National Survey of Sexual Attitudes and Lifestyles 3

Technical Report

Volume 1: Methodology

Bob Erens,[1,2] Andrew Phelps,[3] Soazig Clifton,[1,3] David Hussey,[3] Catherine H Mercer,[1] Clare Tanton,[1] Pam Sonnenberg,[1] Wendy Macdowall,[2] Andrew J Copas,[1] Nigel Field,[1] Kirstin Mitchell,[2] Jessica Datta,[2] Victoria Hawkins,[3] Catherine Ison,[4] Simon Beddows,[4] Kate Soldan,[4] Filomeno Coelho da Silva,[4] Sarah Alexander,[4] Kaye Wellings,[2] Anne M Johnson[1]

1. University College London
2. London School of Hygiene & Tropical Medicine
3. National Centre for Social Research
4. Public Health England

P2930
November 2013

Contents

Note: This version (updated 31/10/2014) includes a few corrections to the equations shown on page 39.

Acknowledgements

The third National Survey of Sexual Attitudes and Lifestyles was undertaken by a large team of researchers based at five different UK organisations: University College London (UCL), the London School of Hygiene & Tropical Medicine (LSHTM), NatCen Social Research (NatCen), Public Health England (formerly the Health Protection Agency), and the University of Manchester.

Our main debt of course is to the 15,162 people who gave up their time to take part in the survey and provide information about their experiences and attitudes, and to the 491 NatCen interviewers who administered the survey.

We would also like to thank NatCen's Operations team, particularly Emma Fenn, Pauline Burge, Pauline Calvert, Sharon Evans and Rita Vaughan, the computing staff Iain Templeton, Sven Sjodin, and Colin Miceli, Michelle Gray from NatCen's Questionnaire Development and Testing Hub, and the NatCen Data Manager Cathy Coshall.

Thanks also to other Natsal-3 team members: Philip Prah, Kyle Jones, Sarah Burkill and Sandy Gale at UCL, Ruth Lewis, Lorna Gibson and Rachael Parker at LSHTM, Frederick Wu at University of Manchester, Tosin Omole and Heather Wardle at NatCen.

We thank the following staff at Public Health England: Sexually Transmitted Bacteria Reference Laboratory (STBRL): Pamela Saunders who oversaw testing for chlamydia, gonorrhoea, and *Mycoplasma genitalium*, and Rumena Khanom who assisted with carrying out the testing. Virus Reference Department: Natasha de Silva, for initial work on the HPV testing protocol, validations and pilot work; Kavita Panwar, Precious Manyenga, Mohammed-Abbas Fazal for HPV testing; Gary Murphy, section head of the HIV diagnostic laboratory, Elaine McKinney who was responsible for HIV testing validation, Bharati Patel for conducting HIV testing and John Parry, for initial discussions on HIV testing strategy. HIV & STI Department: Chinelo Obi, Rebecca Howell-Jones and David Mesher for management of urine specimen data-linkage and anonymisation procedures; Heather Northend, Krishna Gupta and Tracey Cairns for data-entry.

We thank the following staff for their involvement in the testosterone work: Halina MacIntyre, Denis O'Reily and Mike Wallace at Glasgow Royal Infirmary, Brian Keevil at University Hospital South Manchester, Phil MacDonald and Carly Mosely at the University of Manchester, David Baird and John Bancroft.

The research team was advised throughout by an expert advisory group chaired by Anna Glasier (University of Edinburgh), and which included: Nick Partridge (Terrence Higgins Trust), Jackie Cassell (Brighton and Sussex Medical School), Alison Hadley (Director – Teenage Pregnancy Knowledge Unit), Julie Bentley and Audrey Simpson (Family Planning Association), Tony Nardone and Kevin Fenton (Public Health England), Graham Hart (UCL), John Cleland (LSHTM), Nathalie Bajos (Inserm), George Kinghorn (Royal Hallamshire Hospital), and a number of representatives from the Medical Research Council, Wellcome Trust, Economic and Social Research Council and Department of Health.

The survey was jointly funded by grants from the Medical Research Council (G0701757) and the Wellcome Trust (084840), with additional funding provided by the Economic & Social Research Council (ESRC) and the Department of Health. Nigel Field is supported by a NIHR Academic Clinical Lectureship. The study received ethical approval from the Oxford Research Ethics Committee A (reference number 10/H0604/27).

1 Introduction

The third National Survey of Sexual Attitudes & Lifestyles (Natsal-3) was carried out between 2010 and 2012 by a collaborative team from five organisations:
University College London (UCL): Anne Johnson, Bob Erens, Cath Mercer, Pam Sonnenberg, Clare Tanton, Nigel Field, Andrew Copas, Phillip Prah, Sarah Burkill, Kyle Jones
London School of Hygiene & Tropical Medicine (LSHTM): Kaye Wellings, Wendy Macdowall, Jessica Datta, Kirstin Mitchell, Ruth Lewis, Lorna Gibson
National Centre for Social Research (NatCen): Andrew Phelps, Soazig Clifton, Vicki Hawkins, Tosin Omole, Heather Wardle
Public Health England (PHE) (formerly the Health Protection Agency): Cathy Ison, Simon Beddows, Kate Soldan
University of Manchester: Fred Wu

Natsal-3 was closely modelled on the two previous surveys on sexual behaviour in Britain – Natsal-1 in 1990-1991, and Natsal-2 in 1999-2001 – and was carried out by the same core team of investigators. Natsal-1 involved interviews with 18,876 adults aged 16-59 using a combination of face-to-face interview and paper-and-pencil self-completion questionnaire. Natsal-2 interviewed 11,161 adults aged 16-44, along with a further 949 adults from boosted ethnic minority groups; data collection was carried out using computer assisted personal interview (CAPI) techniques along with computer assisted self-interview (CASI) for the more sensitive questions. Details of the methodologies used for Natsal-1 and Natsal-2 have been reported in a number of books, reports and academic papers.[1,2,3,4,5,6,7] A full list of publications can be found on the Natsal website (www.natsal.ac.uk).

Data from Natsal-1 and Natsal-2 have informed sexual and reproductive health policy in Britain.[8,9,10,11] Natsal has been extensively cited in Department of Health (DH) strategic documents; in evidence to the House of Commons Select Committee on Sexual Health; in plans for delivery of the National Chlamydia Screen Programme (NCSP); and for informing a sexual health social marketing intervention. Data from Natsal have been used in the parameterisation of mathematical models of HIV and STI transmission.[12,13]

Natsal-3 aimed to provide up-to-date information on key sexual behaviours and risk factors, as well as to look at trends over time by including comparable measures to those used in the previous two surveys, and to include new questions and an extended age range to take account of current information needs in the field of sexual health. The main objectives of Natsal-3 were to:

- provide a detailed understanding of patterns and variability of sexual behaviour in Britain (including, for example, numbers of sexual partners, frequency of different sexual practices, and homosexual experience)
- provide self-reported estimates of a range of sexual and reproductive health outcomes (including, for example, pregnancy, STI diagnosis, contraception use) and health service use
- explore sexual behaviour and function over the life-course by including an older age group (up to age 74)
- describe changes in sexual activity over time and trends in relationships, reproductive history and patterns of fertility using Natsal-1, Natsal-2 and Natsal-3
- from urine samples, measure the prevalence of *Chlamydia trachomatis, Neisseria gonorrhoeae*, type-specific Human Papillomavirus, HIV antibody and *Mycoplasma genitalium*
- measure the gender-specific distribution of salivary testosterone and its relationship to sexual behaviour, function and ageing.

To enable comparisons over time, Natsal-3 was largely based on methods developed for Natsal-1 and Natsal-2. While Natsal-1 used paper-and-pencil (PAPI) interview and self-completion questionnaires, Natsal-2 moved to the use of computer assisted personal interview (CAPI) and computer assisted self-interview (CASI) techniques, which were also used on Natsal-3. A high proportion of the questions were also identical in all three Natsal surveys, but there were also some methodological differences introduced to Natsal-3:

- Natsal-3 covered the widest age range (ages 16-74 years), compared with Natsal-1 (ages 16-59 years) and Natsal-2 (ages 16-44 years)
- new questions were introduced, including modules on sexual function and satisfaction, health conditions and problems that may affect sexual lifestyles, non-volitional sex, unplanned pregnancy
- the urine specimen measured not only *C.trachomatis* and type-specific HPV, but also *N.gonorrhoeae*, HIV antibody and *M.genitalium*
- a saliva sample was collected to test for salivary testosterone.

Natsal-3 was designed to carry out (approximately) hour-long interviews on average, with a probability sample of 15,000 men and women aged 16-74 years resident in Britain. Young people aged 16-34 were over-sampled in order to provide sufficient statistical power to examine behaviours among the age-group at highest risk for a range of sexual health outcomes, including teenage pregnancy and STIs.

Considerable development work was carried out for Natsal-3, which focused on:

- developing the new questions to be included in the survey
- testing the feasibility and acceptability of the survey in general, and among older respondents (i.e. up to age 74) in particular
- developing an ethical framework for anonymous testing of urine samples without providing results to respondents, and testing the feasibility and acceptability of this approach
- testing the feasibility and acceptability of collecting saliva samples using self-collection methods
- validating an assay for measuring testosterone concentrations in saliva for use in epidemiological studies
- developing psychometrically validated measures for unplanned pregnancy and sexual function.[14,15]

This report describes the methods used for Natsal-3 and deals with sampling, data collection, response, data processing and weighting. More detailed methodological work on the validated measures, anonymised testing of urine samples, the development of salivary testosterone, and the extensive cognitive testing used to develop the new questions for Natsal-3 are reported elsewhere.[14,15,16,17,18] This report does not provide any survey results, which will initially be reported in a series of journal articles in *The Lancet* in November 2013. It also does not describe the in-depth qualitative exploration of particular groups of individuals which involved following-up a purposive sample of men and women from the main Natsal-3 survey.

2 Sampling

2.1 Sample size and structure

Natsal-3 aimed to interview a representative sample of men and women aged 16-74 living in private households in Great Britain. The target number of interviews was 15,000: 10,000 with a 'core' sample of adults aged 16-74; and a further 5,000 with a 'boost' sample of younger adults aged 16-34, so that this age group constitutes approximately half of the achieved sample. The reason for boosting the 16-34 age group was to have sufficient statistical power to permit detailed exploration of behaviours among the age-group at highest risk of a range of sexual health outcomes, including teenage pregnancy and STIs. In order to achieve approximately equal numbers (of around 1900) in five year age bands up to age 35, screening was split into two separate boost samples: boost 1 screened for ages 16-34, and boost 2 screened for ages 16-29.

The sample used a multi-stage, clustered and stratified probability design, with postcode sectors selected as the primary sampling units (PSUs), addresses within them selected at the second stage, and finally one eligible adult was randomly selected at the final stage. The sampling frame was the 'small-user' Postcode Address File (PAF), a list of all addresses (delivery points) in the country. The original sample design involved selecting 56,880 addresses from the PAF, 22,752 for the core sample and 34,128 for the boost sample. (For practical reasons, addresses north of the Caledonian Canal, the Scottish islands, and the Isles of Scilly were excluded from the sample frame. This excludes 0.5% of eligible addresses in Great Britain.)

The sample design is described more fully below.

2.2 Selection of primary sampling units (PSUs)

A list of all postcode sectors (south of the Caledonian Canal) in Great Britain was generated. Postcode sectors with fewer than 1000 PAF 'delivery points' (addresses) were combined with neighbouring sectors so as to avoid any tight clustering of sampled addresses. A total of 1727 postcode sectors were selected as PSUs.

Before selection, postcode sectors were stratified in order to maximise precision of the sample and to ensure that different strata in the population were correctly represented. They were first sorted into 11 Government Office Regions, with a further distinction between inner and outer London. Within each of these 12 regions, they were then listed in increasing order of population density to create three roughly equal sized bands. Next, within each of these 36 region/density bands, they were listed in increasing order of the proportion of the population aged under 60, and again divided into three roughly equal sized bands, giving a total of 108 region/density/age bands. Finally, within these bands, they were listed in increasing order of the proportion of households with a head of household in a non-manual occupation (Socio-Economic Groups 1-6, 13). The data used to create these strata were taken from the 2001 Census of the Population. The sectors were selected systematically, with each postcode sector being given a probability of selection proportional to its total number of delivery points.

Fieldwork was split into eight 'waves', and sectors were selected two waves at a time to allow for adjustments to the number of sectors selected as fieldwork progressed. A new wave was issued roughly every quarter over the two years of data collection (see section 3.3.1). Initially, 225 sectors were selected for wave 1 and another 225 for wave 2. An additional 10 sectors were randomly selected as reserve points for each wave. At wave 1 the reserve points were not used, but the 10 reserve points were issued at wave 2.

For the remaining waves (3 to 8), 217 sectors were selected for each wave, with the exception of wave 8 where the number of issued sectors was scaled back to 182 so that the overall target of 15,000 achieved interviews would not be significantly exceeded (with the excluded sectors chosen at random). Since the allocation of sectors to waves was done randomly, each wave provides a nationally representative sample of the general population.

2.3 Sampling delivery points (addresses) and households

Within each selected PSU, the original design was to systematically select 30 delivery points (addresses). Within each PSU, 12 addresses would be randomly assigned to the core (i.e. screened for ages 16-74), 10 would be boost 1 addresses (screened for ages 16-34), and 8 would be boost 2 addresses (screened for ages 16-29).

In practice, however, the sample design underwent small modifications during the two year period of data collection in order to 'fine tune' the sample to meet the target number of achieved interviews. Thus, because of the higher number of addresses screened out than originally estimated, the number of selected addresses was increased to 36 for waves 3 to 8: of these, 15 were assigned to the core, 11 to boost 1, and 10 to boost 2. Another modification was made in wave 7, as the number of respondents in the younger age groups was exceeding our target, while the number in the older age groups was below the target. In order to rectify this, the allocation of the 36 addresses was altered by increasing the number of core addresses to 16 and decreasing boost 2 addresses to 9 (and leaving boost 1 addresses at 11).

A summary of the number of PSUs and addresses issued per wave is found in Table 2.1.

Table 2.1 Number of PSUs and addresses issued per wave of fieldwork

		Addresses/PSU			
	PSUs issued	Total	Core	Boost 1 (16-34)	Boost 2 (16-29)
Wave 1	225	30	12	10	8
Wave 2	235	30	12	10	8
Wave 3	217	36	15	11	10
Wave 4	217	36	15	11	10
Wave 5	217	36	15	11	10
Wave 6	217	36	15	11	10
Wave 7	217	36	16	11	9
Wave 8	182	36	16	11	9
TOTAL PSUs/ addresses	**1727**	**59412**	**24924**	**18537**	**15951**

When visited by interviewers, 6.8% of the selected addresses were found to contain no private households. Examples include businesses, offices and institutions, vacant properties, demolished or derelict buildings and those still in the process of being built. These addresses were counted as ineligible and were excluded from the survey sample.

A small proportion of addresses on the PAF contain more than one dwelling unit, such as a house which may have been split into two or more flats. (A dwelling unit is a living space with its own front door.) Interviewers may also find dwelling units which contain several resident households, for example, two families living as two separate households within a single house. (A household is defined as one person, or a group of people living in a dwelling unit who either share living accommodation or at least one meal a day.) A relatively small number of such addresses (around 1%) in England and Wales are not reliably identified on the PAF. However, in

Scotland, where this is a more common occurrence, the multiple occupancy indicator (MOI) on the PAF identifies the number of dwelling units at each address. Thus, in Scotland only, before the selection of addresses, PAF was expanded by the MOI to give the extra dwelling units the same chance of being selected as those at single occupancy addresses.

In cases where a selected address contained more than one dwelling unit or household, the interviewer followed a special procedure using random selection digits provided (a Kish grid technique). This procedure resulted in a random selection of one household from among all dwelling units/ households at an address. The selected household was included in the survey sample, and the others were omitted. The final sample was weighted to remove any bias that may arise from the lower chance of selection among dwelling units and/or households at multi-household addresses (see section 7.1.1).

2.4 Sampling individuals within addresses

Since one aim of the sample design was to boost young people aged 16-34 (as described in section 2.1), addresses within each PSU were randomly allocated to one of three sample types:

- the core sample, at which one person aged 16-74 years was selected
- boost 1, at which one person aged 16-34 years was selected
- boost 2, at which one person aged 16-29 years was selected.

At each private residential address, interviewers listed all adults within the relevant age range. Addresses without any residents in the correct age range were ineligible for the survey: 14.5% of core addresses were excluded for this reason, as were 67.1% of boost 1, and 74.2% of boost 2, addresses. At addresses where there was more than one person in the eligible age range, one was randomly selected using a Kish grid technique. Selection weights are used to compensate for this sub-selection of eligible adults (see section 7.1.1).

Around nine in ten (randomly selected) respondents aged 16-44 were asked to provide a urine specimen in order to test for *Chlamydia trachomatis, Neisseria gonorrhoeae*, HPV, HIV antibody and *Mycoplasma genitalium* (see section 3.1).

A random selection of respondents aged 18-74 were asked to provide a saliva sample in order to test for salivary testosterone (see section 3.2).

3 Data collection

3.1 Developing the questionnaire and field procedures

3.1.1 Designing the questionnaire

The Natsal-3 questionnaire was very similar to the questionnaire used in Natsal-2, which in turn was based on the questionnaire first used in Natsal-1. A full description of the development phase of the original Natsal-1 instrument covers issues such as questionnaire wording, terminology, confidentiality, reliability, validity, etc, as well as the feasibility and piloting work that preceded the main stage of the Natsal-1 survey.[7,19] Details of the development of the Natsal-2 questionnaire and piloting are also available, along with the results of an experiment that compared reports of sexual behaviour using PAPI vs CASI methods (which was a key part of the development work for Natsal-2).[4,6]

A number of changes were introduced to the Natsal-3 questionnaire to cover topics that were not included in the earlier surveys, to be relevant to the older age range included in Natsal-3, and to cover new topics of interest raised by members of the Natsal team or by other interested stakeholders. The main changes included adding or amending questions on:

- the general health module, which was extensively expanded to include new questions on health conditions and treatments received that may impact on sexual lifestyles
- non-volitional sex
- most recent partners, in order to better capture the details of the nature and characteristics of recent sexual partnerships
- fertility intentions and history of pregnancies along with the validated London Measure of Unplanned Pregnancy[14]
- the menopause and Hormone Replacement Therapy, reflecting the extended age range
- sexual function, which involved developing a validated measure (the 'Natsal-SF')[15]
- the use of Viagra and recreational drugs
- the PHQ-2, a two item screen for depressive symptoms.

Because of the addition of so many new items, some questions considered less important on Natsal-2 were dropped from the questionnaire in order to keep average interview length under an hour. (All the changes to the questionnaire are documented in the Natsal-3 codebook, which will be made available when the dataset is deposited with the UK Data Archive.)

As in the previous two surveys, Natsal-3 involved a combination of face-to-face interview, using CAPI, and self-completion format, using CASI. The face-to-face interview included questions on general health, family when growing up, learning about sex, first sexual experience, contraception, periods and menopause, attitudes to different kinds of relationships and socio-demographic details including cohabitation history. The most sensitive questions were asked in CASI, which covered: experiences of different types of sexual practices (vaginal, oral and anal intercourse); sex in the last 4 weeks and condom use; number of partners in different time periods (lifetime, 5 years, 1 year, 3 months); same sex sexual experiences (types of sexual practices, sex in the last 4 weeks, number of partners in different time periods); details of most recent partners; having sex with people from other countries, either in the UK or while abroad; non-volitional sex; paying for sex; history of pregnancies, unplanned pregnancy and family formation; fertility intentions and infertility; STI diagnoses and clinic attendance; HIV testing; sexual function and satisfaction; use of Viagra and recreational drugs; and depressive symptoms.

The CAPI and CASI questionnaires underwent thorough development and testing, described below.

3.1.2 Cognitive testing

The development of the questionnaire involved a number of stages, including extensive cognitive testing carried out in collaboration with NatCen's Question Development and Testing (QDT) research team. Cognitive interviews use techniques derived from cognitive psychological theory to explore the ways in which respondents react to and interpret the questions being asked.

Before fieldwork began, an expert panel was assembled to assist the research team prioritise the questions to be cognitively tested, and to advise on the wording of questions that could not be easily tested because they were about rare or particularly sensitive experiences (e.g. non-volitional sex). The cognitive piloting involved two phases:

- the first involved a pilot among members of the public who were respondents to the NatCen Omnibus survey; this included older respondents (ages 59-74) in order to test comprehension and acceptability of the questions among this age group (new to Natsal-3)
- the second phase was a pilot among people who had been recruited by advertisement, in order to find individuals who had more than one sexual partner in the last 5 years; this was necessary to thoroughly test the module of questions on recent sexual partners.

Across the two phases, a total of thirty-two cognitive interviews were conducted among a sample of people designed to reflect the range and diversity of the survey population.

Most modules in the questionnaire underwent cognitive testing with a particular focus on new questions and those where there were concerns about the acceptability of the question, the degree to which the question was understood as intended, the understanding of definitions which were key to the measurement aims of the survey, the ability to recall the information being sought, and respondents' willingness and ability to answer some of the more complicated and potentially burdensome questions.

The detailed results and recommendations of the cognitive testing work are reported elsewhere.[18,20]

The cognitive testing was followed by a series of workshops with members of the public where issues around response, survey documents and giving urine and saliva samples were discussed.

The cognitive testing work was approved by the NatCen Research Ethics Committee (Ref: I9699).

3.1.3 Piloting

An initial pilot of 214 respondents aged 16-74 was carried out in autumn 2009, which also involved the collection of urine and saliva samples. The pilot was mainly designed to test questionnaire flow, routing and length, the survey documents, and agreement rates to collecting the biological samples. Although this initial pilot was not explicitly designed to measure response rates (i.e. it was not a probability sample, the fieldwork period was short, there was no reissuing, and the survey documents mentioned that it was a pilot), the response rate (at 40.4%) was so much lower than anticipated, the research team felt it necessary to undertake a thorough review of the survey recruitment procedures, in particular, how Natsal-3 was presented to potential respondents and whether interviewer motivation might be improved.

As a result of this review, a number of changes were made to survey procedures, documents, and interviewer briefings, including:

- the advance letter was re-designed and re-written
- an advance leaflet was prepared for possible inclusion with the advance letter that was posted to sampled addresses; the leaflet included more detail about the survey and how Natsal data from earlier surveys has informed policy
- a new respondent website was launched, providing further information about the study
- a Natsal logo was designed for use on public documents and the website
- measures to improve interviewer motivation were taken including: concentrating more of the briefing sessions on how to explain the purpose of the survey to respondents, and showing interviewers a DVD which included a brief history of Natsal provided by members of the original Natsal-1 research team, and which described how findings from the previous surveys had been used by policy-makers.

These new procedures were then tested in a second pilot (referred to as a 'dress rehearsal'), along with an experiment to determine if increasing the respondent token of appreciation would influence response. The dress rehearsal involved two split sample experiments:

- respondents were offered either £15 or £30 gift vouchers for taking part in the interview
- half of the selected addresses were sent the advance leaflet with the advance letter in the post, and half received only the advance letter (no leaflet).

So, within each interviewer assignment for the dress rehearsal, addresses were randomly allocated to one of four different groups:

- £30 voucher, leaflet sent in advance
- £30 voucher, leaflet shown on the doorstep
- £15 voucher, leaflet sent in advance
- £15 voucher, leaflet shown on the doorstep

The dress rehearsal, carried out from March to May 2010, achieved interviews with 218 respondents, with a response rate of 50.3%, an increase of more than ten percentage points over the first pilot.

Response rates for the four groups in the split sample experiments were:

- 53.9% for £30 voucher, leaflet sent in advance (n=113)
- 52.9% for £30 voucher, leaflet shown on the doorstep (n=104)
- 47.8% for £15 voucher, leaflet sent in advance (n=109)
- 46.4% for £15 voucher, leaflet shown on the doorstep (n=108)

Overall, response for the £30 voucher was higher than for the £15 voucher (53.4% compared with 47.1% respectively), although the difference was not statistically significant given the small sample sizes in the dress rehearsal. The advantage for sending the leaflet in advance was marginal and also not significant: 50.9% compared with 49.6% for respondents who only received the advance letter.

Although the overall response to the dress rehearsal was still significantly below the rates achieved in Natsal-1 and Natsal-2 (66.8% and 65.4% respectively), there was no reissuing in the dress rehearsal and the fieldwork period was somewhat compressed. A number of further changes were made to fieldwork procedures before the start of the main data collection in summer 2010:

- the logos of the Medical Research Council (MRC) and Wellcome Trust were added to the advance letter and leaflet to emphasise the scientific importance of the survey

- minor wording changes were made to the respondent documents
- interviewers were provided with "Survey fact sheets" describing the achievements of past Natsal studies to equip them with additional information for answering respondent queries.

In terms of the token of appreciation, given the high cost involved of doubling its amount from £15 to £30, it was decided to continue the experiment during the first wave of data collection in order to measure its effect on a larger sample (the experiment is described in section 3.3.3 and results are shown in Table 5.3).

Full reports of the pilot and dress rehearsal will be available on the study website (www.natsal.ac.uk).

3.2 The Natsal-3 questionnaire

The Natsal-3 interview began and ended with face-to-face interviewing (CAPI), with the self-completion component (CASI) in the middle of the interview. The questionnaire comprised five sections, as follows:

- health, family and learning about sex
- first sexual experiences, use of contraception and sexual lifestyle
- the self-completion questionnaire, covering the most sensitive questions on sexual behaviour: number of partners, sexual practices, sexual health, etc.
- attitudes and risks
- socio-demographic questions.

Although the most sensitive questions were asked in the self-completion component, the face-to-face part of the interview also contained some personal questions, in particular those about sexual attraction and first sexual experiences. This was done partly to compare answers given in the face-to-face component with those given in the self-completion, but mainly to determine who was eligible for the self-completion module. The more sensitive questions which were asked face-to-face made use of show cards (included in Volume 2: Appendix A) to help preserve confidentiality (in case there was any chance of being overheard by other household members). It also meant that at no point in the interview did respondents have to verbalise any sexually explicit terms. There were separate versions of the show cards for men and women.

Health, family and learning about sex

As in the previous surveys, the Natsal-3 interview began with neutral questions to allow time for good rapport to develop between respondent and interviewer. The first CAPI component asked about the respondent's general health, whether they have had various health conditions or medical procedures (in particular those that may affect a person's sexual function and lifestyle such as cardio-vascular disease or surgery for various cancers), medications taken, height, weight, smoking and drinking behaviour. It then moved on to family background, and it ended with questions on learning about sex when growing up.

First sexual experiences, use of contraception and sexual lifestyle

The questionnaire then asked respondents to recall their first sexual experience (since the age of 13), including their age at first heterosexual intercourse. These questions were generally asked face-to-face. However, if interviewers felt that respondents might be inhibited from answering the questions (e.g. because of potentially being overheard by a third party), they had the option to defer these questions to the beginning of the self-completion questionnaire when they could be answered by the person more privately. This was done in 11.7% of interviews.

This section then asked about types of contraception used (ever, in the last year and usual use) and use of services for contraception. Women were then asked about menstruation, menopause and hormone replacement therapy. It ended with questions on sexual lifestyle, in the form of opposite-sex and same-sex attraction and experience scales.

The self-completion questionnaire

The most sensitive questions on sexual behaviour were asked in a self-completion questionnaire to reassure respondents of confidentiality and avoid embarrassment. Answers to the questions on early sexual experience determined whether respondents were given this questionnaire. Respondents who had no sexual experience of any kind were not given the self-completion module. Those who had some sexual experience, but not heterosexual vaginal, anal, or oral sex or same-sex experience involving genital contact, were asked a shorter version of the CASI, since many questions were not applicable to these respondents. Overall, 1.9% of respondents were not eligible for this module at all, and 3.9% were given the shortened CASI version.

The self-completion questionnaire was to be completed by the respondent directly on the laptop computer. Respondents with literacy or eyesight problems (or who had difficulty reading English because it was not their first language) had the option to have the questions in the self-completion module read out to them by the interviewer, as did those respondents who did not feel comfortable using a computer.

Respondents were first taken through a few simple practice questions by the interviewer to demonstrate how to use the laptop. They were asked to follow the instructions on the laptop screen and to enter their answers appropriately.

At the beginning of the self-completion questionnaire, a number of key terms were defined in order to ensure respondents interpreted questions in the same way. For example, some heterosexual respondents might interpret the term ‘having sex’ or ‘sexual intercourse’ to refer only to vaginal intercourse. For the survey, however, ‘sexual intercourse’ was defined to include vaginal, oral and anal sexual intercourse. Interviewers emphasised that respondents must read these definitions before answering the self-completion module. If necessary, respondents could refer to these definitions again at relevant questions by pressing a key on the laptop which brought the definitions up on screen.

The self-completion module itself then began with questions on heterosexual practices (e.g. when last had vaginal, oral and anal intercourse), followed by similar questions on same-sex sexual practices. The next section covered the number of opposite-sex and same-sex partners within different time periods (lifetime, last 5 years, year, and 3 months). Next, there was a section of questions on respondents’ most recent sexual partners. Up to four most recent partners could be asked about, including first and last occasions of sex with the partner, use of condoms, how met, and overlapping partnerships. The following set of questions covered having sex with people from other countries, and sex with new partners while abroad. Next was the new section of questions on non-volitional sex, followed by questions about paying for sex. Women were then asked for their history of pregnancies, covering all births, miscarriages and abortions, and the planned status of recent pregnancies, while men were asked about any children they had. There then followed questions on infertility, sexually transmitted infections and HIV testing. The validated module of questions on sexual function and satisfaction, developed specifically for inclusion in Natsal-3, was next, followed by questions on recreational and injecting drug use, and depressive symptoms.

Respondents who were eligible only for the shortened CASI version were asked questions on: opposite-sex experience involving genital contact, masturbation, fertility intentions, chlamydia

and HPV tests/vaccinations, cervical screening, circumcision, HIV testing, satisfaction with sex life, sexual difficulties, use of recreational drugs, and depressive symptoms.

At the end of the CASI questionnaire, as a means of reassuring respondents of the confidentiality of their replies, they were informed that their answers would be 'locked' in the computer so that neither the respondent nor the interviewer could access the self-completion module once it had been finished.

Full advantage was taken of CASI (e.g. compared with a paper self-completion questionnaire) to build in complex filtering (so that respondents are filtered past questions which do not apply to them given their response at earlier questions) as well as a number of range and consistency checks to ensure high data quality. For example, when asking respondents for the number of sexual partners they had in the last year, the programme checked that the number keyed in was not greater than the number of partners they said they had in the last five years. When this did happen, the CASI programme would prompt the respondent to check that they had correctly keyed in their answers to the questions which were inconsistent. Respondents could then change their response to one of the questions or they could elect to leave their responses so the apparent inconsistency would be retained.

Attitudes and risks

At the end of the self-completion component, the interview reverted to face-to-face CAPI with a module of questions on attitudes to sexual relationships (e.g. views on same-sex relationships, young people and sex). As in the previous Natsal surveys, these questions were deliberately asked after the questions on sexual behaviour so that respondents could report their own behaviour before being asked to express their views on different types of sexual relationships.

Respondents were also asked to judge how much they thought they were at risk of becoming infected with STIs and HIV.

Socio-demographic questions

The final part of the interview collected demographic information, including marriage and cohabitation details, economic status, occupation, tenure, education, and religious, ethnic and sexual identity.

The CAPI and CASI questionnaires are fully documented in Volume 2: Appendix B, along with the other Field Documents. The questionnaire topics are summarised in Box 1.

The questionnaire was only available in English and was not translated into any other languages. Translations by interviewers or other household members were *not* permitted, so sampled individuals who did not speak English sufficiently well to carry out the interview were not included in the survey.

BOX 1: Natsal-3 questionnaire content

General health, health conditions, medications taken, medical procedures (that may affect a person's sex life)
Family when growing up
Learning about sex
First heterosexual experience
Contraception used
Periods, menopause and use of hormone replacement therapy
Experience of different heterosexual practices (vaginal, oral and anal intercourse)*
Opposite-sex sex in the last 4 weeks and condom use*
Same-sex sexual experiences (types of sexual practices, sex in last 4 weeks)
Number of opposite-sex partners in different time periods (lifetime, 5 years, 1 year, 3 months)*
Number of same-sex partners in different time periods*
Details of most recent partners*
Having sex with people from other countries and while abroad*
Non-volitional sex*
Paying for sex*
Family formation, pregnancy history and unplanned pregnancy*
Fertility intentions and infertility*
STI diagnoses and clinic attendance, HPV vaccination and cervical screening*
Circumcision*
HIV testing*
Sexual function and satisfaction*
Use of Viagra*
Use of recreational drugs*
Screen for depressive symptoms*
Attitudes to different kinds of relationship and sexual lifestyles
Perceived risk of HIV and other STIs
Previous live-in partnerships
Socio-demographics
(* asked in CASI)

3.3 Fieldwork procedures

3.3.1 Organisation of fieldwork

As described in section 2, the original sample design involved selecting postcode sectors as the PSUs. The 1727 sectors issued during the two years of fieldwork were randomly allocated to one of eight waves. Fieldwork was carried out over 23 months. One wave was issued in each quarter (approximately), as follows:
Wave 1: September 2010
Wave 2: January 2011
Wave 3: April 2011
Wave 4: July 2011
Wave 5: October 2011
Wave 6: January 2012
Wave 7: March 2012
Wave 8: May 2012

Each wave was due to be completed within three months: original issues were normally completed within six weeks, with the following six weeks allocated to re-issuing. Fieldwork was completed in August 2012.

In July 2012, a 'wave 9', consisting of 1142 unproductive addresses from waves 1-6, was re-issued to overlap with the re-issue period of wave 8. Broadly, all original unproductive addresses were eligible to be reissued in 'wave 9', except for 'deadwood' (e.g. vacant, derelict) addresses, office refusals and non-English speakers. They were then batched and covered along with the wave 8 reissues (unless there were only 1 or 2 addresses within an area, in which case they were excluded from reissue). 'Wave 9' reissues resulted in an additional 137 interviews being achieved.

3.3.2 Recruiting respondents

First, every sampled address was sent an advance letter and leaflet which gave some background information about Natsal-3, described its importance and uses to policy-makers and researchers, and stated that an interviewer would soon be calling to seek permission to carry out the interview. It also included the eligible age range being screened for at that address and that each person who participated would be given a £15 gift voucher. The letter included a space for interviewers to write in their name so that respondents knew who would be calling and to make the letters more personal. Copies of the letter and leaflet are included in Volume 2: Appendix A.

Interviewers were notified of any office refusals as a result of the advance letter, and interviewers were not required to visit these addresses.

Soon after the advance letters and leaflets had been sent, interviewers made contact at the address by a personal visit. Interviewers were required to make a minimum of 6 calls, up to a maximum of 9. These calls had to be at different times of day and on different days of the week. If there was still no contact, only then could an interviewer return a case as a 'non-contact'.

At first contact, the interviewer established the number of 'dwelling units' at an address. A dwelling unit is a living space which has its own front door. The front door does not have to be at street level, but it must separate one part of the address from other parts (i.e. only those who live behind the door have access to the area; it is not a communal part of the address).

A dwelling unit need not be fully self-contained - for example, an address may contain four bed-sitters, the inhabitants of whom share a bathroom. Each bed-sitter would count as a dwelling unit as long as it had its own front door. If there was more than one dwelling unit, one would be picked for inclusion in the survey using a random selection procedure (Kish grid).

The interviewer then made contact at the (selected) dwelling unit and established the number of residents aged 16-74 (at core addresses). If there were no residents in the relevant age range, the address was ineligible for interview. If there was one person aged 16-74, an interview was attempted with that person. If there was more than one resident aged 16-74, one was selected at random using a Kish grid. Once the selection was made, no substitutions were allowed. The procedure was the same at the boost addresses, but the eligible residents were in the age range of 16-34 (boost 1) or 16-29 (boost 2).

The survey and its purposes were fully explained to the selected person, and an information leaflet was provided which described the study in more detail and explained how households were chosen. The leaflet also explained that participation was voluntary and that any information they provided would remain confidential and would not be passed on to anyone outside the research team in a form that could be used to identify them. Respondents were provided with a telephone number that they could use if they had any queries. Any substantive queries or complaints were subsequently passed on to researchers to deal with. (A copy of the information leaflet is found in Volume 2: Appendix A.)

If the selected person was aged 16 or 17 and living in the parental home, then the interviewer obtained agreement from one of the parents as well as the young person before arranging an interview. Although not strictly necessary, parental agreement for older teenagers living at home was also obtained if the interviewer felt it to be necessary for reasons of courtesy.

3.3.3 Token of appreciation/gift voucher

Given the preliminary findings from the dress rehearsal that a higher value token of appreciation increased response, albeit non-significantly due to the small sample size (see section 3.1.3), it was decided to continue the experiment on a larger scale. During wave 1 of fieldwork (September to November 2010), a split run experiment was carried out to compare the relative impact on response rates of the £15 and £30 gift vouchers. Although the response rate to the £30 voucher was higher than for the £15 voucher (58.9% vs. 56.3%), it was not judged to be a large enough difference - it was not in fact statistically significant - to warrant the extra costs (of over £200,000) the higher amount would have entailed for this large sample. Detailed response figures are shown in Table 5.3 (in section 5.1). Thus, a £15 gift voucher was provided to all respondents from wave 2 on.

Respondents who provided a urine sample were given an additional £5 gift voucher, as were those who provided a saliva sample.

3.4 Quality control

3.4.1 Interviewer training and briefing

All interviewers working on Natsal-3 attended briefing meetings run by Natsal-3 researchers and field staff. All interviewers had already undergone standard NatCen training in interview techniques, including the use of CAPI. The first briefing session was held in August 2010, and the last was in May 2012. A total of 491 interviewers attended the 42 briefing meetings held around Britain.

Each day-long briefing session covered:

- the background and purpose of the survey
- instructions about sampling procedures and the importance of high response rates
- guidance on introducing the survey to different types of respondents, including “role play” exercises to practice doorstep techniques
- familiarisation with key aspects of the questionnaire, and explanation of important questions and topics
- guidance on how to handle the self-completion section of the interview
- training on all study protocols, including obtaining informed consent
- training in the protocol for collecting urine specimens
- training in the protocol for collecting saliva samples.

The briefing sessions showed a 10 minute video, which included interviews with key members of the research team and with interviewers who worked on the Natsal-3 pilots. Most of the briefing sessions were attended by one or more of the Natsal-3 co-investigators.

Interviewers were also given a pre-briefing exercise, which involved completing a number of practice interviews to familiarise themselves with the questionnaire.

Full sets of written instructions on survey procedures were also provided to all interviewers.

3.4.2 Supervision and quality control

Quality control procedures were built into Natsal-3 at data collection and subsequent stages. Interviewers new to NatCen are accompanied on their first day of working by an experienced interviewer. Moreover, all NatCen interviewers, whatever their level of experience, are supervised in the field every six months in a rolling programme – 161 interviewers were supervised while interviewing on Natsal-3.

Sample recalls to check on the work of interviewers were successfully carried out at 1561 productive addresses: 1527 by telephone and 34 by post. These recalls were carried out on the work of 311 interviewers, which is 63% of those who worked on the study.

The computer program used by interviewers had in-built 'soft' and 'hard' checks, which included messages querying uncommon or unlikely answers, as well as consistency between answers (as described in section 3.2). Extensive consistency checks were also implemented during the data cleaning stage (see section 6.2).

4 The biological samples

4.1 Collecting and testing urine samples

4.1.1 Eligibility for providing a urine sample

Urine samples were collected to test for a number of sexually transmitted infections (STIs). Samples were initially tested for *Chlamydia trachomatis*, *Neisseria gonorrhoeae*, Human Immunodeficiency Virus (HIV) antibody, type-specific Human Papillomavirus (HPV) and *Mycoplasma genitalium,* with residual specimens stored for further testing. All men and women aged 16-44 who reported at least one sexual partner were eligible for providing a urine sample. With a target of 5000 samples, and assuming response would be around 60% to this part of the study, all eligible 16-24 year olds, all men aged 25-44 who reported having sex with another man in the last 5 years, and a randomly selected 85% of eligible respondents aged 25-44 (covering all PSUs), were asked to provide a sample. Urine samples were taken from *all* 16 and 17 year olds, regardless of their sexual experience, so as not to inadvertently alert other family members who may have been present in the household to the young person's sexual experience.

4.1.2 The protocol for collecting urine samples

Urine collection protocol

The purpose and procedure of the urine test was explained to all eligible respondents at the end of the main interview. As well as a verbal explanation, respondents were given a leaflet describing the purpose of the urine test and what would be involved. It was also carefully explained that the test results would be anonymised, and that they would not be told their individual results.

Written consent was obtained from respondents prior to their providing the sample. Samples were not tested until the consent form was received and checked by the NatCen Operations team. Written consent was also requested for storing any remaining urine for potential future measurement of the prevalence of past or current infection with other (unspecified) pathogens. The full protocol for urine sample collection is included in Volume 2: Appendix C.

Up to 5ml of urine was collected in a sterile plastic container, called a 'FirstBurst Urine Collection Device' (see below). The specimens were packaged by the interviewer and posted to the Health Protection Agency (now Public Health England (PHE)) on the same day. No urine collection was to take place on a Friday or Saturday before a bank holiday, to avoid postal delays in samples arriving at the laboratory. Respondents who provided a sample were given a £5 gift voucher by the interviewer.

FirstBurst Urine Collection Device

The FirstBurst Urine Collection Device is designed to collect the first part of the urine stream. It was chosen because it has been shown to yield a specimen with a six-fold higher *C. trachomatis* organism load than the regular urine cup.[21] We also conducted a study comparing first void with flow through urine in a sample of HIV-positive patients, to validate the FirstBurst for HIV antibody and HPV DNA (MREC N° 09/H0715/73 and UKCRN ID 7860).

Anonymity of results

Unlike in Natsal-2, when respondents who tested positive for *Chlamydia trachomatis* were informed of their results, Natsal-3 opted for anonymised testing, with respondents made aware before providing the sample that they would not be given their test results. This anonymised strategy was adopted for Natsal-3 for a number of reasons:

- tests used in the survey do not necessarily reach the level of clinical diagnostic accuracy, largely due to study conditions (e.g. the need to transport, freeze and batch-test the samples), resulting in some loss of sensitivity and the greater likelihood therefore of having both false positive and false negative results
- the widespread availability of STI/HIV testing services in Britain, which provide a more appropriate clinical setting for free testing than that provided by the non-clinical interviewers used on Natsal-3
- the clinical and public health implications for some STIs (e.g. *M. genitalium* and HPV) remain unclear, and a positive result would not necessarily require treatment or partner notification, but may cause unnecessary distress
- providing results for only some STIs but not others may be misinterpreted by respondents that they are negative for all STIs
- it can be very costly to actively follow-up all positive results, and even with staff in place to do so, it would not be possible to contact all positive cases (for example, due to respondents changing address), and there are concerns about the ethical consequences caused by the inability to contact respondents.

A full description of the development of this approach has been published elsewhere.[16]

This approach was tested during the development stages of Natsal-3 and was found to be acceptable. In the dress rehearsal, 61% of respondents agreed to provide a urine sample without receiving results, and few raised this as a concern; those who did were reassured after discussing their concerns with the interviewers.

Maintaining confidentiality of the urine test results

Since three separate organisations were involved in the STI component of Natsal-3 (NatCen, UCL and PHE), careful procedures were put in place to link the STI and survey results while maintaining data confidentiality in such a way that no one institution was able to link STI test results with personal identifying information. See Volume 2: Appendix D for full details of the anonymised linking procedure.

At the end of the study all temporary files will be deleted, making it impossible to connect results to personal identifying information. Respondents had the option to withdraw their urine test results from the study up to three months after the sample was collected.

4.1.3 Testing the urine samples

Upon receipt, specimens were divided into aliquots ahead of their respective testing procedures as follows:

- for *Chlamydia trachomatis* and *Neisseria gonorrhoeae,* a 1mL aliquot was placed into Aptima® buffer and stored at 4°C
- for HIV antibody testing, a 0.5mL aliquot of urine was kept at 4°C
- for HPV and *Mycoplasma genitalium* testing, a 1mL aliquot of urine was subjected to centrifugation at 13,000xg for 20 minutes before the cellular pellet was re-suspended in 300µL ice cold phosphate-buffered saline and stored at -80°C prior to robotic DNA extraction; remaining urine was stored, unprocessed, at -80°C.

Chlamydia trachomatis and *Neisseria gonorrhoeae* testing was undertaken using the Aptima Combo 2 test (Hologic; Genprobe) as an initial screen and any equivocals or positives were subsequently confirmed using Aptima CT or Aptima GC monospecific tests, respectively. For *Mycoplasma genitalium* all urine specimens were tested using an in-house Real-Time PCR assay

which targets the MGPA gene,[22] with positive samples also tested using a Genprobe Mycoplasma test.[23]

An *in house* Luminex®-based genotyping assay was used for the detection of HPV types.[24] High-risk HPV types (HR-HPV) were defined as HPV types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68.[25]

All specimens were examined for the presence of sufficient total IgG antibody concentration by an *in house* total human IgG ELISA before being tested for the presence of HIV-1/-2 antibodies. Samples were first screened for the presence of HIV-1/-2 antibodies using a modified GACPAT (IgG antibody capture particle adherence test) method, as previously described.[26,27] Based on 711 specimens, GACPAT has been shown to reliably detect anti-HIV in urine[28] with high sensitivity (100%) and specificity (99.6%), and has therefore been used to estimate the prevalence of HIV on anonymized specimens in population surveys.[29] HIV-1/-2 antibody positive or equivocal samples were confirmed using the MP Diagnostics (MP Biomedicals, UK) HIV Western Blot 2.2, a qualitative enzyme immunoassay which has been validated for the investigation of anti-HIV specific antibodies in urine.

The laboratories at PHE-Colindale are CPA accredited (Clinical Pathology Accreditation (UK) Limited). This aims to ensure that clinical laboratory test results are accurate, reliable and comparable wherever they are produced.

4.2 Collecting and testing saliva samples

4.2.1 Eligibility for providing a saliva sample

Saliva samples were collected to test for free testosterone in men and women. All men and women aged 18-74 were eligible for providing a saliva sample, except for those who regularly worked at night. With a target of 4400 samples, and assuming a 75% response rate to this part of the study, initially a randomly selected 30% of respondents aged 18-34 and two-thirds of those aged 35-74 (covering all PSUs) were asked to provide a saliva sample. Due to lower than anticipated response, and in order to achieve the target number of samples, these proportions were increased to 36% of those aged 18-34 and 100% of those aged 35-74 for waves 4 and 5, with a further increase to 75% of those aged 18-34 for waves 6 to 8.

4.2.2 The protocol for collecting saliva samples

The purpose and procedure of the saliva test was explained to all eligible respondents at the end of the main interview (and after the urine sampling for the 26% of respondents who were asked to provide both samples). As for the urine test, respondents were given both a verbal explanation and a written leaflet describing the purpose of the saliva test and what would be involved. It was also explained that the saliva test would be anonymised, and that they would not be told their individual results. Written consent was obtained from respondents. Consent was also requested for storing any remaining sample for potential future research studies. The full protocol for saliva sample collection is included in Volume 2: Appendix E.

Respondents who agreed to participate were asked to provide a saliva sample in the morning before 10am, so as to account for the diurnal variation in testosterone. They were asked not to brush their teeth, eat or chew before giving the sample, to reduce potential blood contamination caused by micro-injury. Respondents were asked to drool directly into a polystyrene tube (avoiding potential assay interference from cotton swabs). Respondents were then asked to post the samples on the day of collection to the Department of Clinical Biochemistry, Glasgow Royal Infirmary, for storage and preparation. Once the saliva samples were received by the laboratory, respondents were sent a £5 gift voucher. Samples were only

released for testing once the NatCen Operations team had received and checked the signed consent forms.

Respondents were asked whether they would be willing to be telephoned by the interviewer on the day following the interview in order to remind them to provide the sample and to answer any questions they may have. Interviewers could also inform respondents who may have forgotten to provide the sample before 10am, that they could still do so on the following day. 76% of those who agreed to give a sample also agreed to take the follow-up reminder call.

4.2.3 Testing the saliva samples

Upon receipt by Glasgow Royal Infirmary, samples were frozen at -20°C for at least 24 hours. After thawing, they were centrifuged at 2000g for 10 minutes at 4° to precipitate muco-polysaccharides. The clear supernatant in each sample was then divided into two aliquots of 0.25ml and the remainder in a third tube for long term storage. The two small aliquots were then frozen at -80°C. Once the serial number had been verified and consent had been confirmed by NatCen, the samples were batched and sent in dry-ice to University Hospital South Manchester (USHM) for analysis. The remaining aliquot was stored at -80°C. The testing was carried out by the Department of Clinical Biochemistry, UHSM.

Analysis by Salivary (Sal)-T liquid chromatography tandem mass spectrometry LC-MS/MS

Sal -T in men was initially measured using LC-MS/MS according to a previously published method, [17,30] but due to the low concentrations of testosterone in women it was necessary to improve the assay sensitivity. This was achieved using a research grade LC-MS/MS instrument specifically to improve the sensitivity of testosterone measurement in samples from women.

Sample preparation for both men and women involved the same liquid–liquid extraction procedure by adding sample (200 µL), D5-testosterone internal standard (10 µL, 340 pmol/L) and methyl-tert-butyl ether (1 mL), vortex mixing for 5 minutes and then freezing at -80°C. After 1 hour at this temperature, the organic layer was transferred and evaporated by heating and gentle N_2 gas flow. The residue was reconstituted with a 500 mL/L methanol mobile phase (80 µL) and transferred to a 96-well microtiter plate.

Extracted sample (35 µL) was injected onto the analytical column. Liquid chromatography was performed with an ACQUITY™ Ultra Performance Liquid Chromatography system (Waters Corporation, Manchester, UK) and a C8 Kinetex column 3.0x100mm 2.6 µm (Phenomenex, Macclesfield, UK) maintained at 45°C. The mass spectrometer was a Xevo TQ-S™ mass spectrometer (Waters Corporation, Manchester, UK) operated in positive ionization mode. The assay was validated according to the published Bioanalytical Method Validation guidelines issued by the U.S. Department of Health and Human Services Food and Drug Administration,[31] which include acceptance criteria for linearity, imprecision, recovery and sample stability.

Calibrator accuracy was assessed prior to the study by comparing calibrators, prepared from pure testosterone powder (Cat No; T1500, Sigma-Aldrich Company Ltd, Gillingham, UK.) dissolved in methanol and then diluted into phosphate buffered saline (PBS)/bovine serum albumin as an artificial matrix, with routine LC-MS/MS testosterone calibrators. We have previously shown that measuring serum T samples using calibrators made in this manner has excellent agreement with the reference method.[32]

The lower limit of quantitation (LLOQ) was defined as the concentration for which 10 replicates of Phosphate buffered saline (PBS)-based samples prepared with low concentrations of testosterone gave a coefficient of variation (CV) of less than 20% and bias of less than 20%.

To determine the stability of the extracted saliva, 35 different samples were initially analysed on the LC-MS/MS. The 96-well microtitre plate was then resealed and analysed 24 hours later after overnight storage at 4°C.

Analytical performance

The LC-MS/MS method for measuring Sal-T in men has previously been shown to have good analytical performance.[17] The extracted saliva samples from women were measured using the more sensitive LC-MS/MS method described and produced clean chromatograms with no interfering peaks when using the testosterone quantifier transition m/z 289.3>97.15 (qualifier 289.3>109.2). Following injection of the extracted sample (35 µL), testosterone and D_5-testosterone co-eluted with clean, discrete and identifiable peaks at a retention time of 3.6 minutes. The total run time, injection to injection, was 6.5 minutes. Infusion experiments showed that there was no significant ion suppression present in any of the six chromatograms from extracted saliva when compared with an injected water sample. Standard curves were made by plotting testosterone concentrations on the x-axis and testosterone/ D_5 testosterone peak area ratios on the y-axis. The curve was linear over the standard range and was reproducible between batches. The curves showed good correlation with the assigned standard values with an r^2 value of 0.999. Overnight storage of the extracted saliva at 4°C in 35 different samples showed no decrease in measured testosterone values, indicating good sample stability. The lower limit of quantification of the more sensitive assay was 5 pmol/L. Inter-assay CV (SD) bias was 12.9% (1.7) and 1.2%; 9.8% (2.5) and 0.4%; 4.5% (12.0) and 1.9%, at concentrations of 12.9, 26.0 and 260 pmol/L. Intra-assay CV (SD) bias was 9.5% (1.3) and 0.8%; 5.5% (1.6) and 12.6%; 2.1% (6.2) and 11.1%; at concentrations of 12.9, 26.0 and 260 pmol/L. Recovery was 104% (range 98.3–108.9%).

Sample stability during transport and storage

One-way repeated measures ANOVA showed that Sal- T concentrations in different saliva samples from men stored at both room temperature and at 4°C for a period of 5 days did not differ significantly (F=0.88, P=0.50 and F=1.03, P=0.42, respectively). There was also no significant difference in Sal-T measurements over five freeze-thaw cycles (F=1.29, P=0.31) or after 3 years storage (mean T before storage 217 pmol/L and after storage 202 pmol/L). Paired-sample t-tests showed that LC-MS/MS Sal-T (t=2.88, P<0.01) were significantly lower at the second (mean decrease =15 pmol/L) as compared with the first measurement. However, the mean decrease in testosterone over the 3 year period was 4.7% (95% CI 0.4–9.0) which is within the analytical variation of the assay; this confirmed the stability of testosterone in saliva after 3 years storage at 80°C.

Details of the testing and validation of the LC-MS approach have been published elsewhere.[30]

5 Response

5.1 Response to the interview

The sample design, described in section 2, required that one randomly selected adult aged 16-74 (16-34 in the boost sample) be interviewed in each eligible sampled address. Table 5.1 shows the response rate calculations. Response rates can be calculated in different ways. Natsal-3 calculations follow the standards set by the American Association for Public Opinion Research (AAPOR), which sets out a range of different formulas for calculating response rates, cooperation rates, refusal rates and contact rates. In this report, we provide calculations for AAPOR's response rates 1, 3 and 5 (RR1, RR3, RR5) and its cooperation rate 2 (COOP2).[33]

Ineligibles

Of the 59,412 issued addresses, 6.8% were not residential, occupied or traceable. At 46.9%, there were no residents in the eligible age range. Ineligibility by age varied considerably by address sample type: the lowest proportion of age ineligibles was at core addresses (14.5% of issued addresses), while boost addresses had much higher rates of ineligibles (67.1% of issued addresses for boost 1 and 74.2% for boost 2) because of the much narrower age range for eligibility. Removing these addresses from the base leaves 27,503 potentially eligible addresses.

Of these, no contact was made at 1056 addresses (1.8% of all issued addresses) after repeated calls (at least six), and at 2501 addresses (4.2%) insufficient information was obtained about whether there were any eligible residents to enable a selection of an individual to be made. (These include addresses where the residents did not speak English well enough to establish eligibility.) Overall, eligibility could not be established at 4143 addresses, and some of these would be ineligible because the residents would be outside the relevant age range.

Recommended practice is to use the best evidence available for estimating the proportion of ineligibles at addresses where ineligibility is unknown.[34] The best estimate is that 1229 of these addresses would be ineligible. This is derived by assuming that the percentage of age ineligibles at the addresses where eligibility is unknown will be the same as for the addresses where eligibility is *known* (i.e. issued minus unknown addresses): 16.6% at core addresses, 69.4% at boost 1 addresses and 76.0% at boost 2 addresses. These percentages were used to estimate, separately for each sample type, the number of ineligible households where eligibility was not known due to non-contact, refusal of all information about the address, or the presence of non-English speakers only.

Unproductives

The (estimated) total number of eligible households is 26,274: 23,360 known eligibles plus 2914 assumed eligibles. This number forms the denominator for calculating the response rate. At these addresses, there was no contact with the selected person at 327 (1.2% of estimated eligibles), while at 6343 (24.1%) the selected person refused to participate (including proxy refusals). The selected person was ill or away, did not speak English or did not participate for some other reason at a further 1528 (5.8%) addresses.

Productives

Interviews were completed with 15,162 respondents. Following AAPOR's formula for calculating RR3, which estimates the number of ineligibles at addresses of unknown eligibility (as described above), Natsal-3 achieved a response rate of 57.7%. RR3 was lowest in the core addresses (54.5%) and considerably higher in both sample boosts: 64.8% in boost 1 and 67.3% in boost 2.

RR3 varied only slightly by wave: from 59.3% in wave 5 to 55.7% in wave 8 (Table 5.2). (The 137 respondents interviewed as part of a 'wave 9' reissue – see section 3.3.1 – are counted in the figures for the wave of their first issue in Table 5.2.)

It is also possible to show response as a range by assuming that all of the addresses where eligibility is unknown are either eligible (lower limit) and included in the unproductive outcomes (AAPOR RR1), or ineligible (upper limit) and excluded from the response calculation (AAPOR RR5). If it is assumed that all 4147 addresses of unknown eligibility were eligible, the lower limit of response (RR1) is 55.1%; if it is assumed that they are all ineligible, the upper limit of response (RR5) is 64.9%.

An alternative calculation is the 'Cooperation Rate', which is the proportion of all cases interviewed of all eligible units ever contacted. Natsal-3 achieved a cooperation rate of 65.8%, when using AAPOR's formula for Cooperation Rate 2 (COOP2). (COOP2 is used instead of COOP1, as the former includes partial interviews, and Natsal-3 productives include a small number of partial interviews.)

Table 5.1 Response rate (RR3) for Natsal-3 for core and boost sample types

	All		**Core (16-74)**		**Boost 1 (16-34)**		**Boost 2 (16-29)**	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Sampled addresses	**59412**		**24924**		**18537**		**15951**	
Out of scope addresses:								
Vacant/ derelict	3137	5.3	1620	6.5	828	4.5	689	4.3
Non-residential	710	1.2	310	1.2	203	1.1	197	1.2
Not traced built/ other	177	0.3	92	0.4	39	0.2	46	0.3
Not eligible age range	27885	46.9	3613	14.5	12438	67.1	11834	74.2
Total known ineligibles	**31909**	**53.7**	**5635**	**22.6**	**13508**	**72.9**	**12766**	**80.0**
Unknown eligibility:								
No contact	1056	1.8	698	2.8	206	1.1	152	1.0
All information refused	2501	4.2	2048	8.2	291	1.6	162	1.0
Other	586	1.0	418	1.7	106	0.6	62	0.4
Total unknown eligibility	**4143**	**7.0**	**3164**	**12.7**	**603**	**3.3**	**376**	**2.4**
Estimated ineligible	1229		525		418		286	
Total estimated eligible addresses:	**26274**	**100**	**18764**	**100**	**4611**	**100**	**2899**	**100**
No interview:								
No contact with selected person	327	1.2	190	1.0	86	1.9	51	1.8
Refused (including proxy refusal)	6343	24.1	4668	24.9	1049	22.8	626	21.6
Other reason	1528	5.8	1043	5.6	303	6.6	182	6.3
No information about address	2914	11.1	2639	14.1	185	4.0	90	3.1
Total unproductive	**11112**	**42.3**	**8540**	**45.5**	**1623**	**35.2**	**949**	**32.7**
Completed interviews	**15162**	**57.7**	**10224**	**54.5**	**2988**	**64.8**	**1950**	**67.3**

Table 5.2 Response rate (RR3) for Natsal-3 by wave of fieldwork

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8	All
	N *(%)*	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Sampled addresses	**6750**	**7050**	**7812**	**7812**	**7812**	**7812**	**7812**	**6552**	**59412**
Out of scope addresses:									
Vacant/ derelict	329 *(4.9%)*	375 *(5.3%)*	449 *(5.7%)*	454 *(5.8%)*	404 *(5.2%)*	422 *(5.4%)*	357 *(4.6%)*	347 *(5.3%)*	3137 *(5.3%)*
Non-residential	80 *(1.2%)*	80 *(1.1%)*	87 *(1.1%)*	91 *(1.2%)*	118 *(1.5%)*	84 *(1.1%)*	96 *(1.2%)*	74 *(1.1%)*	710 *(1.2%)*
Not traced built/ other	15 *(0.2%)*	16 *(0.2%)*	41 *(0.5%)*	19 *(0.2%)*	26 *(0.3%)*	19 *(0.2%)*	14 *(0.2%)*	27 *(0.4%)*	177 *(0.3%)*
Not eligible age range	3192 *(47.3%)*	3329 *(47.2%)*	3612 *(46.2%)*	3669 *(47.0%)*	3661 *(46.9%)*	3718 *(47.6%)*	3698 *(47.3%)*	3006 *(45.9%)*	27885 *(46.9%)*
Total known ineligibles	**3616**	**3800**	**4189**	**4233**	**4209**	**4243**	**4165**	**3454**	**31909**
Unknown eligibility:									
No contact	138 *(2.0%)*	130 *(1.8%)*	145 *(1.9%)*	132 *(1.7%)*	106 *(1.4%)*	137 *(1.8%)*	130 *(1.7%)*	138 *(2.1%)*	1056 *(1.8%)*
All information refused	236 *(3.5%)*	292 *(4.1%)*	316 *(4.0%)*	349 *(4.5%)*	329 *(4.2%)*	303 *(3.9%)*	356 *(4.6%)*	320 *(4.9%)*	2501 *(4.2%)*
Other	83 *(1.2%)*	62 *(0.9%)*	79 *(1.0%)*	81 *(1.0%)*	74 *(0.9%)*	71 *(0.9%)*	76 *(1.0%)*	60 *(0.9%)*	586 *(1.0%)*
Total unknown eligibility	**457**	**484**	**540**	**562**	**509**	**511**	**562**	**518**	**4143**
Estimated ineligible	139	149	167	165	139	161	156	154	1229
Total estimated eligible addresses:	**2995 *(100%)***	**3101 *(100%)***	**3456 *(100%)***	**3414 *(100%)***	**3464 *(100%)***	**3408 *(100%)***	**3491 *(100%)***	**2944 *(100%)***	**26274 *(100%)***
No interview:									
No contact with selected person	52 *(1.7%)*	39 *(1.3%)*	47 *(1.4%)*	47 *(1.4%)*	23 *(0.7%)*	35 *(1.0%)*	44 *(1.3%)*	40 *(1.4%)*	327 *(1.2%)*
Refused (including proxy refusal)	710 *(23.7%)*	764 *(24.6%)*	867 *(25.1%)*	798 *(23.4%)*	831 *(24.0%)*	834 *(24.5%)*	831 *(23.8%)*	708 *(24.0%)*	6343 *(24.1%)*
Other reason	165 *(5.5%)*	169 *(5.5%)*	184 *(5.3%)*	224 *(6.6%)*	187 *(5.4%)*	191 *(5.6%)*	215 *(6.2%)*	193 *(6.6%)*	1528 *(5.8%)*
No information about address	318 *(10.6%)*	335 *(10.8%)*	373 *(10.8%)*	397 *(11.6%)*	370 *(10.7%)*	350 *(10.3%)*	406 *(11.6%)*	364 *(12.4%)*	2914 *(11.1%)*
Total unproductive	**1245 *(41.6%)***	**1307 *(42.1%)***	**1471 *(42.6%)***	**1466 *(42.9%)***	**1411 *(40.7%)***	**1410 *(4142%)***	**1496 *(428%)***	**1305 *(44.3%)***	**11112 *(42.3%)***
Completed interviews	**1750 *(58.4%)***	**1794 *(57.9%)***	**1985 *(57.4%)***	**1948 *(57.1%)***	**2053 *(59.3%)***	**1998 *(58.6%)***	**1995 *(57.1%)***	**1639 *(55.7%)***	**15162 *(57.7%)***

Response by area deprivation

The Index of Multiple Deprivation (IMD) is used to classify small geographic areas in terms of their level of relative deprivation using a range of measures from different domains including income, employment, health, education, housing, crime, and others.[35] Response (RR3) showed little variation by IMD quintile:

IMD quintile	Response rate (RR3)
1 (least deprived)	57.5%
2	58.4%
3	58.1%
4	57.1%
5 (most deprived)	57.3%

Response rates including those completing the Natsal-3 web questionnaire

In waves 1 to 4, respondents who refused to participate in the personal interview, but who had provided an email address to interviewers, were invited (by email) to complete a shorter web version of the Natsal-3 questionnaire. Numbers were limited to those for whom we had a full name to write to, and those who were not 'double refusers' – i.e. had not refused at the initial interviewing stage and the reissue stage. As a result only 174 letters were sent. In all, 25 respondents completed the web version of the questionnaire.

Given that the number of respondents completing the Natsal-3 web questionnaire was so low, that the web version was much shorter than the CAPI/CASI Natsal-3 questionnaire, and that there could be differences in response due to differences in mode of data collection, it was decided *not* to include the web respondents in the final dataset. However, since these individuals did agree to participate in a (modified and shorter) version of Natsal-3, they could be included in the response rate calculations for the survey. With the inclusion of the additional 25 cases, response would marginally increase to 57.8% (AAPOR RR3).

Wave 1 token of appreciation experiment

As described in section 3.3.3, following the experiment developed in the dress rehearsal (see section 3.1.3), it was decided that the first wave of data collection should continue the experiment on whether the amount of the gift voucher offered to respondents affected response. This involved randomly allocating addresses within each sample point to either a £15 or a £30 gift voucher, to be given to respondents who completed the Natsal-3 interview. The amount of the gift voucher was included in the advance letter and leaflet posted to sampled addresses, and could also be mentioned by interviewers on the doorstep when recruiting participants. The results of the experiment are shown in Table 5.3. Overall, and for the core and boost 1 sample types (but not boost 2), response was higher at addresses offered the £30 gift voucher than it was at addresses which were offered the lower amount of £15. The difference in response overall (58.9% vs 56.3%) was not statistically significant.

Table 5.3 Wave 1 response rates (RR3) by gift voucher amount and sample type

	£15 gift voucher				£30 gift voucher			
	All	**Core**	**Boost 1**	**Boost 2**	**All**	**Core**	**Boost 1**	**Boost 2**
	N *(%)*	N *(%)*	N *(%)*	N *(%)*	N *(%)*	N *(%)*	N *(%)*	N *(%)*
Sampled addresses	**3375**	**1350**	**1125**	**900**	**3375**	**1350**	**1125**	**900**
Out of scope addresses:								
Vacant/ derelict	169 *(5.0%)*	80 *(5.9%)*	50 *(4.4%)*	39 *(4.3%)*	155 *(4.6%)*	75 *(5.6%)*	47 *(4.2%)*	33 *(3.7%)*
Non-residential	37 *(1.1)*	14 *(1.0%)*	10 *(0.9%)*	13 *(1.4%)*	43 *(1.3%)*	19 *(1.4%)*	9 *(0.8%)*	15 *(1.7%)*
Not traced built/ other	11 *(0.3%)*	6 *(0.4%)*	4 *(0.4%)*	1 *(0.1%)*	4 *(0.1%)*	4 *(0.3%)*	0 *(-)*	0 *(-)*
Not eligible age range	1576 *(46.7%)*	170 *(12.6%)*	735 *(65.3%)*	671 *(74.6%)*	1606 *(47.6%)*	200 *(14.8%)*	737 *(65.5%)*	669 *(74.3%)*
Total known ineligibles	**1793**	**270**	**799**	**728**	**1808**	**298**	**793**	**717**
Unknown eligibility:								
No contact	82 *(2.4%)*	57 *(4.2%)*	11 *(1.0%)*	14 *(1.6%)*	83 *(2.5%)*	47 *(3.5%)*	24 *(2.1%)*	12 *(1.3%)*
All information refused	128 *(3.8%)*	94 *(7.0%)*	22 *(2.0%)*	12 *(1.3%)*	123 *(3.6%)*	98 *(7.3%)*	17 *(1.5%)*	8 *(0.9%)*
Other	40 *(1.2%)*	29 *(2.1%)*	7 *(0.6%)*	4 *(0.4%)*	43 *(1.3%)*	31 *(2.3%)*	7 *(0.6%)*	5 *(0.6%)*
Total unknown eligibility	**250**	**180**	**40**	**30**	**226**	**176**	**48**	**25**
Estimated ineligible	76	26	27	23	72	30	33	19
Total estimated eligible addresses:	**1506**	**1054**	**299**	**153**	**1490**	**1022**	**299**	**164**
No interview:								
No contact with selected person	32 *(2.1%)*	18 *(1.7%)*	9 *(3.0%)*	5 *(3.3%)*	24 *(1.6%)*	14 *(1.4%)*	6 *(2.0%)*	4 *(2.4%)*
Refused (including proxy refusal)	369 *(24.5%)*	272 *(25.8%)*	69 *(23.1%)*	28 *(18.3%)*	350 *(23.6%)*	255 *(25.0%)*	61 *(20.4%)*	34 *(20.7%)*
Other reason	86 *(5.7%)*	57 *(5.4%)*	18 *(6.0%)*	11 *(7.2%)*	74 *(5.0%)*	38 *(3.7%)*	23 *(7.7%)*	13 *(7.9%)*
No information about address	174 *(11.5%)*	154 *(14.6%)*	13 *(4.3%)*	7 *(4.5%)*	167 *(11.2%)*	146 *(14.3%)*	15 *(5.1%)*	6 *(3.6%)*
Total unproductive	**661** ***(43.9%)***	**501** ***(47.5%)***	**109** ***(36.4%)***	**51** ***(33.3%)***	**615** ***(41.4%)***	**453** ***(44.3%)***	**105** ***(35.2%)***	**57** ***(34.7%)***
Completed interviews[1]	**848** ***(56.3%)***	**553** ***(52.5%)***	**193** ***(64.6%)***	**102** ***(66.7%)***	**874** ***(58.9%)***	**569** ***(55.7%)***	**198** ***(66.2%)***	**107** ***(65.3%)***

[1]The number of completed interviews (1722) in Table 5.3 is less than the number shown for Wave 1 in Table 5.2, as Table 5.3 is based on the achieved outcomes *before* Wave 9 – see section 3.3.1.

Comparison with Natsal-2 response

At 57.7%, the response rate for Natsal-3 is lower than the 65.4% achieved ten years earlier in Natsal-2.[4] The lower response is due to three factors:

- First, in general, surveys have been experiencing declining response rates over the past decade. For example, ONS has shown that response rates have fallen significantly on a number of major government surveys including: the Labour Force Survey (wave 1), which achieved 74% response in 2005/06 and 60% in 2008/09; the Living Cost and Food Survey, which dropped from 58% to 47% over this period; and the Family Resources Survey, which showed a similar decrease from 64% to 51% over these four years.[36]
- Second, industry standards for calculating response have been further developed since Natsal-2,[33] resulting in an alteration of the method for estimating ineligibles. In Natsal-2, individuals who were ill, away or non-English speakers were categorised as ineligible along with those who did not meet the age requirement. However, in Natsal-3 these individuals are categorised as eligible and unproductive. If these individuals were categorised as ineligible (in line with the Natsal-2 method), the response rate for Natsal-3 would be 60.5%.
- Third, and perhaps most importantly, is the extension of the age range in Natsal-3 up to 74 years, whereas Natsal-2 only included individuals aged 16-44 years. All three Natsal surveys have consistently shown that response tends to decrease with age, so the inclusion of an older age group in Natsal-3 in itself appears to account for the lower response than Natsal-2. This is best illustrated by looking at response rates for the young person's boost samples, which are much higher than for the core sample. In particular, the Natsal-3 boost 1 sample, which covers a similar age range to Natsal-2 (16-34 and 16-44 respectively), achieved a 64.8% response rate, which is only marginally lower than that achieved 10 years earlier on Natsal-2.

5.2 Response to the self-completion questionnaire

The self-completion (CASI) module was asked about half-way through the interview. Overall, 98.1% of respondents were eligible for the CASI (see section 3.2). Among the eligible, 2.2% refused to complete any of the CASI. The vast majority answered the questions directly on the laptop computer. Those who did not wish to use CASI, or could not do so due to a problem with literacy, eyesight, or inability to read English because it was not their first language, could ask the interviewer to read the questions out to them. (Unlike in Natsal-2, a paper version of the self-completion questionnaire was *not* available as an additional option for those wishing to not use the laptop.) The response to the self-completion questionnaire is shown in Table 5.4.

The proportion of respondents who completed the CASI without assistance (80.9%) was very similar to Natsal-2 (where it was 79.6%) in spite of the maximum age for Natsal-3 being considerably higher than Natsal-2 (74 and 44 years respectively). However, the likelihood of requiring assistance to complete the CASI or of having the CASI questions read out by the interviewer increased with age, such that in the oldest age group (65-74 years) over two in five respondents received such assistance.

Refusing to complete the CASI also increased with age. Reasons given by the 2.2% of eligible respondents for refusing the CASI included not wishing to use the laptop, objections to the subject matter, having insufficient time, and having problems with literacy, eyesight or reading English.

Table 5.4 Response rate to the self-completion questionnaire by age

	16-19	20-24	25-34	35-44	45-54	55-64	65-74	All
	N *(%)*	N *(%)*	N *(%)*	N *(%)*	N *(%)*	N *(%)*	N *(%)*	N *(%)*
Eligible for self-completion questionnaire	**1670**	**1975**	**3981**	**2017**	**1913**	**1783**	**1531**	**14870**
Completed CASI								
- without assistance	1554 *(93.1%)*	1798 *(91.0%)*	3484 *(87.5%)*	1726 *(85.6%)*	1458 *(76.3%)*	1234 *(69.2%)*	773 *(50.6%)*	12027 *(80.9%)*
- with assistance	94 *(5.6%)*	119 *(6.0%)*	340 *(8.5%)*	197 *(9.8%)*	301 *(15.7%)*	363 *(20.4%)*	443 *(29.0%)*	1857 *(12.5%)*
- partially completed	3 *(<0.5%)*	6 *(<0.5%)*	13 *(<0.5%)*	8 *(<0.5%)*	10 *(<0.5%)*	2 *(<0.5%)*	14 *(0.7%)*	56 *(<0.5%)*
Questions read out by interviewer	8 *(<0.5%)*	32 *(<0.5%)*	76 *(1.9%)*	55 *(2.7%)*	102 *(5.3%)*	119 *(6.7%)*	209 *(13.7%)*	601 *(4.0%)*
Refused to complete	11 *(0.7%)*	20 *(1.1%)*	68 *(1.7%)*	31 *(1.5%)*	42 *(2.2%)*	65 *(3.6%)*	92 *(6.0%)*	329 *(2.2%)*

5.3 Response to the urine sample

Eligibility criteria for providing a urine sample are described in section 4.1.1. Response rates for collecting urine samples are shown in Table 5.5 and Figure 5.1. (Samples for 16-17 year olds who were not sexually experienced are excluded from the response figures shown in Table 5.5.)

Overall, of the 8,047 eligible respondents who were asked to provide a urine sample, 4828 (60.0%) provided a sample. The reasons respondents refused to provide a sample included feeling uncomfortable with the idea of giving a urine sample, a feeling that they had done enough for the study already and 'just don't want to' (provide a sample).

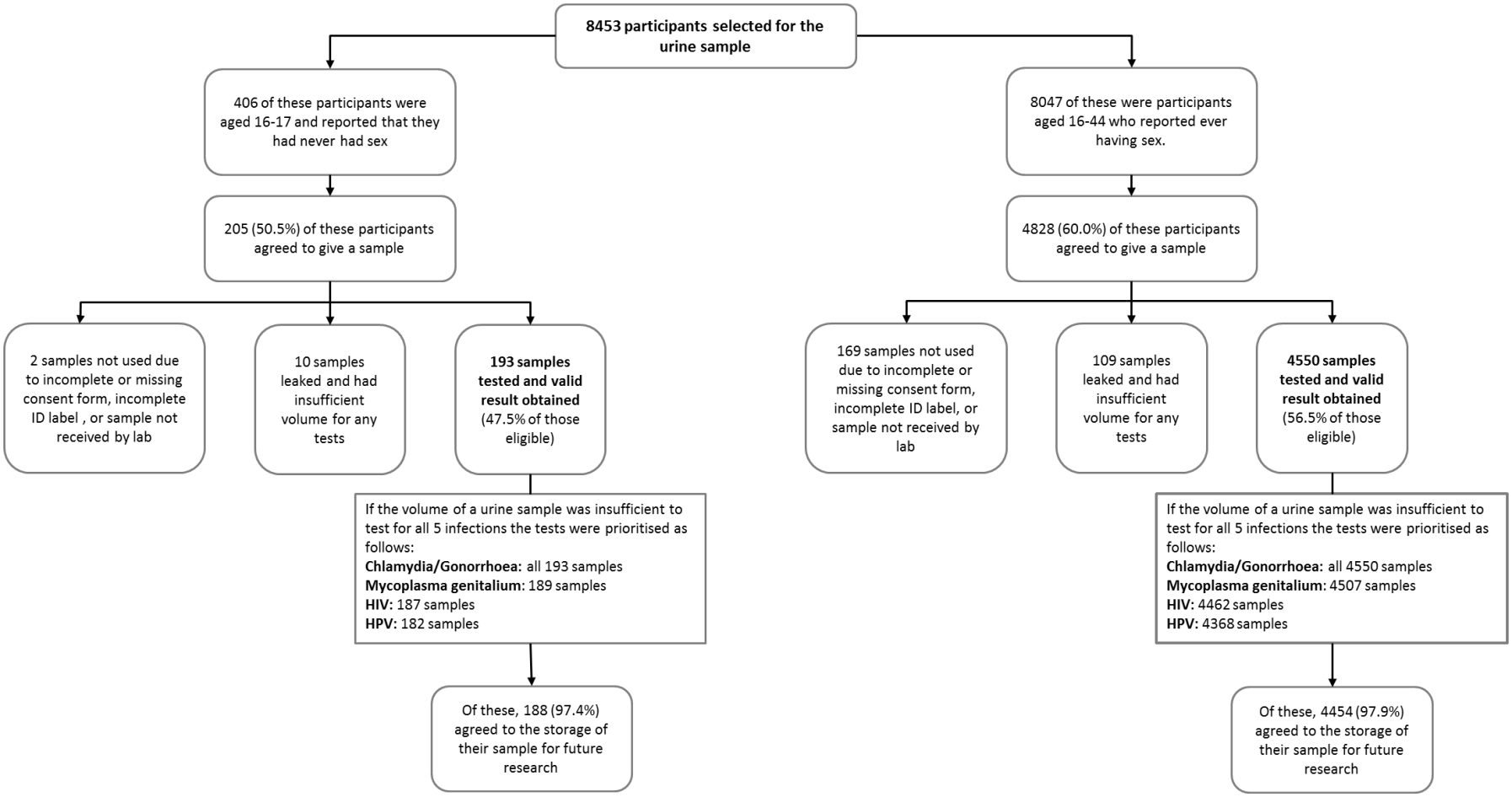
The urine samples were not always useable for a number of reasons - e.g. insufficient sample was provided, leakage before reaching the laboratory, or the sample being lost in the post. A total of 4550 urine samples were analysed and included on the dataset (56.5% of eligible respondents). Of those who provided a useable urine sample, 4454 (97.9%) agreed that it could be stored for possible future analysis.

Table 5.5 Response rate for providing a urine sample by sex

	Men	Women	All
	N *(%)*	N *(%)*	N *(%)*
Eligible for urine sample	**3285**	**4762**	**8047***
Provided urine sample	1983 *(60.4%)*	2845 *(59.7%)*	4828 *(60.0%)*
Useable samples	1885 *(57.4%)*	2665 *(56.0%)*	4550 *(56.5%)*
Agreed to sample storage *(% of useable samples)*	1855 *(98.4%)*	2602 *(97.6%)*	4454 *(97.9%)*

* Includes 4 samples collected in error from respondents (with their consent) who were not selected to provide a urine sample but nonetheless provided a useable sample which was kept in the dataset.

Figure 5.1: Response to the urine sample



5.4 Response to the saliva sample

As described in section 4.2.1, a random selection of respondents aged 18-74 were asked to provide a saliva sample. Response rates for collecting saliva samples are shown in Table 5.6 and Figure 5.2.

Of the 9170 respondents eligible to provide a sample, 71.0% agreed to do so. The reasons respondents refused to provide a sample included the procedure being too much hassle, concerns about linking to DNA, concerns about confidentiality, and a feeling that they had done enough for the study already.

Because saliva samples had to be provided first thing in the morning, this was carried out after the interviewer had left, with the result that some respondents who agreed to provide the sample did not in fact do so. In all, 4591 saliva samples were received in the laboratory, which is 70.5% of those who agreed to provide a sample (and 50.1% of eligible respondents). About 10% of these samples were excluded for various reasons – e.g. insufficient sample, delay in the post – so that 4128 samples were anlaysed and included on the dataset (45.0% of eligible respondents).

While agreeing to provide a saliva sample did not vary much by age group, there was a marked age gradient in those who actually provided useable samples: whereas only 34.0% of eligible respondents aged 18-24 provided useable samples, this percentage increased in each age group to 57.0% of eligible respondents aged 65-74.

Of those who provided a useable sample, nearly all (97.6%) agreed that the sample could be stored for possible future analysis.

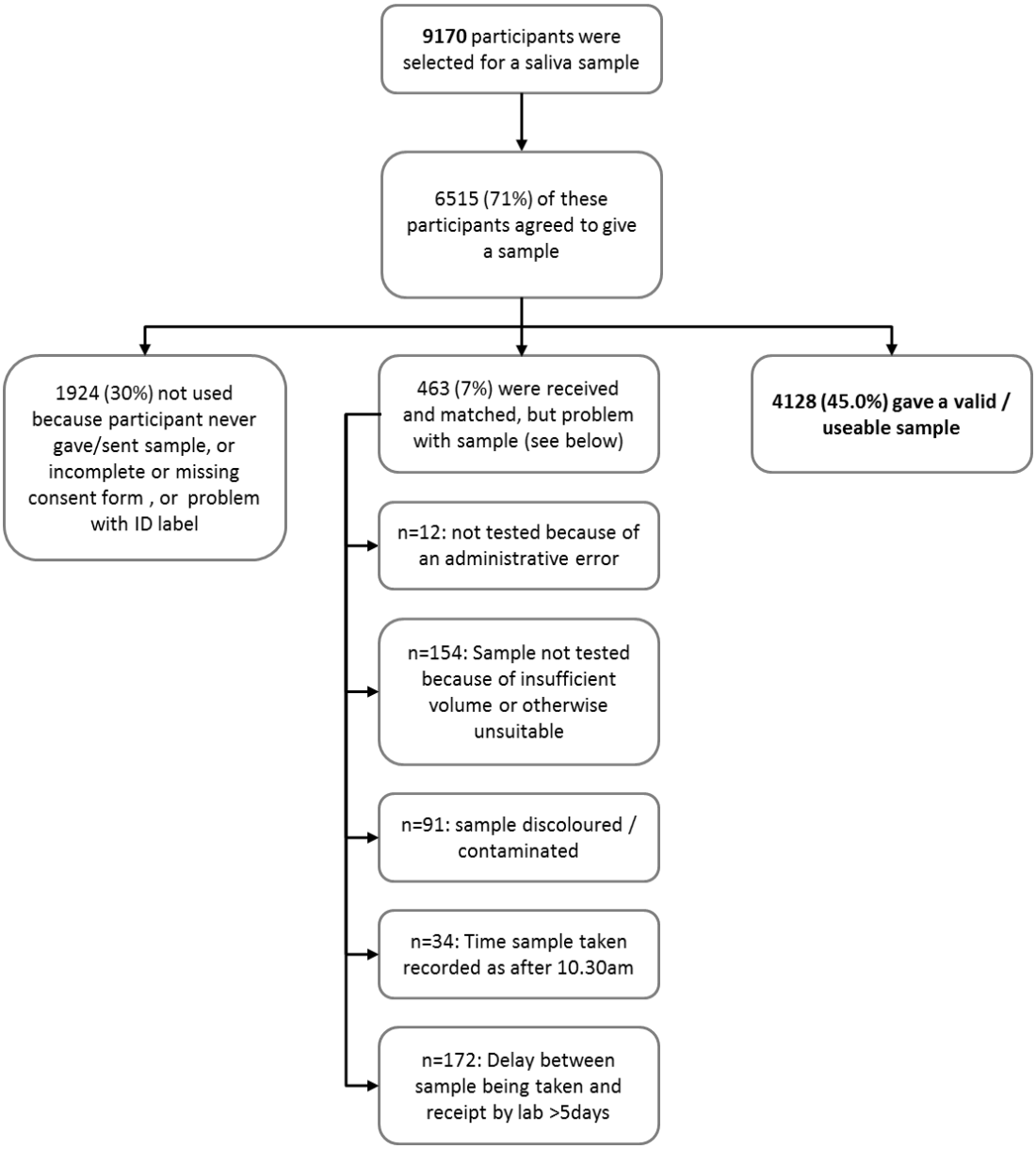
Table 5.6 Response rate for providing a saliva sample by age

	18-24	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**	**All**
	N *(%)*	N *(%)*	N *(%)*	N *(%)*	N *(%)*	N *(%)*	N *(%)*
Eligible for saliva sample	**1277**	**1789**	**1628**	**1602**	**1522**	**1352**	**9170**
Agreed to provide saliva sample	949 *(74.3%)*	1278 *(71.4%)*	1097 *(67.4%)*	1160 *(72.4%)*	1092 *(71.7%)*	939 *(69.5%)*	6515 *(71.0%)*
Provided saliva sample [1]	497 *(38.9%)*	673 *(42.6%)*	756 *(46.4%)*	836 *(52.2%)*	905 *(59.5%)*	834 *(61.7%)*	4591 *(50.1%)*
Useable saliva samples[2]	434 *(34.0%)*	690 *(38.6%)*	669 *(41.1%)*	756 *(47.2%)*	809 *(53.2%)*	770 *(57.0%)*	4128 *(45.0%)*
Agreed to sample storage *(% of useable samples)*	421 *(97.0%)*	671 *(97.2%)*	660 *(98.7%)*	732 *(96.8%)*	793 *(98.0%)*	752 *(97.7%)*	4029 *(97.6%)*

[1] Sample received by the lab and matched to survey data, and completed consent form returned.

[2] Samples were excluded from analysis for the following reasons: not tested due to error (n=12); insufficient volume (n=154); sample discoloured/bloody (n=91); sample recorded as taken after 10.30am (n=34); period between sample being taken and received by the laboratory more than 5 days (n=172).

Figure 5.2: Response to the saliva sample



5.5 Presence of others during the interview

In the briefing sessions, interviewers were reminded that the Natsal-3 interviews should be done in private, partly to reduce any embarrassment the respondent may feel, but also to encourage honest reporting of sensitive behaviours. It is not always possible, however, to arrange a completely private interview, so interviewers were asked to record if anyone else was present at any time during the interview. Table 5.7 shows the responses to this question.

Nearly three in four interviews were completed without any other person present or passing through the room where the interview was conducted. The interviewer and respondent were alone in the home in 43.3% of cases, which is less than the 45.7% of cases in Natsal-2. However, even when someone else was at home, the interviewer was able to ensure that there was no-one else present or passing through the room in 30.8% of cases, which is higher than the 27.9% of cases in Natsal-2. Where someone else was present or passing through the room during the interview, this was most likely to be a husband/wife/ partner (14.5% of all interviews).

Although one in four interviews were conducted when another adult or child was present in the house or room, in only 148 households (1.0%) was it coded by interviewers that the self-completion questionnaire was seen by another person in the household, and in another 198 households (1.3%) interviewers said that the self-completion module was discussed with another household member.

Table 5.7 Whether anyone else was present during the interview by age

	16-19	**20-24**	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**	**All**
	N *(%)*	N *(%)*	N *(%)*	N *(%)*	N *(%)*	N *(%)*	N *(%)*	N *(%)*
Total sample	**1836**	**2032**	**4009**	**2017**	**1912**	**1798**	**1539**	**15143***
No-one else present in home	432 *(23.5%)*	703 *(34.6%)*	1694 *(42.3%)*	870 *(43.1%)*	971 *(50.8%)*	1036 *(57.6%)*	845 *(54.9%)*	6551 *(43.3%)*
Other person at home, but no-one else present in interview room	877 *(47.8%)*	701 *(34.5%)*	1077 *(26.9%)*	625 *(31.0%)*	551 *(28.8%)*	454 *(25.3%)*	377 *(24.5%)*	4662 *(30.8%)*
Others present/ passing through:								
- husband/wife/ partner	132 *(7.2%)*	277 *(13.6%)*	698 *(17.4%)*	313 *(15.5%)*	263 *(13.8%)*	237 *(13.2%)*	273 *(17.7%)*	2193 *(14.5%)*
- parents	398 *(21.7%)*	190 *(9.4%)*	148 *(3.7%)*	27 *(1.3%)*	21 *(1.1%)*	22 *(1.2%)*	5 *(<0.5%)*	811 *(5.4%)*
- child(ren) aged 0-5	63 *(3.4%)*	229 *(11.3%)*	693 *(17.3%)*	201 *(10.0%)*	48 *(2.5%)*	17 *(0.9%)*	6 *(<0.5%)*	1257 *(8.3%)*
- child(ren) aged 6-15	90 *(4.9%)*	42 *(2.1%)*	224 *(5.6%)*	179 *(8.9%)*	97 *(5.1%)*	13 *(0.7%)*	6 *(<0.5%)*	651 *(4.3%)*
- young adult aged 16-21	101 *(5.5%)*	96 *(4.7%)*	34 *(0.8%)*	48 *(2.4%)*	69 *(3.6%)*	18 *(1.0%)*	3 *(<0.5%)*	369 *(2.4%)*
- other adult aged 22 and over	88 *(4.8%)*	198 *(9.7%)*	258 *(6.4%)*	78 *(3.9%)*	97 *(5.1%)*	81 *(4.5%)*	88 *(5.7%)*	888 *(5.9%)*

*19 cases had missing data on this variable.

6 Data preparation

6.1 The CAPI and CASI programs

NatCen uses Blaise survey software (www.blaise.com) for its CAPI and CASI questionnaires. As in previous Natsal surveys, virtually all the questions used in Natsal-3 were pre-coded, so that a list of answer categories is shown to the interviewer on the screen (or to the respondent when completing the CASI module). The appropriate response is selected by entering a numeric code, which is then highlighted on the screen to help confirm that the interviewer/respondent has selected the correct value. In this situation, the computer assigns a single variable to the question. For questions which allow more than one response, the interviewer/respondent can enter a list of values separated by spaces. In this case, Blaise creates a series of variables, one for each response category available. Each of these variables has a name which appears on the laptop screen.

Blaise is very efficient at handling dates and numeric values, which are assigned a range of 'columns' for the number of digits required. 'Hard' and 'soft' checks can be included for numeric entries, which query potentially out-of-range responses with the interviewer. With a 'hard' check, the response is definitely out-of-range and must be altered to an acceptable value before proceeding with the interview; with a 'soft' check, the interviewer/respondent must either alter the value or confirm that it is correct before proceeding.

The default position during the interview is that at any question, the interviewer (or respondent, during the CASI) can enter CTRL-K for 'don't know' or CTRL-R for 'refused'. For a small number of questions, explicit codes were included on the screen for 'don't know' or 'refused'.

Another feature of Blaise is that it allows 'textfills' within the question text to appear on the screen, which adapts the exact question wording to the context of a particular respondent. Given the importance of gender differences for questions which ask about sexual practices, this facility was particularly useful for Natsal-3 and was used throughout the CASI questionnaire, where question wordings were often slightly different for male and female respondents. The CAPI and CASI questionnaires included in Volume 2: Appendix B show many of these features of Blaise.

6.2 Editing and coding

All completed interviews underwent an office-based edit which comprised two stages: first, all open-text, or 'verbatim', answers were coded to prepare them for analysis, then each interview was checked for inconsistent answers.

With the use of CAPI and CASI, there tends to be very little missing data. Item non-response in Natsal-3 was typically below 0.5% in the CAPI and around 1%-3% in the CASI.

The verbatim 'other' answers to questions, and the occupational coding (see section 6.3) were coded in the office by a team of five coders, all of whom were fully trained in the edit program before beginning the process. The coders also examined any additional comments entered by interviewers as notes during the interview, and occasionally data were edited manually on the basis of these comments. NatCen's policy is that every coder's first batch of work must be 100% verified by staff in the Operations Unit. Coders cannot undertake more work on the project until the first batch has been checked and confirmed to be of an acceptable standard. Any concerns identified during verification are fed back to the coder. If the issues are considered to be minor,

a 10% random check is undertaken in the future. If the work is considered not to be satisfactory, the coder is required to attend a further training session before they can undertake more work on the project and their next batch of work will be 100% checked again. This process continues until either the coder reaches the required standard or is withdrawn from the project.

A second editing stage was carried out, in which the data were checked for internal consistencies between the responses to different questions. As mentioned in section 3.2, with the use of CASI it is possible to build consistency checks into the program so that respondents can rectify any problems immediately. The extent to which such consistency checks can be built into the questionnaire itself is limited both by the degree to which respondents are willing to put up with reminders that they are being inconsistent as well as by the reason for the inconsistency, of which there may be several. For example, a respondent may have reported having a higher number of sexual partners in the last year than the number reported for lifetime partners. This may be a simple keying error which could be easily corrected when drawn to the respondent's attention. Alternatively, it may be memory error, with the respondent remembering an additional partner when answering the subsequent question and such cases can also be rectified if it is pointed out during the interview.

Clearly, the capacity for respondents to correct such inconsistencies on the spot results in higher data quality. However, there are also instances when the inconsistency may have been deliberate. For example, respondents were asked, using show cards, whether they had any same-sex sexual experience during the face-to-face interview as well as at various points in the self-completion questionnaire. An inconsistency could arise if the respondent felt more confident about reporting same-sex sexual experiences in the more anonymous CASI questionnaire than in the face-to-face interview. Pointing out such an inconsistency to the respondent would not have been appropriate, nor would programming the interview in such a way to ensure this sort of inconsistency could not arise.

An automated computer edit program, based on the CAPI and CASI programs but with additional checks and editing instructions, was run to check for inconsistencies during the data cleaning stage. Researchers then examined these cases to attempt to resolve any inconsistencies found. In some instances, the inconsistencies could be resolved, although in many cases no action was taken except to 'flag' the inconsistency on the data file. The default approach was not to change a respondent's answer except where there was evidence (e.g. from elsewhere in the questionnaire) to do so.

In all, 137 potential inconsistencies were checked for each respondent during data cleaning. Volume 2: Appendix F shows the details of each of these checks and the corrective action taken in each case. For many of these checks, Natsal-3 researchers examined cases individually in order to make a decision about how to edit a particular inconsistency.

As well as checking for inconsistent answers, an additional stage of manual checking of the CASI data was carried out to identify possible cases where the respondent may have been trying to get through the self-completion module as quickly as possible, by just answering all questions with the first available code (i.e. answering all as '1'). A check of the first four CASI modules was carried out on all interviews and any cases with over 65% of questions answered as 1 were manually checked in full. A small number of cases (n=26) were manually edited as a result of this checking stage (e.g. by editing one or more answers to be missing).

6.3 Occupation coding

The only open-ended questions in the interview asked for respondents' (and their partners') occupations. Details were collected about current occupation, or the most recent occupation for those not currently in work. These were coded in the office using the Standard Occupational Classification (SOC2010) at three digit level. The National Statistics Socio-Economic Classification (NS-SEC) was derived from SOC2010 and employment status, and was used as a social class measure.[37]

The SOC2010 classification has the following categories:

Code	Description
1	Managers, directors and senior officials
2	Professional occupations
3	Associate professional and technical occupations
4	Administrative and secretarial occupations
5	Skilled trades occupations
6	Caring, leisure and other service occupations
7	Sales and customer service occupations
8	Process, plant and machine operatives
9	Elementary occupations

The National Statistics Socio-Economic Classification (NS-SEC) was derived from SOC2010 and employment status, and was used as a social class measure. NS-SEC has the following categories:

Code	Description
1	Managerial and professional occupations
2	Intermediate occupations
3	Small employers & own account workers
4	Lower supervisory & technical occupations
5	Semi-routine and routine occupations
6	Never worked or long-term unemployed
7	No job of 10+ hours/week or not worked in last 1- years
8	Student in full-time education

In order to estimate social class when respondents were growing up, interviewers asked about the type of work their father did (or mother, if the father was not in the household) and whether he had any managerial or supervisory responsibilities, when respondents were aged 14. These answers were used to derive a proxy measure of social class.

7 Weighting

7.1 Interview sample weights

7.1.1 Selection probability weighting

As described in section 2, respondents selected for inclusion in the survey did not have equal chances of selection for three reasons: first, younger respondents aged 16-34 had a greater chance of selection than respondents aged 35-74 as the sample aimed to over-represent younger respondents; second, at addresses containing more than one household, one was selected at random; third, in households with two or more adults in the eligible age range (16-74 in core households, 16-34 in boost 1 and 16-29 in boost 2), only one was randomly selected.

Therefore, before the data can be used as a representative sample of the general population of Britain, the imbalances created by different probabilities of selection must be accounted for, otherwise the sample would over-represent young people aged 16-34 and those living alone. This was done by applying two sets of weights: a 'household weight' to correct for the selection of one household at a multi-household address; and a 'person weight' to correct for the selection of one adult at an address (which also corrects for the unequal probabilities of selection by age). These adjustments were made by applying weights which were inversely proportional to the selection probabilities for the number of households and adults within the eligible age range at each selected address.

The household selection weight was equal to the number of households found at the address. This weight was trimmed at 3, so that respondents at addresses containing 3 or more households received a household selection weight of 3.

The person selection weight depended on the age of the respondent, as the probability of selection varied for the three sample types: the core sample of adults aged 16-74, the 16-29 age boost, and the 16-34 age boost. Adults aged 16-29 could be selected in all three sample types; adults aged 30-34 could be selected in the core or the 16-34 boost samples; and adults aged 35-74 could only be selected in the core sample. The probability of selection for any sample type depended on the proportion of total addresses (across all waves) that were selected for that type and the number of people in the household in the relevant age group. The selection probabilities for each age group were:

Aged 16-29

Pr(Selection) = Pr(Core) + Pr(Boost16-34) + Pr(Boost16-29) =

$$\frac{24924}{59412} * \frac{1}{A1 + A2 + A3} + \frac{18537}{59412} * \frac{1}{A1 + A2} + \frac{15951}{59412} * \frac{1}{A1}$$

Aged 30-34

Pr(Selection) = Pr(Core) + Pr(Boost16-34) =

$$\frac{24924}{59412} * \frac{1}{A1 + A2 + A3} + \frac{18537}{59412} * \frac{1}{A1 + A2}$$

Aged 35-74

Pr(Selection) = Pr(Core) =

$$\frac{24924}{59412} * \frac{1}{A1 + A2 + A3}$$

A1= the number of people aged 16-29 in the household
A2 = the number of people aged 30-34 in the household
A3 = the number of people aged 35-74 in the household

The person selection weights were then calculated as the reciprocal of these selection probabilities.

The household selection weight and the person selection weight were then multiplied together to create the final selection weight (Selwt).

7.1.2 Non-response weighting

After application of the selection probability weights (Selwt), the profile of Natsal-3 respondents was compared with 2011 census data for Britain by age, sex and government office region (Table 7.1, section [a] compared with section [c]). Any differences are most likely to be due to differential non-response, but there could also be a small amount of sampling error (i.e. small differences between the profile of those at the selected addresses after correction for unequal selection probabilities and the profile of the general population).

In order to adjust for the differences in the profile by age, sex and region, a post-stratification weight was calculated as follows:

$$Post\_wt = \frac{proportion\ of\ eligible\ \ population\ within\ age/sex/region\ category}{proportion\ of\ sample\ (weighted\ by\ Selwt)\ within\ age/sex/region\ category}.$$

Due to small numbers, and in order to avoid extreme weights, three pairs of cells (males 35-39 in Inner and Outer London, males 40-44 in Inner and Outer London and males 55-64 in Inner and Outer London) were merged for the purposes of creating this weight.

The final weight (*Total_wt*) was calculated as the product of the selection weight and the post-stratification weight:

*Total_wt = Selwt * Post_wt*

Total_wt was scaled to have a mean of 1. After inspection of the scaled weights, one extreme outlying weight (8.38) was trimmed back to the next highest weight (6.84) and the weights were re-scaled to have a mean of 1. This gives a weighted sample size equal to the unweighted sample size of 15,162.

7.1.3 Representativeness of Natsal-3 data

The distributions of age and region by sex after applying the final weight is shown in the middle columns (section [b]) in Table 7.1. After post-stratification, the sample matches the population closely on age, sex and region. There are two areas with minor discrepancies: the first (men 35-39) is due to trimming the final weight; the second is due to merging the cells for men in Inner and Outer London.

Table 7.1 Comparison of Natsal-3 weighted sample with 2011 census population figures

	[a] Natsal-3 sample after selection weighting			[b] Natsal-3 sample after final weighting			[c] 2011 Census figures		
Age	**Men**	**Women**	**All**	**Men**	**Women**	**All**	**Men**	**Women**	**All**
	%	%	%	%	%	%	%	%	%
16-19	7.8	6.6	7.1	7.1	6.7	6.9	7.1	6.7	6.9
20-24	8.9	8.0	8.4	9.4	9.1	9.3	9.4	9.1	9.3
25-29	8.3	8.5	8.4	9.3	9.2	9.3	9.3	9.2	9.3
30-34	7.9	8.4	8.2	9.0	8.8	8.9	9.0	8.8	8.9
35-39	8.5	9.8	9.3	9.0	9.0	9.0	9.1	9.0	9.0
40-44	9.6	10.7	10.2	9.9	10.0	10.0	9.9	10.0	10.0
45-54	18.7	20.4	19.7	18.8	18.9	18.8	18.8	18.9	18.8
55-64	16.8	16.2	16.5	15.9	16.1	16.0	15.9	16.1	16.0
65-74	13.5	11.4	12.3	11.5	12.2	11.8	11.5	12.2	11.8
Row %	42.6	57.4		49.5	50.5		49.5	50.5	
GOR	**Men**	**Women**	**All**	**Men**	**Women**	**All**	**Men**	**Women**	**All**
	%	%	%	%	%	%	%	%	%
North East	5.1	4.9	5.0	4.2	4.3	4.3	4.2	4.3	4.3
North West	12.9	13.1	13.0	11.5	11.5	11.5	11.5	11.5	11.5
Yorkshire & the Humber	8.2	8.0	8.1	8.6	8.6	8.6	8.6	8.6	8.6
East Midlands	8.6	7.9	8.2	7.4	7.4	7.4	7.4	7.4	7.4
West Midlands	8.6	9.1	8.9	9.1	9.0	9.0	9.1	9.0	9.0
South West	8.6	8.2	8.4	8.6	8.5	8.6	8.6	8.5	8.6
East	10.7	10.4	10.5	9.4	9.4	9.4	9.4	9.4	9.4
Inner London	3.6	3.7	3.7	4.9	5.5	5.2	5.6	5.5	5.6
Outer London	7.1	7.0	7.1	8.6	8.1	8.3	8.0	8.1	8.0
South East	13.1	14.2	13.7	13.9	13.9	13.9	13.9	13.9	13.9
Wales	5.5	5.1	5.3	5.0	5.0	5.0	5.0	5.0	5.0
Scotland	7.9	8.4	8.1	8.7	8.9	8.8	8.7	8.9	8.8

Table 7.2 compares Natsal-3 distributions with three other variables from the 2011 census (in England and Wales only and limited to ages 16-74). Generally, there is a close match between Natsal-3 and the census on these three variables. Looking at marital status, it appears that Natsal-3 over-represents respondents who are living with their marital spouse or civil partner, and under-represents men and women who are single. This discrepancy could be partly explained by the census figures including individuals aged 16-74 living in institutions, such as care homes, and who are much more likely to be single. In terms of ethnicity, there is a slight under-representation of Asian men and women in Natsal-3, while for self-reported general health, it seems that Natsal-3 respondents (especially men) are more likely to classify themselves in 'fair' health.

Table 7.2: Natsal-3 distributions compared with 2011 census population figures

England & Wales, ages 16-74	Natsal-3			Census 2011		
	Men	**Women**	**All**	**Men**	**Women**	**All**
Marital status[1]	%	%	%	%	%	%
Single, never married	38.1	32.3	35.1	41.0	34.3	37.7
Married, living with spouse	50.0	49.4	49.7	46.7	47.1	46.9
Separated/divorced/widowed	11.6	17.7	14.8	12.0	18.3	15.2
Civil partnership, living with partner	0.4	0.6	0.5	0.3	0.2	0.2
Ethnic group[1]						
White	86.7	86.9	86.8	86.6	86.8	86.7
Mixed	1.7	2.0	1.8	1.6	1.6	1.6
Asian	7.0	5.7	6.4	7.6	7.4	7.5
Black	3.4	4.0	3.7	3.0	3.3	3.2
Other	1.2	1.4	1.3	1.2	0.9	1.0
Self-reported general health[1]						
Very good/good	81.1	80.7	80.9	82.1	81.3	81.7
Fair	14.9	14.3	14.6	12.5	13.3	12.9
Bad/very bad	3.9	4.9	4.4	5.4	5.3	5.4

[1] Census data 2011, all usual residents aged 16-74 in England & Wales only.

Another check on the representativeness of the Natsal-3 data involved comparing the birth rate for Natsal-3 respondents with birth rate data in Britain. Using figures provided by the Office for National Statistics (ONS) on the number of live births registered in Britain in 2011, we calculated the birth rate for the population in Britain aged 16-74 as 17.32 per 1000 persons. Calculated on the same basis, the birth rate for Natsal-3 respondents works out as 18.98 per 1000 persons aged 16-74, with 95% confidence intervals (17.11 to 21.04 per thousand) that overlap the population birth rate.

7.2 Urine sample weights

7.2.1 Urine sample selection probability weighting

Since not all respondents were asked to provide a urine sample, the calculation of an additional "urine weight" - i.e. *in addition to* the main survey weight – is necessary in order to provide unbiased estimates of the urine sample test results. Those eligible for urine sample collection included:

- all respondents aged 16-24 who reported ever having sex
- all men aged 25-44 who reported having sex with another man in the last 5 years
- a random selection of 85% of respondents aged 25-44 who reported ever having sex.

To avoid risk of disclosure of sexual activity within the household, urine samples were requested from all respondents aged 16-17, including those who did not report ever having sex (see section 4.2.3). Results for respondents who never had sex are not used in the main analysis and therefore were not given a urine sample weight.

The probabilities of selection for these groups therefore are:

- 0.85 for women aged 25-44 who reported ever having sex
- 0.85 for men aged 25-44 who reported ever having sex, but who did not report having sex with another man in the last 5 years
- 1 for respondents aged 16-24 who reported ever having sex
- 1 for men aged 25-44 who reported having sex with another man in the last 5 years

The urine selection probability weights (*urine_selwt*) were then calculated as the reciprocal of these selection probabilities.

7.2.2 Urine sample non-response weighting

As described in section 5.3, 8047 Natsal-3 respondents were eligible for the urine study and 4550 (56.5%) provided a valid sample. With the aim of minimising any potential bias in the achieved urine sample data arising from differential response, a non-response weight was calculated specifically for those respondents providing a useable urine sample.

Response to the urine sample was modelled, separately for men and women, using logistic regression (weighted by *total_wt * urine_selwt* and scaled within gender to have a mean of 1), with the dependent variable indicating whether or not a useable urine sample was obtained. Using data available for both responders and non-responders to the urine sample, age was included in the model along with a wide range of other demographic and behavioural indicators. The variables found to be related to urine response were:

- for men: ethnicity, highest educational qualification, marital status, number of opposite- or same-sex partners without a condom in the last year, same-sex experience (ever), attended a sexual health clinic (ever), overlap in partners in the last 5 years, injected non-prescribed drugs (ever).
- for women: region, total number of lifetime opposite- and same-sex partners, heterosexual anal sex (last 5 years), blood test for HIV (ever), same sex experience (ever), other people were present in the household during the Natsal-3 interview, children aged 6-15 were present or passing through during the interview.

Both models were then run again including all variables that were significant (or borderline non-significant) in either the men's or women's model. The coefficients for the final model are shown in Tables 7.3 (men) and 7.4 (women).

Table 7.3 Urine sample: model of non-response for men

	Coefficient (log odds)	Odds ratio[1]	p
Age			**<0.01**
16-17	0	1	-
18-19	-0.29	0.75	0.22
20-24	-0.13	0.88	0.52
25-34	-0.58	0.56	<0.01
35-44	-0.61	0.54	<0.01
Government Office Region			**0.28**
North East	0	1	-
North West	0.21	1.23	0.32
Yorkshire & the Humber	0.21	1.24	0.32
East Midlands	-0.05	0.95	0.82
West Midlands	0.06	1.06	0.79
South West	-0.24	0.79	0.27
East	-0.04	0.96	0.85
London	-0.01	0.99	0.95
South East	0.07	1.07	0.75
Wales	0.22	1.25	0.36
Scotland	0.10	1.11	0.64

Ethnicity			**<0.01**
White	0	1	-
Mixed, Chinese, Other	0.06	1.07	0.74
Asian/Asian British	-0.72	0.49	<0.01
Black/Black British	-0.55	0.58	<0.01
Highest educational qualification			**0.04**
Degree	0	1	-
Higher education, <degree but A-level/equivalent	-0.23	0.80	0.02
GCSE, O-level or equivalent, Foreign or other	-0.15	0.86	0.13
None	-0.41	0.67	0.02
Marital status			**<0.01**
Single & never married	0	1	-
Married/civil partner & living with partner	0.03	1.03	0.74
Separated, divorced, widowed	0.68	1.97	<0.01
Total number of opposite- or same-sex partners in lifetime			**0.88**
1-2	0	1	-
3-4	-0.05	0.95	0.67
5-6	-0.10	0.91	0.46
7-8	0.14	1.15	0.34
9-10	-0.06	0.95	0.73
11-15	-0.01	0.99	0.96
16-30	-0.08	0.92	0.55
31+	-0.01	0.99	0.93
Number of opposite- or same-sex partners without a condom in the last year			**<0.01**
0	0	1	-
1	0.32	1.38	<0.01
2+	0.20	1.23	0.16
Heterosexual anal sex in the last 5 years			**0.94**
No	0	1	-
Yes	-0.01	0.99	0.94
Blood test for HIV (ever)			**0.60**
Yes	0	1	-
No	0.10	1.11	0.31
Maybe/not sure	0.07	1.08	0.68
Same sex experience with genital contact (ever)			**<0.01**
Yes	0	1	-
No	-0.50	0.61	<0.01

Attended a sexual health clinic (ever)			**<0.01**
Yes	0	1	-
No	-0.33	0.72	<0.01
Overlap between partners in last 5 years			**<0.01**
Yes	0	1	-
No	-0.38	0.68	<0.01
Other people in the household during the interview			**0.84**
Yes	0	1	-
No	0.02	1.02	0.84
Children aged 6-15 years present/passing through during the interview			**0.73**
Present	0	1	-
Not present	-0.06	0.94	0.73
Ever injected non-prescribed drugs			**0.04**
Yes	0	1	-
No	-0.63	0.53	0.04
Intercept	**2.21**	**9.13**	**<0.01**

[1]The dependent variable in the logistic regression model was: 1=provided a urine sample; 0=did not provide a urine sample. All respondents who were eligible for the urine study were included in the model.

Table 7.4 Urine sample: model of non-response for women

	Coefficient (log odds)	Odds ratio[1]	p
Age			**0.11**
16-17	0	1	-
18-19	-0.26	0.77	0.20
20-24	-0.35	0.70	0.04
25-34	-0.37	0.69	0.03
35-44	-0.48	0.62	<0.01
Government Office Region			**0.01**
North East	0	1	-
North West	-0.08	0.92	0.63
Yorkshire & the Humber	0.04	1.04	0.82
East Midlands	0.20	1.22	0.29
West Midlands	-0.25	0.78	0.16
South West	-0.16	0.85	0.37
East	-0.16	0.85	0.36
London	-0.33	0.72	0.05
South East	-0.19	0.82	0.25
Wales	0.00	1.00	0.98
Scotland	0.04	1.04	0.83

Ethnicity			**0.08**
White	0	1	-
Mixed, Chinese, Other	-0.10	0.90	0.51
Asian/Asian British	-0.32	0.72	0.02
Black/Black British	-0.23	0.79	0.16
Highest educational qualification			**0.07**
Degree	0	1	-
Higher education, <degree but A-level/equivalent	0.17	1.19	0.03
GCSE, O-level or equivalent, Foreign or other	0.00	1.00	0.99
None	-0.08	0.93	0.58
Marital status			**0.25**
Single & never married	0	1	-
Married/civil partner & living with partner	0.12	1.13	0.14
Separated, divorced, widowed	0.15	1.17	0.21
Total number of opposite- or same-sex partners in lifetime			**0.06**
1-2	0	1	-
3-4	0.10	1.10	0.30
5-6	0.19	1.21	0.06
7-8	0.23	1.26	0.05
9-10	0.40	1.49	<0.01
11-15	0.29	1.34	0.02
16-30	0.27	1.31	0.04
31+	0.21	1.23	0.31
Number of opposite- or same-sex partners without a condom in the last year			**0.20**
0	0	1	-
1	0.03	1.03	0.67
2+	0.24	1.27	0.08
Heterosexual anal sex in the last 5 years			**<0.01**
No	0	1	-
Yes	0.29	1.33	<0.01
Blood test for HIV (ever)			**<0.01**
Yes	0	1	-
No	-0.18	0.84	0.01
Maybe/not sure	-0.34	0.71	<0.01
Same sex experience with genital contact (ever)			**<0.01**
Yes	0	1	-
No	-0.46	0.63	<0.01

Attended a sexual health clinic (ever)			**0.55**
Yes	0	1	-
No	-0.04	0.96	0.55
Overlap between partners in last 5 years			**0.27**
Yes	0	1	-
No	-0.11	0.89	0.27
Other people in the household during the interview			**0.05**
Yes	0	1	-
No	0.13	1.13	0.05
Children aged 6-15 years present/passing through during the interview			**<0.01**
Present	0	1	-
Not present	0.34	1.40	<0.01
Ever injected non-prescribed drugs			**0.57**
Yes	0	1	-
No	0.27	1.31	0.57
Intercept	**0.35**	**1.42**	**0.53**

[1]The dependent variable in the logistic regression model was: 1=provided a urine sample; 0=did not provide a urine sample. All respondents who were eligible for the urine study were included in the model.

The non-response weights for the urine sample (*urine_nrwt*) were calculated, separately for men and women, as the inverse of the model-predicted probability of obtaining a useable urine sample.

The final urine weight (*urine_wt*) for the 4550 cases with a useable test result was then calculated as: *urine_wt = total_wt * urine_selwt * urine_nrwt.*

The value of *urine_wt* for two cases was much higher (15.39 and 12.39) than the maximum of the urine selection weight (9.02), so these two cases were trimmed to this maximum. After trimming, the final urine weights were scaled to have a mean of 1, so that the weighted urine sample size is the same as the achieved urine sample size. The distribution of the different components of the weighting are shown in Table 7.5 (all weights have been scaled to have a mean of 1 for comparison).

Table 7.5 Distribution of different components of urine weights after trimming

	Sample size	Minimum	Maximum	Mean	Standard deviation
Main survey weight (*total_wt*) for urine respondents only	4550	0.24	8.57	1.00	0.69
Urine selection weight (*total_wt * urine_selwt*)	4550	0.25	9.02	1.00	0.72
Urine non-response weight (*urine_nrwt*)	4550	0.62	1.79	1.00	0.19
Urine weight (*urine_wt*)	4550	0.17	9.02	1.00	0.81

The non-response weighting significantly reduced any bias in the profile of urine sample respondents. For example, 40.2% of those eligible to provide a urine sample reported heterosexual anal sex in the last 5 years. At 43.5%, respondents who provided a urine sample were more likely to report heterosexual anal sex, i.e. a bias of 3.3%. After applying the non-response weights, this becomes 40.3%, thereby reducing the bias to only 0.1% of the weighted sample of urine respondents who report this behaviour.

7.3 Saliva sample weights

7.3.1 Saliva sample selection probability weighting

As described in section 4.2.1, a random sub-sample of respondents was asked to provide a saliva sample, which means that it is necessary to calculate an additional "saliva weight" in order to provide unbiased estimates. Those eligible for saliva sample collection included all adults aged 18-74 (except for those who worked night shifts), with the aim of over- sampling those aged 35-74. As explained earlier, the sampling fractions changed during the course of fieldwork, so the probabilities of selection were as follows:

- 0.30 for respondents aged 18-34 (waves 1 – 3)
- 0.36 for respondents aged 18-34 (waves 4 – 5)
- 0.75 for respondents aged 18-34 (waves 6 – 8)
- 0.66 for respondents aged 35-74 (waves 1 – 3)
- 1.00 for respondents aged 35-74 (waves 4 – 8)

The saliva selection probability weights (*saliva_selwt*) were calculated as the reciprocal of these selection probabilities.

7.3.2 Saliva sample non-response weighting

Overall, 9170 respondents were eligible to provide a saliva sample and 4128 (45.0%) provided useable samples (see section 5.4). In order to reduce possible bias in the achieved saliva sample data due to differential non-response, a non-response weight was calculated specifically for the saliva sample test results. This followed a similar procedure to that described for the urine sample test results.

With the dependent variable indicating whether a useable saliva sample was provided, logistic regression models were run separately for men and women (weighted by *total_wt* * *saliva_selwt* and scaled within gender to have a mean of 1). Age and region were included in the models along with a range of other demographic and behavioural variables. Those found to be related to saliva response were:

- for men: ethnicity, highest educational qualification, total number of lifetime opposite- or same-sex partners, self-reported general health, longstanding illness or disability, sexual attraction scale, depressive symptoms, sexual function, children aged 0-5 present or passing through the room during the interview.
- for women: NS-SEC, ethnicity, highest educational qualification, self-reported general health, longstanding illness or disability, BMI, currently using a hormonal method of family planning, depressive symptoms, children aged 6-15 present or passing through the room during the interview.

Both models were then run again including all variables that were significant in either the men's or women's model. In the final model for women, menstrual cycle was borderline non-significant (p=0.06), but was considered an important covariate and thus was included in the final model. The coefficients for the final model are shown in Tables 7.6 (men) and 7.7 (women).

Table 7.6 Saliva sample: model of non-response for men

	Coefficient (log odds)	Odds ratio[1]	p
Age			**<0.01**
18-19	0	1	-
20-24	-0.18	0.84	0.43
25-34	0.13	1.14	0.58
35-44	0.22	1.24	0.35
45-54	0.49	1.64	0.04
55-64	0.74	2.10	<0.01
65-74	1.07	2.90	<0.01
Government Office Region			**0.06**
North East	0	1	-
North West	-0.14	0.87	0.48
Yorkshire & the Humber	-0.01	0.99	0.98
East Midlands	-0.02	0.98	0.91
West Midlands	0.14	1.16	0.49
South West	-0.16	0.85	0.44
East	0.13	1.13	0.54
London	-0.02	0.98	0.90
South East	0.12	1.12	0.55
Wales	-0.21	0.81	0.37
Scotland	0.33	1.39	0.11
NS-SEC			**0.54**
Managers & professional	0	1	-
Intermediate	-0.02	0.98	0.82
Semi-routine/routine	-0.11	0.90	0.27
No job in last 10 years	0.12	1.12	0.46
Student in full-time education	-0.13	0.88	0.51
Ethnicity			**<0.01**
White	0	1	-
Mixed, Chinese, Other	0.17	1.18	0.43
Asian/Asian British	-0.38	0.68	0.03
Black/Black British	-1.18	0.31	<0.01
Highest educational qualification			**0.05**
Degree	0	1	-
Higher education, <degree but A-level/equivalent	-0.06	0.94	0.52
GCSE, O-level or equivalent, Foreign or other	0.07	1.07	0.53
None	-0.28	0.76	0.05

Total number of opposite- or same-sex partners in lifetime			**<0.01**
0	0	1	-
1-2	-0.28	0.76	0.27
3-4	-0.45	0.64	0.09
5-6	-0.62	0.54	0.02
7-8	-0.50	0.61	0.08
9-10	-0.42	0.66	0.14
11-15	-0.64	0.53	0.03
16-30	-0.77	0.46	<0.01
31+	-0.41	0.66	0.15
Children aged 0-5 years present/ passing through during interview			**<0.01**
Present throughout	0	1	-
Present some of the time	-0.39	0.68	0.24
Not present	0.29	1.33	0.26
Children aged 6-15 years present/passing through during the interview			**0.90**
Present some/all of the time	0	1	-
Not present	-0.03	0.97	0.90
Self-reported health			**<0.01**
Very good	0	1	-
Good	-0.26	0.77	<0.01
Fair	-0.41	0.66	<0.01
Bad or very bad	-0.72	0.49	<0.01
Longstanding illness/ disability			**0.04**
Yes	0	1	-
No	-0.18	0.83	0.04
Body Mass Index			**0.07**
Underweight (BMI <18.5 kg/m^2)	0	1	-
Normal (BMI 18.5 – 25 kg/m^2)	0.91	2.48	<0.01
Overweight (BMI over 25 – 30 kg/m^2)	0.92	2.52	<0.01
Obese I (BMI over 30 – 35 kg/m^2)	1.08	2.94	<0.01
Obese II (BMI over 35 – 40 kg/m^2)	0.96	2.61	0.02
Obese III (BMI >40 kg/m^2)	1.18	3.25	0.01
Not answered	0.66	1.93	0.10
Currently using a hormonal method of family planning			**0.18**
No	0	1	-
Yes	0.13	1.14	0.18
Sexual attraction			**<0.01**
Opposite sex only	0	1	-
More often opposite sex, and at least once same sex	0.72	2.05	<0.01
All other categories	0.37	1.45	0.07

Depressive symptoms			**0.04**
No	0	1	-
Yes	0.10	1.11	0.42
Not answered	-0.81	0.45	0.02
Sexual function			**<0.01**
Normal function	0	1	-
Low function	0.27	1.31	<0.01
Not sexually active	0.16	1.17	0.17
Not available	-0.78	0.46	<0.01
Intercept	**-0.97**	**0.38**	**0.10**

[1]The dependent variable in the logistic regression model was: 1=provided a saliva sample; 0=did not provide a saliva sample. All respondents who were eligible for the saliva study were included in the model.

Table 7.7 Saliva sample: model of non-response for women

	Coefficient (log odds)	Odds ratio[1]	p
Age			**<0.01**
18-19	0	1	-
20-24	-0.07	0.93	0.70
25-34	0.15	1.16	0.45
35-44	0.28	1.33	0.16
45-54	0.53	1.70	<0.01
55-64	0.76	2.15	<0.01
65-74	1.08	2.95	<0.01
Government Office Region			**0.28**
North East	0	1	-
North West	0.04	1.04	0.82
Yorkshire & the Humber	0.15	1.17	0.36
East Midlands	0.06	1.07	0.71
West Midlands	0.10	1.11	0.53
South West	0.29	1.34	0.08
East	0.15	1.16	0.37
London	0.00	1.00	0.98
South East	-0.02	0.98	0.92
Wales	0.29	1.34	0.11
Scotland	0.11	1.12	0.50
NS-SEC			**<0.01**
Managers & professional	0	1	-
Intermediate	0.02	1.02	0.81
Semi-routine/routine	-0.19	0.83	0.03
No job in last 10 years	-0.34	0.71	<0.01
Student in full-time education	-0.14	0.87	0.35
Ethnicity			**<0.01**
White	0	1	-
Mixed, Chinese, Other	-0.40	0.67	0.02
Asian/Asian British	-0.46	0.63	<0.01
Black/Black British	-0.63	0.53	<0.01

Highest educational qualification			**<0.01**
Degree	0	1	-
Higher education, <degree but A-level/equivalent	0.07	1.07	0.41
GCSE, O-level or equivalent, Foreign or other	-0.17	0.84	0.04
None	-0.33	0.72	<0.01
Total number of opposite- or same-sex partners in lifetime			**0.12**
0	0	1	-
1-2	-0.19	0.83	0.39
3-4	-0.28	0.76	0.22
5-6	-0.12	0.89	0.61
7-8	-0.22	0.80	0.37
9-10	-0.14	0.87	0.57
11-15	0.08	1.08	0.74
16-30	0.02	1.02	0.93
31+	0.06	1.06	0.83
Children aged 0-5 years present/ passing through during the interview			**0.30**
Present throughout	0	1	-
Present some of the time	0.19	1.21	0.38
Not present	0.22	1.25	0.12
Children aged 6-15 years present/ passing through during the interview			**<0.01**
Present	0	1	-
Not present	0.43	1.54	<0.01
Self-reported health			**<0.01**
Very good	1	0	-
Good	-0.18	0.83	<0.01
Fair	-0.48	0.62	<0.01
Bad or very bad	-0.66	0.52	<0.01
Longstanding illness/ disability			**<0.01**
Yes	0	1	-
No	-0.24	0.79	<0.01
Body Mass Index			**<0.01**
Underweight (BMI <18.5 kg/m^2)	0	1	-
Normal (BMI 18.5 – 25 kg/m^2)	0.12	1.13	0.51
Overweight (BMI over 25 – 30 kg/m^2)	0.16	1.18	0.38
Obese I (BMI over 30 – 35 kg/m^2)	0.20	1.22	0.31
Obese II (BMI over 35 – 40 kg/m^2)	0.49	1.63	0.03
Obese III (BMI >40 kg/m^2)	0.61	1.83	0.02
Not answered	-0.10	0.90	0.64

Currently using a hormonal method of family planning			**0.03**
No	0	1	-
Yes	0.18	1.19	0.03
Menstrual cycle			**0.06**
In last 7 days	0	1	-
Between 7 days and 4 weeks	0.23	1.26	0.01
Between 4 weeks and 6 months	0.30	1.35	0.03
Between 6 months and 1 year	0.12	1.13	0.45
Between 1 year and 5 years	0.11	1.11	0.40
Not answered	0.30	1.35	0.03
Sexual attraction			**0.15**
Opposite sex only	0	1	-
More often opposite sex, and at least once same sex	0.13	1.13	0.20
All other categories	0.28	1.33	0.12
Depressive symptoms			**<0.01**
No	0	1	-
Yes	0.09	1.09	0.35
Not answered	-0.92	0.40	<0.01
Sexual function			**0.62**
Normal	0	1	-
Low	0.09	1.10	0.25
Not sexually active	0.02	1.02	0.85
Not available	-0.11	0.89	0.59
Intercept	**-1.04**	**0.35**	**0.01**

[1]The dependent variable in the logistic regression model was 1=provided a saliva sample; 0=did not provide a saliva sample. All respondents who were eligible for the saliva study were included in the model.

The non-response weights for the saliva sample (*saliva_nrwt*) were calculated, separately for men and women, as the inverse of the model-predicted probability of obtaining a useable saliva sample. A small number of large weights resulted in a large increase in variance, so the top 2% of the non-response weights (separately for men and women) were trimmed.

The final saliva weight (*saliva_wt*) for the 4128 cases with a useable test result was then calculated as: *saliva_wt = total_wt * saliva_selwt * saliva_nrwt.*

The different components of the saliva weighting are shown in Table 7.8 (with all weights scaled to have a mean of 1 for comparison).

Table 7.8 Distribution of different components of saliva weights after trimming

	Sample size	Minimum	Maximum	Mean	Standard deviation
Main survey weight (*total_wt*) for saliva respondents only	4128	0.17	4.58	1.00	0.57
Saliva selection weight (*total_wt * saliva_selwt*)	4128	0.17	4.37	1.00	0.59
Saliva non-response weight (*saliva_nrwt*)	4128	0.52	2.18	1.00	0.31
Saliva weight (*saliva_wt*)	4128	0.20	9.46	1.00	0.72

The non-response weighting significantly reduced the bias in the profile of saliva sample respondents. For example, 16.4% of those eligible to provide a saliva sample were classified as having 'low sexual function'. At 18.0%, respondents who provided a saliva sample were more likely to be classified as having 'low sexual function', i.e. a bias of 1.6%. After applying the non-response weights, this figure becomes 16.4%%, thereby reducing the bias to less than 0.1% of the weighted sample of saliva respondents who were classified as having 'low sexual function'.

References

[1]Johnson AM, Wadsworth J, Wellings K, Field J (1994) *Sexual Attitudes and Lifestyles*. Oxford: Blackwell Scientific Press.

[2] Wellings K, Field J, Johnson AM, Wadsworth J (1994) *Sexual Behaviour in Britain*. London: Penguin.

[3] Wadsworth J, Field J, Johnson AM, Bradshaw S, Wellings K (1993) Methodology of the National Survey of Sexual Attitudes and Lifestyles. *J R Statist Soc* 156:407-421.

[4] Erens B, McManus S, Field J, *et al*. (2001) *National Survey of Sexual Attitudes and Lifestyles II: Technical Report*. London: National Centre for Social Research.

[5] Johnson A, Mercer CH, Erens B, *et al*. (2001) Sexual behaviour in Britain: partnerships, practices and HIV risk behaviours. *The Lancet* 358:1835-1842.

[6] Johnson AM, Copas AJ, Erens B, *et al*. (2001) Effect of computer–assisted self-interviews on reporting of sexual HIV risk behaviours in a general population sample: a methodological experiment. *AIDS* 15:111-115.

[7] Wellings K, Field J, Wadsworth J, Johnson AM *et al*. (1990) Sexual lifestyles under scrutiny. *Nature* 348:276-278.

[8] Department of Health. (2001) The national strategy for sexual health and HIV. London.

[9] Scottish Executive. (2005) Respect and responsibility: strategy and action plan for improving sexual health. Edinburgh.

[10] The National Assembly for Wales. (2001) A strategic framework for promoting sexual health in Wales.

[11] Johnson AM, Wadsworth J, Wellings K, *et al*. (1992) Sexual lifestyles and HIV risk. *Nature* 360:410-412.

[12] Turner KME, Adams EJ, LaMontagne DS *et al*. (2006) Modelling the effectiveness of Chlamydia screening in England. *Sex Transm Infect* 82:496-502.

[13] Choi YH, Jit M, Gay N. (2010) Transmission dynamic modelling of the impact of human papillommavirus vaccination in the UK. *Vaccine* 28:4091-4102.

[14] Barrett G, Smith SC, Wellings K. (2004) Conceptualisation, development, and evaluation of a measure of unplanned pregnancy. *J Epidemiol Community Health* 58(5):426-33.

[15] Mitchell KR, Ploubidis GB, Datta J, Wellings K. (2012) The Natsal-SF: a validated measure of sexual function for use in community surveys. *Eur J Epidem* 27(6):409-418.

[16] Field N, Tanton C, Mercer CH *et al.* (2012) Testing for sexually transmitted infections in a population-based sexual health survey: development of an acceptable ethical approach. *J Medical Ethics* 38:380-82.

[17] Macdonald PR, Owen LJ, Wu FC *et al*. (2011) A liquid chromatography-tandem mass spectrometry method for salivary testosterone with adult male reference interval determination. *Clinical Chemistry* 57(5):774-775.

[18] Aicken RH, Gray M, Clifton S, *et al*. (2012) Improving questions on sexual partnerships: lessons learned from cognitive interviews for Britain's Third National Survey of Sexual Attitudes and Lifetsyles ('Natsal-3'). *Arch Sex Behav* (DOI 10.1007/s10508-012-9962-2).

[19] Chapter 2 in Johnson AM, Wadsworth J, Wellings K, Field J (1994) *Sexual attitudes and lifestyles*. Oxford: Blackwell Scientific Press.

[20] Gray M, Nicholson S (2009) *National survey of sexual attitudes and lifestyles 2010: Findings and recommendations from cognitive question testing.* London: National Centre for Social Research.

[21] Wisniewski CA, White JA, Michel C-EC *et al*. (2008) Optimal method of collection of first-void urine for diagnosis of Chlamydia trachomatis infection in men. *J Clin Microbiol* 46:1466-69.

[22]Jensen JS, Bjornelius E, Dohn B, *et al*. (2004) Use of TaqMan 5_ nuclease real-time PCR for quantitative detection of Mycoplasma genitalium DNA in males with and without urethritis who were attendees at a sexually transmitted disease clinic. *J Clin Microbiol* 42:683–692.

[23] Hardick J, Giles J, Hardick A, *et al.* (2006) Performance of the Gen-Probe transmission-Mediated Amplification Research Assay compared to that of a Multitarget Real-time PCR for Mycoplasma genitalium detection. *J Clin Microbiol* 44: 1236-40

[24] Bissett SL, Howell-Jones R, Swift C, *et al.* (2011) Human papillomavirus genotype detection and viral load in paired genital and urine samples from both females and males. *J Med Virol* 83:1744-51

[25] Bouvard V, Baan R, Straif K, *et al*. (2009) A review of human carcinogens-Part B: biological agents. *Lancet Oncol* 10:321-2.

[26] Parry JV, Mortimer PP. (1989) An immunoglobulin G antibody capture particle-adherence test (GACPAT) for antibody to HIV-1 and HTLV-I that allows economical large-scale screening. *AIDS* 3:173-176.

[27] Parry JV, Connell JA, Reinbott P, *et al*. (1995) GACPAT HIV 1 + 2: a simple, inexpensive assay to screen for, and discriminate between, anti-HIV 1 and anti-HIV 2. *J Med Virol* 45:10-16.
[28] Connell JA, Parry JV, Mortimer PP, *et al*. (1990) Preliminary report: accurate assays for anti-HIV in urine. *The Lancet* 335(8702):1366-1369.
[29] Mabey D, Changalucha J, West B, *et al*. (1998) Validation study of HIV testing in urine samples. *Int Conf AIDS* 12:764 (abstract no. 41108).
[30] Keevil BG, Macdonald PR, Macdowall W, *et al*. (2013, in press) Salivary testosterone measurement by liquid chromatography tandem mass spectrometry in adult males and females. *Annals of* Clinical *Biochemistry.*
[31] Food and Drug Administration, Department of Health and Health Services. (2001) *Guidance for industry: bioanalytical method validation*. http://www.fda.gov/downloads/Drugs/Guidances/ucm070107.pdf
[32] Thienpont LM, Van Uytfanghe K, Blincko S, *et al*. (2008) State-of-the-art of serum testosterone measurement by isotope dilution-liquid chromatography-tandem mass spectrometry. *Clin Chem* 54: 1290-1297.
[33] The American Association for Public Opinion Research. (2011) *Standard definitions: final dispositions of case codes and outcome rates for surveys*. 7th edition. AAPOR. http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions2&Template=/CM/ContentDisplay.cfm&ContentID=3156
[34] Lynn P, Beerten R, Liaho J, et al. (2001) *Recommended standard final outcome categories and standard definitions of response rate for social surveys.* ISER Working papers 2001-23. Colchester: University of Essex. https://www.iser.essex.ac.uk/publications/working-papers/iser/2001-23.pdf
[35] Payne RA, Abel GA. (2012) UK indices of multiple deprivation – a way to make comparisons across constituent countries easier. *Health Statistics Quarterly* 53.
[36] Betts P, Lound C. (2010) *The application of alternative modes of data collection on UK Government social surveys: Literature review and consultation with National Statistical Institutes*. Office for National Statistics. (http://www.ons.gov.uk/ons/guide-method/method-quality/general-methodology/data-collection-methodology/reports-and-publications/alternative-modes-of-data-collection/index.html)
[37] Office for National Statistics. (2010) *Standard Occupational Classification 2010. Volumes 1 – 3*. Basingstoke: Palgrave Macmillan.