	1.36	460	43.2	31.2
Whether medical research will lead to an improvement in quality of life in the UK over the next 20 years	% definitely will	44	3.32	1.43	460	50.5	37.5
Optimism about medical advances as a result of genetics research	% at least somewhat optimistic	79	2.53	1.34	460	84.1	74.2
Level of interest in science lessons at school	% at least fairly interested	82	1.98	1.11	460	86.2	78.4
Whether science is a good area of employment for young people to go into	% yes	82	2.10	1.18	460	86.6	78.4

7. Data processing and management

Editing

A number of checks were included in the CAPI programme (Table 8.1) 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).

Table 8.1 Checks included in the CAPI programme

Question	Check
RadNews, TVNews, TVHrs, WWWHrs	Query respondents who have previously said that they get information from each media saying they never use these media in follow up questions
R, AgeIf	Query any unusual age/relationship combinations. Confirm any impossible age/relationship combinations.
AgeIf	Check to confirm ages greater than 97
DtJbLv	Confirm date left last job if date entered is before 1900. Query if date provided is before 1980. Query if date provided is in the future. Query if date provided is before date provided at StartDat

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.

Coding

Post-interview coding was undertaken by members of Ipsos MORI's coding department using our coding software, Ascribe. Coders were briefed by researchers and provided with full instructions (Appendix C). In total, nine coders worked on the coding of the open-ended questions.

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-ended questions

The adult interview contained eleven open-ended questions and the young person interview contained ten open-ended questions. These open-ended questions were mainly designed to measure respondents' awareness and knowledge of 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-ended

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.

Where questions were consistent between waves, code frames were carried over from the first Wellcome Trust Monitor. Five new open-ended questions were inserted into the questionnaire for the second Wellcome Trust Monitor. Researchers developed code frames for these new questions based on the data received from the first 500 interviews. As far as possible, these new code frames were created to closely reflect the style of pre-existing codes.

The first batch of 50 interviews coded by each interviewer was fully checked by the Operations team, and all coding was quality checked at the end of the project with responses re-assigned where necessary.

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).

Audio recorded questions

As outlined in chapter 4, the Wellcome Trust is interested in seeing whether recording respondents’ answers to three questions of particular interest (DNAMean, GenMean, GMMean) would generate higher-quality data. Responses to these questions were first transcribed and then coded using the same code frame for standard typed responses. The same team was used to code audio and typed responses. At DNAMean, the sample was split so that half of respondents had their answers recorded using a microphone, while the other half had their responses typed in as usual. This allowed the responses provided by the two groups to be compared to see whether those providing audio responses gave fuller or more considered answers.

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.

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 (1,396) and all 14-to-18-year-old respondents (460). Analysis should always be conducted separately for adults and young people; as the dataset contains two samples from 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.

Appendices

Appendix A – Cognitive testing and pilot survey documents

Appendix B – Main stage fieldwork documents

Appendix C – Editing and coding documents

Appendix D – Questionnaire and showcards

Wellcome Trust

We are a global charitable foundation dedicated to achieving extraordinary improvements in human and animal health. We support the brightest minds in biomedical research and the medical humanities. Our breadth of support includes public engagement, education and the application of research to improve health.

We are independent of both political and commercial interests.

Wellcome Trust
Gibbs Building
215 Euston Road
London NW1 2BE, UK
T +44 (0)20 7611 7221
F +44 (0)20 7611 8254
E education@wellcome.ac.uk
www.wellcome.ac.uk

The Wellcome Trust is a charity registered in England and Wales, no. 210183. Its sole trustee is The Wellcome Trust Limited, a company registered in England and Wales, no. 2711000 (whose registered office is at 215 Euston Road, London NW1 2BE, UK). PS-5444.3/05-2013/MC