



Ipsos MORI
Social Research Institute

May 2017

Active Lives Survey: Year 1

Technical Report

Ipsos MORI



Contents

1	Introduction	2
1.1	Survey Background.....	2
1.2	Summary of the survey	2
1.3	Signposting for the technical report	3
1.4	Terminology used in the report.....	3
2	Questionnaire development and piloting	5
2.1	Overview.....	5
2.2	Phase 1 development	6
2.3	Phase 2 development	9
2.4	Survey mode and questionnaire design	11
2.5	Overview of survey content by phase and mode.....	12
2.6	Specific questionnaire content and design issues	14
2.7	Design for the young person questionnaire	15
3	Sampling.....	20
3.1	Sample design	20
3.2	Sample size	21
3.3	Household sample design	21
3.4	Sample design for young people	22
4	Fieldwork.....	25
4.1	Letter design	25
4.2	Letters and reminder regime.....	26
4.3	Online questionnaire	26
4.4	Postal questionnaire	27
4.5	Incentives	28
4.6	Survey website	29
4.7	Accessibility	29
4.8	Handling queries.....	30
4.9	Fieldwork for young people	30
5	Response rates	32
5.1	Method and assumptions	32
5.2	Response rates.....	32
5.3	Break offs	35

5.4	Profile of achieved sample	38
5.5	Response rate for young people	40
6	Weighting.....	43
6.1	Design and purpose.....	43
6.2	Methods for creating weights.....	44
6.3	Outline of weight variants.....	45
6.4	Technical information	46
7	Data validation and management.....	51
7.1	Overview.....	51
7.2	Data editing	52
7.3	Back-coding of other sports.....	56
7.4	Deduplication	57
7.5	Selection of cases to be included	58
7.6	Checking of combined final file.....	58
7.7	Imputation and cleaning.....	59
7.8	Creation of derived variables	60
7.9	The checks on the derived variables	61
8	Experiments in the design	64
8.1	Response related experiments.....	64
8.2	Student survey experiment	66
8.3	Progress bar/ progress text experiment	70
	Appendices	73

List of Figures

Figure 7.1:	Example of the paper questionnaire editing tool	54
Figure 8.1:	Approaches considered for contacting students	67
Figure 8.2:	Example of the progress bar indicator.....	70
Figure 8.3:	Example of the progress text indicator	71

List of Tables

Table 2.1:	KPIs built into the questionnaire design	6
Table 2.2:	Outcomes built into the questionnaire design from phase 2	9
Table 2.3:	Questionnaire content of the adult survey in order of questions.....	12
Table 2.4:	Questionnaire content for young people.....	16

Table 3.1: Issued sample size by wave	21
Table 5.1: Response rate by wave and phase	33
Table 5.2: Showing the cumulative response rate by mailing for waves 7-12	34
Table 5.3: Break off by section for the two phases in year 1. Showing the proportion of all break-offs that broke off at particular points in the survey	36
Table 5.4: Showing the proportion of all break offs that are ACTIV3 or ACTYR	37
Table 5.5: Break off rate by phase and device	38
Table 5.6: Proportion of questionnaires on each device type by completion status of the questionnaire (whether complete or a break off)	38
Table 5.7: Profile of the achieved adults sample (unweighted) in year one by mode	39
Table 5.8: Profile of the achieved adults sample (unweighted) in year one by mode (continued)	40
Table 5.9: Profile of the achieved 14-15s sample (unweighted) in year one by mode	41
Table 6.1: Date range of the four quarters	43
Table 6.2: Weights needed for the different types of analyses.	46
Table 6.3: Design effect	46
Table 6.4: Control totals used for the weighting	47
Table 8.1: Named sample experiment results – address return rate	65
Table 8.2: Online address return rate by wave and URL type	66
Table 8.3: Showing the breakdown of the student survey	68
Table 8.4: Materials provided to each group in the student survey	69
Table 8.5: Online return rate (number of online questionnaires completed / number of households surveyed) by progress indicator	71

Introduction

1 Introduction

1.1 Survey Background

Sport England is an increasingly insight led organisation that is tasked with the behavioural challenge of getting more people active. To create the right conditions to increase participation, to decide who we invest in, and understand how sport can deliver wider objectives we need both a broad and deep understanding of sports participation.

Through our commitment to the measurement of sports participation over the past decade we have largely unparalleled participation data both internationally and amongst other sporting and cultural agencies in the UK. Through the Active Lives Survey, we have sought to strengthen this commitment whilst ensuring our data reflects the best available methods of data collection and aligns with our new strategy. To achieve this we piloted new methods, consulted widely with the sector and sought advice from independent experts such as the Office of National Statistics and the Institute for Social and Economic Research in the development process.

1.2 Summary of the survey

In this section we provide a brief overview of the key points in relation to the survey design, fieldwork and data. Full details are provided in the rest of the report.

The Active Lives survey was conducted by Ipsos MORI on behalf of Sport England, which commissioned the survey with additional funding from Public Health England, Arts Council England and the Department for Transport.

1.2.1 Design

The Active Lives survey is a 'push-to-web' survey involving four postal mail outs designed to encourage participants to complete the survey online. The survey is fully 'device-agnostic' and so can be completed on mobile devices or a desktop PC/laptop. The first two mail outs are letters containing information about how to access the survey online and a passcode for accessing the survey. At the third mailing a paper self-completion questionnaire is sent out to maximise response. A final letter reminder is sent at the fourth mailing which includes a reminder of how to access the online questionnaire (it does not include a paper questionnaire).

The data were collected between 16 November 2015 and 15 November 2016. The data were collected using an online questionnaire (52%) and a paper self-completion questionnaire (48%). There were two types of questionnaire: one for adults (16+) and a modified version of the adult questionnaire for young people aged 14-15. Both the adult questionnaire and the young person's questionnaire could be completed either online or by filling in a paper version. Valid responses were received from 201,579 people in total.

1.2.2 Questionnaire

There were two phases to the data collection in Year 1 of the survey (see section 2). The core questions needed to produce activity-related measures (such as 'inactivity' and 'participation in sports and physical activity at least twice in the last 28 days') were included from the start of the survey year. In addition, key demographic variables (age, gender, ethnicity, occupational status) and measures such as height/weight and fruit/vegetable consumption were included from the beginning.

Halfway through the year (in May 2016), phase 2 data collection started with new questions added, to gain further insight into sport and physical activity participation in England. These questions include asking the setting in which activities take place, people's habits, motivations, membership of clubs and levels of volunteering. The questions also included some new classification questions such as household living arrangements, sexual identity and religion. Some questions were included across the sample from phase 2, some were included for online cases only and others were included for half the online sample; in order to keep the questionnaire to a reasonable length, while ensuring there were enough data for analyses. Later in this report we show which questions were asked of each group.

1.2.3 Sampling

The sample is selected from the Postcode Address File using random probability sampling and one letter is sent to each address inviting up to two adults from the household to take part. Young people aged 14-15 years old are identified through responses to the adult questionnaire and an additional mailing is sent to named 14-15 year olds, where applicable. In total, during the first year of the survey, initial invitation letters were sent to 845,230 addresses in England.

1.2.4 Weighting

Weighting is required to make the achieved sample match the population as closely as possible. For the Active Lives survey, the weights correct for the disproportionate selection of addresses across local authorities (related to target samples for each local authority) and for the selection of adults and youths within households. They also adjust the achieved sample by month to control for seasonality. In addition, by weighting to population estimates and national estimates from the Annual Population Survey (2015), the weights should also reduce bias in the survey estimates.

1.3 Signposting for the technical report

This report is organised broadly in the order in which survey processes were completed: questionnaire design, sampling, fieldwork, response, weighting, data management. A final section describes some experiments which were conducted during the first year of the survey.

14 and 15 year olds were also surveyed as part of Active Lives. This was carried out in a similar manner to the adult survey, but there are instances where this differed. The details for the this are presented in the final section in each chapter, with this element referred to as the *young person questionnaire*.

1.4 Terminology used in the report

Those who respond to the survey are referred to as *respondents*. The reason for not using the conventional term *participants* is that, in the context of sports participation, *participant* has a different meaning. When describing the development of the questionnaire using focus groups and cognitive interviews, those who took part are referred to as *participants*, since *respondent* is not a suitable term to use in the context of qualitative methods. The term *young person questionnaire* is used to refer to the separate questionnaire administered to 14-15 year olds identified in the households of those completing the adult questionnaire.

Questionnaire development and piloting

2 Questionnaire development and piloting

2.1 Overview

There was a phased approach to questionnaire development. The survey started in November 2015 in order to be able to provide baselines for key tracking measures in early 2017. The Government's *Sporting Future* Strategy was not published until December 2015 and so content which needed to be developed after the publication of this was added six months after the start of the survey, once it had been developed.

Initial questionnaire for phase 1: all the key sports and physical activity participation questions were included from the start of the survey. This included questions about whether each activity had been done in the last year, how many days in the last 28 the activity was done, how long each session lasts and the intensity of the session (whether it raises your breathing rate, whether it makes you out of breath or sweaty). In addition, key demographic variables were included.

Further development for phase 2: new questions on volunteering, engagement, settings in which activities are done wellbeing, self-efficacy and community trust (which are Sport Future outcomes), and more detailed demographics were introduced at phase 2. A small number of questions, for example, on walking pace and body image were removed at this phase in order to make space for the new content and because they were lower priority.

Mode: The questionnaire was offered online and in a paper version, so all questionnaire development involved designing and testing instruments for these two modes. At phase 2, because the questionnaire content was greater than the time and space available in the questionnaire, the online sample was split into two groups. Some of the additional content was asked to *group 1* and some to *group 2*. Which group a respondent in is determined by when they entered the online questionnaire – respondents are allocated alternately to group 1 and group 2.

Young person's questionnaire: The survey included 14- 15 year olds, through a separate invitation to named young people after their parents had taken part and had given permission for their child or children to be subsequently contacted. The questionnaire for this group was similar, but excluded questions such as working status and occupation status, height and weight and information on people in household. Both online and paper questionnaires for this age group were developed using the adult survey as the starting point.

2.2 Phase 1 development

2.2.1 Design of questions

The Government's Sport Future Strategy includes some Key Performance Indicators (KPIs) in relation to participation in sports and physical activity. The Active Lives survey is designed to provide data to measure progress towards those indicators. The KPIs for which data have been collected since the beginning of phase 1 are outlined in the table below:

Table 2.1: KPIs built into the questionnaire design

KPI	Summary definition	Precise definition
1a ¹	KPI 1a – The percentage of the adult population undertaking at least 150 minutes of moderate physical activity a week	Physically active with at least 150 minutes per week of moderate intensity activity in bouts of 10 minutes or more spread over several days. Periods of vigorous activity are counted as double.
2a	KPI 2a – The percentage of adults physically inactive	Physically inactive with less than 30 minutes per week of moderate intensity activity in bouts of 10 minutes or more spread over several days. Periods of vigorous activity are counted as double.
3a	KPI 3a – The percentage of the adult population taking part in sport and physical activity at least twice in the last 28 days	Participate in at least two sessions of at least moderate intensity activity in the last 28 days combining to at least 60 minutes in total over the 28 days.
11a	KPI 11a – The percentage of adults who have attended a live sporting event at least twice in the past year	Attended a live sporting event at least twice in the past year to include watching professional and amateur sport

In addition, Sport England has a responsibility to collect data on participation in specific sports and disciplines, not just overall participation in activity.

In order to produce data for these measures, the questionnaire needed to ask questions to capture information on:

- which activities people take part in,
- how often they take part,
- how long they take part for each time, and
- the intensity of the activity.

Key demographic information was needed, such as age, sex, ethnicity, disability, socio-economic status, education level. The survey was also being used to gather data on other health measures including obesity and fruit and vegetable

¹ KPI 1a and KPI 2a are also treated as measures of the physical wellbeing outcome in the strategy (outcome one).

consumption for Public Health England, and information on participation in some key arts and cultural activities was also collected for Arts Council England.

During the design phase, questions from other surveys including the previous Active People survey, the Health Survey for England and IPAQ (International Physical Activity Questionnaires)² were considered. Active Lives is a new survey with a different methodology and intended to measure different outcomes from those in the Active People survey. Nonetheless it was important that data could be reproduced on some of the key measures (e.g. overall participation, participation in swimming, running, golf and other key activities).

In addition, a limited review of literature on physical activity measurement was conducted to explore issues such as the most suitable reference period for collecting data, and the best approach to capturing intensity.

A draft questionnaire was produced through collaboration between Ipsos MORI and Sport England with Sport England focussing on the requirements in terms of their data needs and Ipsos MORI focussing on questionnaire design issues.

2.2.2 Cognitive testing

Cognitive testing is a method derived from cognitive psychology to test survey questions by exploring how people understand and respond to survey questions. It also offers an opportunity to observe how participants react to the subject matter: their level of interest and any sensitivities. By uncovering problems which may not be apparent from a standard field pilot, questions can be improved so that the questions are understood by participants in the way intended by researchers.

Three rounds of cognitive testing were carried out on the questionnaire prior to fieldwork. All the interviews took place at the Ipsos MORI offices in London using an observation suite, in which the client and colleagues could observe the interviews, with participant permission. The individuals were recruited using Ipsos MORI's team of recruiters. Quotas were used to ensure that the individuals interviewed included men and women from a range of ages, ethnic backgrounds, those with and without a long-term impairment and people who are active (both through sport, walking and gardening) and who are inactive.

The questions were initially drafted on paper, but with the online design in mind. An initial round of cognitive testing with five participants took place to explore whether the terms and language being used were suitable, how well people could recall the information being requested and how they grouped different sports and physical activities. The questions looking at when people participate in sport across the year were also tested as these had not been included in previous surveys.

The second round of testing included six participants where the postal version of the questionnaire was tested. This round of testing was also used to explore participant views on the recruitment materials (the four separate mail outs) and on their ability to answer the questions and navigate through the postal questionnaire.

The third round of testing involved five interviews to test the online questionnaire. The initial two rounds of testing had involved a seven-day reference period for the detailed information about activities but this was increased to 28 days for the online questionnaire testing in the third round of interviews.

² HSE: <http://content.digital.nhs.uk/catalogue/PUB19295/HSE2014-Methods-and-docs.pdf>

IPAQ: http://www.sdp.univ.fvg.it/sites/default/files/IPAQ_English_self-admin_long.pdf

Some of the key findings which impacted on the design of the questionnaire and should be considered in data analysis are:

- participants found it difficult to understand the question about which time periods they have taken part in each activity over the course of a year (split into three time periods – last 3 months, 4-6 months ago, 7-12 months ago). For the online questionnaire, use of explanations and checks when people only selected one period can be used to aid understanding but this is not possible on the postal questionnaire. Data on time periods throughout the year were included on the postal questionnaire during phase 1 but analysis of the data confirmed people's confusion. As a result, these questions were removed from the postal questionnaire and only online data about time periods over the year (ACTYR) should be used.
- Although a recall period of a week is recommended, participants were able to recall the activities they did in the past 28 days and so this period was used for the main survey in line with the KPI 3a reference period.
- There is a balance to be struck between providing participants with all the information they need in order to know what is included in each activity and making the questionnaire so wordy and complicated that it is hard to complete. Therefore, on the postal questionnaire additional clarification detail was provided at the bottom of some pages. For the online questionnaire, information buttons which could be clicked on for more information were used as an alternative. When analysing the data, users should be clear that not necessarily everyone will have read all the clarifying information on what should and should not have been included in each activity.
- Participants make errors in reporting how long they spend doing activities, thinking about overall time over 28 days rather than a 'session', not thinking about walking there and back on journeys etc. Analysis of the data had confirmed these and a range of other measurement errors. As described in later sections of this technical report, measures such as capping extreme values have been used to correct for implausible information.
- When reporting activities done, some people include times when they were coaching others even they should not be counted under the Sport England definitions used for this survey. Therefore, it is necessary to provide specific instructions not to include coaching others in reported activities.

2.3 Phase 2 development

2.3.1 Design of questions

The strategy document “Sporting Future: A New Strategy for an Active Nation”³ was published by the Department for Media, Culture and Sport in December 2015, (after fieldwork had started) and included some KPIs which required data not included from the start of the survey. The following two KPIs require data from the Active Lives survey to monitor progress and so, from phase 2, additional questions were added about volunteering alongside some additional classification questions (religion, sexual identity and household living arrangements).

- KPI 9a – The percentage of adults volunteering in sport at least twice in the last year
- KPI 10a – The demographics of adult volunteers in sport compared to society as a whole

In addition, Sport England needed information on the settings in which people take part in activities, particularly outdoor settings as participating in sport and activity in the natural environment is a key element of *Sporting Future*.

There was a requirement for questions on membership of clubs, attitudes to sport and activity, the extent to which people have physical activity as a habit and readiness to participate (for those who are less active). New questions were added to collect data on the strategy outcomes which include ONS wellbeing, self-efficacy and community trust.

Table 2.2: Outcomes built into the questionnaire design from phase 2

Outcome	Summary definition	Precise definition
2	Mental Wellbeing	% of population reporting positive subjective wellbeing as measured by the ONS wellbeing questions
3	Individual development	% of population reporting positive perceived self-efficacy as measured in a question about goals
4	Social/community development	% of population reporting positive levels of social trust

Since some of the new questions needed are complex, and there were few existing questions to draw from, Sport England commissioned behavioural scientists TBA (The Behavioural Architects) to do a review of the concepts and how they could be operationalised.

Sport England also commissioned SIRC (Social Issues Research Centre) to conduct a review of volunteering and club membership literature and existing questions. Alongside the SIRC review, focus groups were carried out to understand how people describe their motivations and habits in relation to sports and physical activity to ensure that the resulting questions correspond to the way in which people think about these issues.

Changes were also made to the questionnaire at phase 2 as a result of examining the data from phase 1 and identifying areas for improvement. For example, it was clear that people did not always include themselves in the number of adults in the household so a change was made to make it clearer they should include themselves. As mentioned above, we also

³ <https://www.gov.uk/government/publications/sporting-future-a-new-strategy-for-an-active-nation>

found that the data about when activities were completed across the year had not worked well on the paper questionnaire and so it was removed from the paper version from phase 2. We also discovered that people were mentioning a wide range of other activities they had done which were not counted as sport and physical activity (e.g. wood chopping and playing with toddlers etc.). The wording of the request for other activities was changed to help people understand what should be included. The wording was changed from 'other sport, fitness and recreation activities' to 'other sport or fitness activities'.

2.3.2 Focus Groups

Four focus groups were carried out in the offices of Ipsos MORI in central London and Manchester, using the observation suite in London and a meeting room in Manchester. The composition and location of the four groups is described below.

- **London:** Males and females aged 18-34 – all levels of activity
- **Manchester:** Males and females aged 35-59 – active
- **London:** Males and females aged 35-59 – inactive
- **Manchester:** Males and females aged 60+ – all levels of activity

During the focus groups, participants were asked to discuss motivations and barriers to participation in activities, the stages they go through in preparing to do activities, their ability to take part, the types of settings in which they take part and with whom, and their views on 'organised sport'. There was also a discussion about volunteering and the ways in which people describe the volunteering activities they do.

The findings from these groups were used alongside the report from TBA and SIRC to inform the design of questions on these new topics. They informed the design of questions in terms of the topics which could be covered and the language which people used and would understand.

2.3.3 Cognitive testing

Two rounds of cognitive testing were carried out as part of the development of the phase 2 questionnaire. The first round was carried out in people's homes in Kent and Norwich. The sample was designed to include men and women from a range of ages, both active and inactive and living in different household types. The testing was carried out using a mixed mode approach with some questions tested online and some on paper.

The testing included questions which were not certain to be included in the survey and, for efficiency, those which were more speculative were tested on paper. Questions on "who they do activities with", and "the extent to which participants try multiple activities which they do not continue with", did not get included in the final questionnaire. Questions about settings, habits, readiness for activity, motivation, and volunteering were included in the final questionnaire with modifications.

The second round of testing was carried out in Ipsos MORI's offices in London with six participants. Again, the sample was designed to include a range of men and women of different ages, household types, ethnic groups and with differing activity levels. This testing focussed on the new questions which were planned for phase 2 including household living arrangements, sense of community and individual development as well as settings, habits, motivations, and readiness for activity. The testing was also used to explore whether the activity list for fitness and gym activities could be improved to

reduce duplication in reporting and correspond more closely to the measures required by Sport England. A new format for the activity questions which did not involve using a grid format was also tested and found to work successfully.

The key findings were that the new format for collecting details of activities in the last 28 days worked and that the new questions being suggested about settings, habits, readiness for activity, motivation and volunteering were things that people were able and willing to answer questions about. There were specific recommendations, such as the need to only show settings which are relevant to the activity being asked about, the fact that sport and exercise should not be asked about in the same motivation question, and that specific activities were missing from the volunteering question.

Other findings included avoiding double negatives in a question about motivation, and only asking a question about activities being pointless to those whose previous answers were not at odds with this view. These findings were used to modify the questions prior to their inclusion in the main survey.

2.4 Survey mode and questionnaire design

The survey was offered both online and as a paper version which could be returned by post. The first two invitation letters provided the link/unique passwords to the online survey and a paper questionnaire was offered at mailing 3 (alongside a repeat of the online log in details). Mailing 4 provided a final reminder with details of the online log in details included again. Although the paper questionnaire was only sent unprompted at mailing 3, participants could request a paper questionnaire after any of the mailings. In the final dataset, 52% of the completed surveys were online and 48% were on paper.

The paper questionnaire is 16 pages long and involves some simplification of the online version. For example, it is not possible to offer the complete list of activities which can be shown online. This is because for the online questionnaire, activities are grouped into categories and respondents can click to see a drop down of the activities within each group. This allows a larger number of activities to be presented than on the paper questionnaire.

In addition, more complex questions (such as those asking which settings activities take place in) are not included on the paper questionnaire, since on the online version only the settings tailored to each activity are shown and this sophistication is not possible on the paper questionnaire. The following section shows the survey content for each phase and mode and so highlights where data are not available from the paper questionnaire.

As well as allowing the inclusion of more complex questions and options than the paper questionnaire, the online questionnaire also allows for checking of responses. For example, unexpectedly or implausibly high responses for number of sessions per month or length of session can be checked online, giving the respondent a chance to correct their answer. It is also possible to query logical inconsistencies such as an answer of zero adults in their household and to design questions to prevent people selecting multiple answers on questions where only one answer is required (e.g. gender). On an online survey it is also possible to prevent people accidentally missing questions. This means that the data from the online survey can be considered more complete than those from the paper questionnaire. An ongoing challenge for the survey is increasing the percentage of questionnaires completed online.

Efforts to maximise the online response to the survey have been made by offering a device agnostic online questionnaire. This means it can be completed on a desktop or laptop computer, a tablet or a smartphone. For simpler questions the

format is the same on any device, but for more complex questions (for example, grids) the presentation is automatically changed according to the device on which it is completed.

The drop-down response lists (for example on activities and ethnic groups) work equally well on all types of device. The in-house testing of the online questionnaire involved testing on a range of devices including laptops, desktop PCs, standard and mini tablets (on devices with android, iOS, and Linux operating systems), and the most popular types of smartphones (including the latest and older models). It also involved thorough testing on the most popular browsers (Internet Explorer, Chrome, Firefox and Safari).

2.5 Overview of survey content by phase and mode

As explained in the previous sections, the survey was conducted in two modes (online and paper) and there were two phases with new content added at phase 2 and a small number of questions removed. The table below summarises the questions asked and data available from each mode in each phase. In addition, owing to the required questionnaire content being greater than could be included in a single interview, within the phase 2 online questionnaire there were two routes through the questionnaire and the sample was split into two groups (determined at the moment they entered the questionnaire) each of which took one of the routes through the questionnaire.

Table 2.3: Questionnaire content of the adult survey in order of questions

Content	Phase 1 online	Phase 1 paper	Phase 2 online group 1	Phase 2 online group 2	Phase 2 paper
CULTURAL ACTIVITIES					
<i>Cultural activities in last year</i>	✓	✓	✓	✓	✓
<i>Which period in the last year activities were done in</i>	✓	✓ (data not available)	✗	✗	✗
<i>Frequency of cultural activities</i>	✓	✓	✓	✓	✓
<i>Cultural activities in last 4 weeks</i>	✓	✗	✓	✓	✓
SPORT AND PHYSICAL ACTIVITIES					
<i>Walking, cycling, dance, and gardening in the last year</i>	✓	✓	✓	✓	✓
<i>Sports and activities in the last year</i>	✓	✓	✓	✓	✓
<i>Which period in the last year activities were done in</i> (for all activities reported in last year)	✓	✓ (data not available)	✓	✓	✗
<i>Whether done activities in the last 28 days</i> (for all activities reported in last year)	✓	✓ (from number of days in last 28 days)	✓	✓	✓
<i>Number of days/sessions of activities in the last 28 days</i> (for all activities reported in last 28 days)	✓	✓	✓	✓	✓

<i>Time spent doing activity in each session</i> (for all activities reported in last 28 days)	✓	✓	✓	✓	✓
<i>Intensity of activities done</i> (for all activities reported in last 28 days unless intensity is assumed)	✓	✓	✓	✓	✓
<i>Settings in which the activities are done</i> (for all activities reported in last 28 days unless setting is assumed)	✗	✗	✓	✗	✗
<i>Membership of clubs for each activity</i> (for all activities reported in last year)	✗	✗	✓	✓	✗
FEELINGS ABOUT SPORT AND ACTIVITIES					
<i>Agreement to statements about habit about up to two specific activities</i> (see questionnaire for routing rules)	✗	✗	✗	✓	✗
<i>Agreement to statements about readiness for sport/physical activity in terms of ability and opportunity</i> (see questionnaire for routing rules)	✗	✗	✓	✓	✗
<i>Agreement to statements about motivation</i> (see questionnaire for routing rules)	✗	✗	✓	✓	✓
VOLUNTEERING					
<i>Which of 7 volunteering activities have been done in last 12 months</i>	✗	✗	✗	✓	✓
<i>Whether volunteered on more than one occasion in last 12 months</i>	✗	✗	✗	✓	✓
<i>Whether volunteered in last 4 weeks</i>	✗	✗	✗	✓	✗
DIET					
<i>Fruit and vegetable consumption</i>	✓	✓	✓	✓	✓
DEMOGRAPHICS					
<i>Gender</i>	✓	✓	✓	✓	✓
<i>Sexual identity</i>	✗	✗	✓	✗	✗
<i>Age</i>	✓	✓	✓	✓	✓
<i>Number of adults in household</i>	✓	✓	✓	✓	✓
<i>Number of children in household</i>	✓	✓	✓	✓	✓
<i>Age of children in household</i>	✓	✓	✓	✓	✓
<i>Household living arrangements</i> (Relationship to people in household and whether children have left home)	✗	✗	✓	✓	✓
<i>Ethnicity</i>	✓	✓	✓	✓	✓
<i>Religion</i>	✗	✗	✓	✗	✗

HEALTH AND DISABILITY					
<i>Height and weight</i>	✓	✓	✓	✓	✓
<i>Perceptions of height and weight</i>	✓	✓	✗	✗	✗
<i>Disability</i> (conditions lasting 12 months or more, whether limiting, and area which disability affects)	✓	✓	✓	✓	✓
<i>Wellbeing questions</i> (ONS)	✗	✗	✗	✓	✗
<i>Self-efficacy (goals)</i>	✗	✗	✗	✓	✗
<i>Community Trust</i>	✗	✗	✗	✓	✗
SOCIO-ECONOMIC STATUS					
<i>Highest educational qualification</i>	✓	✓	✓	✓	✓
<i>Working status</i>	✓	✓	✓	✓	✓
<i>Current studies</i>	✓	✗	✓	✓	✗
<i>Socio-economic variables</i> (for NS-SEC classifications) on employment and education	✓	✓	✓	✓	✓

2.6 Specific questionnaire content and design issues

The main section of the questionnaire is designed to gather information on which activities people have taken part in, how much time has been spent doing them and the level of intensity. This information can then be used to create measures which correspond to the KPIs in the *Sporting Future Strategy*. For this reason, the key reference periods used are the last year and the last 28 days since these are the reference periods in the KPIs.

It should be noted that there are recall errors associated with long reference periods for these types of behaviour. Errors associated with use of a 28-day and a year-long reference period include *telescoping* (where people bring in activities which actually took place earlier) and omission of activities which took place in the period because they have been forgotten. In gathering detailed activity data for a 28-day period, some respondents will have multiplied what they did in the last week by four, which may have led to errors in the calculation, or in reporting information for the 28-day period which was not strictly correct if not all four weeks were like the most recent week.

In gathering information on participation in activities, respondents were not required to select 'no' for activities they had not done. Given the volume of activities to select from, this would have been an excessive burden on respondents. This means that in analysing activity and deriving the key measures, any activities which have not been ticked are regarded as not having been done. This contrast with other questions, where the absence of a tick is treated as missing data.

An important measure for analysis of the activity data is socio-economic status. The key measure used is NS-SEC (National Statistics Socio-Economic Classification). This is derived from information about people's occupations and their workplace. In other official surveys, this is derived using detailed questions about job description, qualifications needed to do the job, the place where they work as well as information on number of employees and supervisory responsibilities which is then coded to SOC (Standard Occupational Classification) which is in turn used to derive NS-SEC. However, this full approach to the collection of the data used to code SOC and then NS-SEC relies on interviewer administration (in person or by telephone). The Office for National Statistics (ONS) have created a short self-coded version which can be

used to create five classes of NS-SEC⁴. This is not as accurate as the interviewer-coded counterpart, with 75% agreement with the full version. ONS ran an experiment in 2001 on their Omnibus comparing interview-coded NS-SEC (the measure collected in the Labour Force Survey) against self-coded NS-SEC (the measure collected in the Active Lives Survey) for the same participants. They found that there was a tendency for people to overstate their NS-SEC level, with a lower proportion coding themselves as semi-routine and routine than in the full version.

Nonetheless, because of the importance of socio-economic differences in participation and the under-representation of some groups in the survey responses, NS-SEC has been used for weighting, even though it is not directly comparable with the NS-SEC version in the Annual Population Survey against which we are weighting. To mitigate the impact of this, highest educational qualification, which is more comparable, between the Active Lives survey and the Annual Population Survey has also been included in the weighting. The weighting is explained in more detail in the weighting chapter.

In the dataset, where follow up questions have only been asked for a subset of respondents, the data have been used to create derived variables which are re-based to the population. For example, volunteering more than one in the last year, and number of long term limiting disabilities.

2.7 Design for the young person questionnaire

The survey design included a questionnaire for those aged 14-15 years old. The questionnaire for young people was based closely on the adult questionnaire. Six cognitive interviews were carried out with young people in the relevant age groups to test the questionnaire. The young people were recruited from among the children of staff members and young people known to the research team. The interviews were carried out in the young person's home or a neutral venue familiar to the young person.

The key findings from the testing with young people were that young people participate in a larger number and wider range of activities than adults, meaning that follow-up questions about activities done in the last 28 days are more burdensome for them. However, for consistency with the adult survey, the reference period of 28 days and follow-up on all activities was retained.

The survey data confirmed the fact that young people take part in more activities so, going forward, a way to reduce the burden on young people is something to be considered. The cognitive testing also found that certain activities may need different explanations or examples for young people, for example: to clarify that music practice is a creative, artistic activity; to clarify whether PE lessons or school matches are to be included and to give examples relevant to the types of activities young people participate in such as Scouts.

The table below shows the content of the young person questionnaire using the same framework as the table above for the adult questionnaire. Owing to the exclusion of questions in the young person questionnaire, it was not necessary to have two groups for the phase 2 online questionnaire.

⁴ Section 14 in the ONS Socio-economic classification guidance.

<http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/classifications/current-standard-classifications/soc2010/soc2010-volume-3-ns-sec--rebased-on-soc2010--user-manual/index.html>

Table 2.4: Questionnaire content for young people

Content	Phase 1 online	Phase 1 paper	Phase 2 online	Phase 2 paper
CULTURAL ACTIVITIES				
<i>Cultural activities in last year</i>	✓	✓	✓	✓
<i>Which period in the last year activities were done in</i>	✓	(data not available)	✗	✗
<i>Frequency of cultural activities</i>	✓	✓	✓	✓
<i>Cultural activities in last 4 weeks</i>	✓	✗	✓	✓
SPORT AND PHYSICAL ACTIVITIES				
<i>Walking, cycling, dance, and gardening in the last year</i>	✓	✓	✓	✓
<i>Sports and activities in the last year</i>	✓	✓	✓	✓
<i>Which period in the last year activities were done in</i> (for all activities reported in last year)	✓	(data not available)	✓	✗
<i>Whether done activities in the last 28 days</i> (for all activities reported in last year)	✓	(from number of days in last 28 days)	✓	✓
<i>Days/ sessions of activities in the last 28 days</i> (for all activities reported in last 28 days)	✓	✓	✓	✓
<i>Time spent doing activity in each session</i> (for all activities reported in last 28 days)	✓	✓	✓	✓
<i>Intensity of activities done</i> (for all activities reported in last 28 days unless intensity is assumed)	✓	✓	✓	✓
<i>Settings in which the activities are done</i> (for all activities reported in last 28 days unless setting is assumed)	✗	✗	✓	✗
<i>Membership of clubs for each activity</i> (for all activities reported in last year)	✗	✗	✓	✗
FEELINGS ABOUT SPORT AND ACTIVITIES				
<i>Agreement to statements about habit about up to two specific activities</i> (see questionnaire for routing rules)	✗	✗	✓	✗
<i>Agreement to statements about readiness for sport/physical activity in terms of ability and opportunity</i> (see questionnaire for routing rules)	✗	✗	✓	✗
<i>Agreement to statements about motivation</i> (see questionnaire for routing rules)	✗	✗	✓	✓
VOLUNTEERING				

<i>Which of 6 volunteering activities</i> (excluded transport from the adult questionnaire) have been done in last 12 months	✗	✗	✓	✓
<i>Whether volunteered on more than one occasion in last 12 months</i>	✗	✗	✓	✓
<i>Whether volunteered in last 4 weeks</i>	✗	✗	✓	✗
DIET				
<i>Fruit and vegetable consumption</i>	✓	✓	✓	✓
DEMOGRAPHICS				
<i>Gender</i>	✓	✓	✓	✓
<i>Sexual identity</i>	✗	✗	✗	✗
<i>Age</i> (including month of birth)	✓	✓	✓	✓
<i>Number of adults in household</i>	✗	✗	✗	✗
<i>Number of children in household</i>	✗	✗	✗	✗
<i>Age of children</i>	✗	✗	✗	✗
<i>Household living arrangements</i> (though not asked for young people, they can be linked to parent data on household living arrangements)	✗	✗	✗	✗
<i>Ethnicity</i>	✓	✓	✓	✓
<i>Religion</i>	✗	✗	✓	✗
HEALTH AND DISABILITY				
<i>Height and weight</i>	✗	✗	✗	✗
<i>Perceptions of height and weight</i>	✗	✗	✗	✗
<i>Disability</i> (conditions lasting 12 months or more, whether limiting, and area which disability affects)	✓	✓	✓	✓
<i>ONS wellbeing questions</i>	✗	✗	✗	✗
<i>Goals and trust in community</i>	✗	✗	✗	✗
SOCIO-ECONOMIC STATUS				
<i>Highest qualification</i>	✗	✗	✗	✗
<i>Working status</i>	✗	✗	✗	✗
<i>Current studies</i>	✓	✓	✓	✓

<i>NS-SEC variables</i> (Employment and education)	x	x	x	x
--	---	---	---	---

Sampling

3 Sampling

3.1 Sample design

The sample for the Active Lives survey was selected from the Postcode Address File (PAF) – a list of all postal addresses in the United Kingdom maintained by The Royal Mail. There are no other sampling frames that give the coverage and ease of use of the PAF. The PAF is a list of addresses and contains no information about residents at addresses. Hence the need to send the survey materials to the address without names, and to issue instructions about selecting up to two adults within each selected address (see the section on household sample design on the next page for further information on this process). Because the samples were unclustered within each local authority, the sample was selected as a systematic sample with the PAF addresses in postcode order. This was to ensure a good spread of addresses across the local authorities.

The sampling was designed to achieve fixed numbers of returns from adults within each local authority across the year of the survey, to permit local level analysis. For the majority of the local authorities (267 from 326) the target number was 500 returns. For the two smallest (City of London and Isles of Scilly) the target was reduced to 250. There were also boosts for 57 of the local authorities to between 750 and 2,000 returns in total (see Appendix for more details).

As this survey used a new and relatively untested methodology with the ‘push to web’ design, we estimated the initial return rates for each local authority using estimates from the GP Patient Survey, which is a broadly comparable large-scale postal survey. Once we had obtained process data for a few waves of the Active Lives survey, we could estimate the predicted return rates directly and use that when selecting the issued sample for future waves. The design was checked for each wave of the survey and the issued sample adjusted according to the latest estimates of return rates so that we would reach the targets for each local authority.

3.2 Sample size

There was considerable variation in the issued sample size for each wave. This varied in order to meet the target number of returns nationally (198,250) and at a local authority level. The table outlining the variation in issued sample size can be seen below.

Table 3.1: Issued sample size by wave

Wave	Issued sample size
1	45,620
2	65,620
3	55,620
4	65,620
5	65,620
6	78,165
7	78,165
8	78,165
9	78,165
10	75,024
11	85,794 *
12	73,644

* The sample size for wave 11(*) was increased by an additional 6,000 as part of an experiment conducted at wave 11.

3.3 Household sample design

Addresses are selected on a random basis from the Postcode Address File (PAF) to take part in the Active Lives survey. At each address, up to two adults can take part in the survey. Two unique login codes for the online survey are provided on each invitation, and at the paper questionnaire mailing (M3), up to two questionnaires are sent (if one adult in a household has already completed the questionnaire online, then only the unused password and one questionnaire will be sent in reminders, which follow the initial invitation) Each adult who completes the questionnaire receives a £5 paper voucher/or online gift code for doing so. This methodological approach was used in preliminary testing by Sport England, prior to commissioning Ipsos MORI to undertake the Active Lives survey.

In the preliminary testing work on this approach carried out by Sport England, four log-in codes were provided in the invitation letter (each with a conditional incentive). While this removed the need for any selection of adults in nearly all households (only 1% of households have more than 4 adults resident), it has the potential for fraud whereby a household can complete the survey for more adults than actually live there in order to obtain the incentives on offer. The work for Sport England found that 4% of addresses filled in the survey for more adults than they listed as living at the address, and the average size of participating households was higher than would be expected (2.19 compared to an average household size of around 1.8).

Collecting up to two completed questionnaires per household reduces the chance of receiving fraudulent responses. There is still scope for fraud using this approach, but this is reduced. Therefore, for the main Active Lives survey, two log-ins were provided on each letter.

3.3.1 Process for selecting adults within a household

There are many approaches that could be taken to selecting adults within households. For instance, the two adults with the most recent birthdays or the adults with the two next birthdays could be selected to complete the questionnaire. These are commonly referred to as *quasi-random approaches*, as they are roughly equivalent to a fully random approach. While this would randomise the selection process to a degree in households where there are more than two adults, in self-administered surveys it just adds another barrier to completing the survey and has shown to not be carried out correctly in about 20% to 25% of cases⁵.

With this previous point in mind, it was decided not to apply any selection criteria to taking part in the survey. Instead, any two members of the household (aged 16+) can take part. Under this approach, it is estimated that 93% of the sample are the ones that would be selected using a quasi-random approach (compared to 57% of the sample if only one adult was selected).

3.3.2 Impact of up to two adults per household taking part in the survey

There is a small impact on data from obtaining more than one response per household, which results from the increased clustering effects – we would expect people in the same household to have on average similar levels of sporting participation, which would result in a small loss of precision. However, this loss in precision is more than outweighed by the gain in precision from the resulting less variable within household selection weights. For any estimates that are reported by sex, then the clustering effect within the household is more or less negligible as for the vast majority of households with more than one respondent, one will be male and one female

3.4 Sample design for young people

Sampling for the 14-15 year olds followed a different approach from that used on the adult main survey. During the adult survey, questions are asked on the composition of a respondent's household. If a respondent indicates that they have a child or children between the ages of 0-15 living in their household, they are then asked the age of their child/children. If a respondent indicates that they have a 14 or 15-year-old in their household, they are then asked follow up questions later in the survey.

The follow up questions ask if the participant is the parent or guardian of any 14-15 year olds in their household and, if so, whether they would be willing for us to send their 14 or 15-year old a separate invitation to take part in their own survey. Those who agree are asked for the child's/children's name. Only if the parent/guardian provide informed consent for this and they provide a name(s), is their child selected to take part in the follow-up 14-15 survey.

⁵ TNS BMRB (2013). Community Life Survey: Summary of web experiments. Report prepared for the Cabinet Office. City University London (forthcoming). Feasibility Report – costs, response rates, design effects & quality of alternative data collection. ESS-DACE, Deliverable 7.4

As there is a chance that two 16+ adults from a household may have taken part in the survey and both of these may have provided information on the same child/children, a wide range of checks are carried out to ensure that the same 14-15-year old is not sampled for the survey twice. Duplicate checks are carried out on household ID and on name combined (on both perfect and imperfect spellings) – some of this checking is automated and some manual, to account for any spelling errors or shorthand versions of a name.

Once this has been done an ID is created for the sampled member to be sent the young person questionnaire. This contains the same first 13 digits of the parent ID, with only the last two differing. As the first 13 digits of the ID are unique to a household then the address can be pulled through from the adult's sample based on this. Additionally, this allows for the possibility of adult and youth responses from the same household to be linked.

During year 1 there were only 11 waves of fieldwork for 14-15 year olds, as the year ended before the sample of young people could be followed up from wave 12. In total 4,949 14-15 year olds were sampled for the survey across these 11 waves in year 1.

Fieldwork

4 Fieldwork

This section outlines the approach used for fieldwork for the adult survey.

4.1 Letter design

The principles for designing the invitation and reminder letters were based on the Tailored Design Method⁶, along with a host of literature and best practise based on previous studies, that had been reviewed by the research team. The main aim of the letters is to provide all the relevant information a respondent requires, to complete the survey, and to answer immediate questions which they may have.

Our guiding principles for designing the letter were:

- Use simple and easy to understand language, with no unnecessary complicating text
- To cover key messages that needed to be conveyed in the letters including:
 - (a) Importance
 - (b) Motivators for taking part
 - (c) How to take part
 - (d) Your personal data are safe
- Importance (a) was to be conveyed in all four letters in the following ways:
 - Sport England, Public Health England, and Arts Council England logos should to be prominent;
 - Visual clutter which could distract from the logos and the importance of the survey to be avoided;
 - Professional letter format with address of recipient and full date;
 - Signed by someone with authority (although the public may not have heard of this person, their title should reflect that they are important).
 - Highlight key messages in the text; using these to break up the text and make it easier to read
- The main motivational statements (b) vary across the four letters, thus increasing the likelihood of converting non-respondents.
 1. 1st letter: Help us improve leisure activities in [geography]
 2. 2nd letter: Taking part will benefit you, your family and your community in [geography]
 3. 3rd letter: We want to hear from as many people as possible
 4. 4th letter: An opportunity to help shape the provision of leisure activities in [geography]
 5. In addition, letters 2-4 we include the following secondary motivational statement, to try and encourage participation - We will send you a £5 gift voucher as a 'thank you'

⁶ Dillman, DA. Smyth, JD. Christian, LM. Internet, Phone, Mail and Mixed-Mode Surveys: The Tailored Design Method (2014). Wiley.

In addition to this the letters also provides key information on the organisations involved with the study and contact details for Ipsos MORI, should the respondent wish to get in contact about any matter relating to the survey.

4.2 Letters and reminder regime

One of the biggest challenges when conducting self-completion surveys is maximising the survey response rate – we aimed to do this through effective design of materials and optimising our mail-out strategy.

We adopted a push-to-web approach and the following mailings were despatched on a monthly basis (with the day of the week that mailings land varying):

1. Letter sent to all sampled addresses inviting up to two adults to go online and complete the online questionnaire (C5 white envelope, 2nd class postage)
2. A reminder letter was sent one to two weeks after the first invitation to non-responding addresses and addresses where only one adult had completed the online questionnaire but not a second adult (presence of an eligible second adult was determined in the first questionnaire). (C5 white envelope, 2nd class postage);
3. Another reminder letter was sent one to two weeks after the first reminder letter to non-responding addresses and addresses where only one adult had completed the online questionnaire but not a second adult (presence of an eligible second adult was determined in the first questionnaire); this mailing included one or two postal questionnaires (C4 white envelope, 2nd class postage);
4. A final reminder letter was sent two weeks after the second reminder letter to non-responding addresses and addresses where only one adult had completed the online or paper questionnaire but not a second adult (presence of an eligible second adult was determined in the first questionnaire). Note that in only a small number of cases are paper questionnaires received back in time to feed into the mailing 4 selection (C5 white envelope, 2nd class postage).

4.3 Online questionnaire

The Active Lives survey was hosted using our Global Dimensions platform in Rackspace, a managed hosting facility and Europe's most successful managed hosting company. The security features offered by Rackspace, and Ipsos MORI are listed below:

At Rackspace:

- Rackspace has SAS 70 type II and Safe Harbor certifications
- The servers and network infrastructure are physically located in England
- The servers and network components are fully redundant
- Rackspace guarantees recovery of hardware failures within one hour

At Ipsos MORI:

All access to Dimensions' questionnaires and data was password protected. Only a small number of online survey experts had access.

- Survey data and any respondent personal information were stored in separate databases.
- Penetration testing was carried out on our installation to check that there were no problems.
- Paper questionnaires were stored securely and destroyed at the end of the project.

4.3.1 Survey URL

When deciding on a suitable URL for the survey we had to strike a balance between several factors.

- Availability: We had to find an address that was not already in use. This ruled out options such as ALS.com, ALS.co.uk and ActiveLives.com
- Brevity: We had to choose an address that was short enough for people to remember as we thought this would improve accessibility and would help maximise response rate
- Engagement/appeal: We wanted the URL to give some indication of the content of the survey as well as linking to the overall branding. The theory being that this would help promote engagement with the survey for all types of respondents.

After considering all of our options we opted for <http://www.activelivessurvey.org/> as it was the most suitable option based on the criteria listed above.

Based on internal timings we estimated that the questionnaire would take around 15 minutes to complete. The median actual completion time was 14 minutes 23 seconds for phase 1 compared with 15 minutes and 34 seconds for group 1 in phase 2 and 17 minutes and 22 seconds for group 2.

4.4 Postal questionnaire

The push-to-web design aims to maximise online responses, but we were aware that a significant proportion of people would not complete the survey. The four most common reasons for non-completion were:

- Participant not opening the letter;
- Participant not then going online to complete the questionnaire;
- Participant not having online access; and
- Participant would prefer to complete a postal questionnaire.

Inclusion of a postal questionnaire at the second reminder (mailing 3) increased the likelihood of converting these reluctant respondents, though we recognise that a significant proportion of people would still not respond to the survey at all. As detailed in the previous chapter, for practical reasons the postal questionnaire was much shorter than the online questionnaire and covered 16 pages.

The postal questionnaire was designed in conjunction with our survey methodologists and graphic designers as well as being subject to cognitive testing with members of the public. The aim was to produce a document that was as clear and concise as possible thus reducing burden on participants and maximising response rates. The questionnaire was printed on white, A4 paper in a landscape orientation and bound along the top edge (the young person questionnaire was printed on yellow paper in order for scanners to distinguish between the two surveys). We included several design elements to help participants navigate the questionnaire including:

- The questionnaire was printed mainly using three Sport England brand colours, with highlighted information symbols throughout. These were used to provide definitions or instructions where necessary.
- Arrows were included on rows of grids to indicate if an entire row needed to be completed as this had caused some confusion in the cognitive testing
- Key instructions were underlined or emboldened. In some cases, a different colour was used to highlight specific information.
- Tick boxes were positioned as closely as possible to row text so there was no confusion around which option was being ticked.
- Where multiple options were available, alternative shading was used on each row.

Before sign off, we asked the printer to provide physical copies of the questionnaire so we could establish the quality and the look and feel of the questionnaire.

4.4.1 Storage of scanned images and survey results

All scanned images and survey data were stored on a secure server, which is isolated from the Ipsos MORI network and has restricted access controls. Our secure file servers are housed in server rooms/data centres with appropriate physical access controls and monitoring procedures. The network is protected by appropriate use of firewalls, DMZ and intrusion detection systems. Public facing servers are also appropriately protected and are based on a secure (minimum) two tier or, our general standard, three-tier architecture. All sub-contractors are subject to appropriate quality checks and second party information security audits by our in-house Data Compliance team. We used AES256 as a minimum standard for encryption.

4.5 Incentives

Incentives were used to encourage participation in the survey and boost response rates. £5 conditional vouchers were offered upon completion of the survey.

For online completions, respondents wishing to receive a £5 voucher, entered their email address at the end of the survey. and are then sent a unique link to a dedicated survey website where they were able to select a voucher from a list of alternative suppliers. Participants are then emailed a link/code for their voucher which allows them to redeem it.

Respondents who return a postal questionnaire received a £5 Love2Shop paper gift voucher, which can be redeemed in a wide range of high street stores, within two weeks of Ipsos MORI receiving the questionnaire. Respondents were asked to give their name, in order to address the incentive to the correct person, but even without a name an incentive would be sent to that address.

4.6 Survey website

The main survey landing page (www.activelivessurvey.org) included general information about the survey in order to reassure respondents about the authenticity of the site. The main questionnaires pages were only accessible to those who were invited to take part in the survey by inputting a six-digit alphanumeric password.

The Active Lives survey landing page also included the following:

- A link to Ipsos MORI's main website
- Information on the Active Lives survey privacy policy
- The Active Lives survey accessibility statement
- A dedicated email address (activelives@ipsos.com) should participants have any queries before or after completing the survey
- A frequently asked questions (FAQs) section with a comprehensive list of key questions about the survey designed to reassure respondents and provide help with completing the survey
- A selection of relevant FAQs translated into British Sign Language

The Active Lives survey website was designed to be accessed using a variety of devices, including: desktop computers, laptops, tablets and smartphones. The aim of this 'device agnostic' approach was to improve accessibility and maximise response rates. A google analytics tool was setup to monitor the proportion of those accessing the survey through particular devices. This allowed, traffic to the survey website to be monitored throughout the fieldwork period in order to analyse response patterns and help identify problems with online services or postal deliveries.

The survey website could also be accessed through search engines such as Google and Bing. Most common searches for the survey returned the Sport England Active Lives survey page as the top result. This presents general information and background information about the survey. The link to access the survey website was included towards the bottom of the page.

4.7 Accessibility

The Active Lives survey is designed to be as accessible as possible for all respondents. The functionality of the survey was improved to ensure that content was written and structured using accessible language; the website and documents were assessed by an independent organisation, who specialise in reviewing documents to ensure that they were as easy to use as possible. Lessons learned were shared with the GP Patient Survey Team, which already had extensive experience of reviewing content to ensure full accessibility. They offered further advice about maximising accessibility.

Additionally, the online questionnaire was designed in a way that made it easy for people to adjust colour contrasts and increase font size; Key FAQs were translated into British Sign Language and content was made available in alternative formats and languages when required. This included: large prints of the paper questionnaires and the possibility to complete the survey on the telephone in English, Arabic, Polish, Portuguese, Mandarin and Slovak. The language options available were mentioned on the letters sent to potential respondents and shown using the relevant script and language.

The English Federation for Disability Sport (EFDS) made recommendations to try and improve the accessibility of the survey. Based on these, an accessibility statement was added to the questionnaire homepage and translations of key FAQs into British Sign Language.

4.8 Handling queries

From the beginning of fieldwork, the survey website contained a list of FAQs and provided information about the survey. A telephone helpline and dedicated email address (activelives@ipsos.com) were available for participants to call if they had any queries about the survey.

On a typical month, we received and responded to around 400-450 enquiries from members of the public; of which 65% were emails and 35% were telephone calls to the Active Lives helpline. Of these, less than 2% constituted a 'complaint'.

We receive a significant number of queries via telephone, email and by post from participants and other members of the public. Common queries received included: participants asking to opt out of the survey, requests for paper questionnaires, general queries about the survey and questions regarding the voucher incentive.

Telephone queries were first recorded by an answer machine and a member of the research team returned the call when they had identified an appropriate solution to their request; emails sent to the Active Lives inbox were first answered with automatic responses which included the most commonly asked questions and answers. Each query was then followed up individually.

After the 7th wave of the survey, a dedicated helpline team was set up and trained by the core research team to respond to both telephone and email queries to ensure the highest levels of customer service to those enquiring about the Active Lives survey.

The research team had collected a large amount of data on the most common queries and their solutions and wanted to streamline the process of responding to participant feedback. To do so, the helpline team used a specific list of FAQs to help participants with their enquiry or alternatively forwarded them to the core research team on a case by case basis. A third party supplier was also involved with queries relating to issues with specific voucher codes. Each query was logged in an Excel spreadsheet and a specific code assigned to identify the nature of it. Opt-outs were recorded separately and removed from the sample using their unique reference number to avoid further mailings.

Postal queries were received by TNT Business Solutions, which were then forwarded to Ipsos MORI for sorting. They were then processed in the same way as email or telephone queries depending on the nature of the enquiry and the information available.

4.9 Fieldwork for young people

Fieldwork for young people followed the same format and the same reminder regime as the adult survey. The same survey website was used as for the adult survey, and when young people entered their password they were taken to the young person questionnaire.

Response rates

5 Response rates

5.1 Method and assumptions

The Active Lives survey seeks to survey up to two adults (aged 16+) in every household. This makes it difficult to calculate an individual response rate as the number of adults in non-participating households is unknown. Therefore, it is more sensible to calculate the proportion of sampled addresses that have responded and additionally calculate the number of respondents per responding address. This method has been utilised previously in similar studies.

In order to calculate the response rate, it is also important to be able to estimate the proportion of sampled households that are eligible to respond. This is estimated at 92% (derived from face-to-face PAF surveys and also used by TNS BMRB for the Active People Survey online pilots and the Community Life online survey). Therefore, the calculation for response rates in the Active Lives survey is as follows:

Response Rate = (Addresses where at least one response is received / Total addresses sampled) / 0.92

With the calculation for the respondents per responding households as follows:

Respondents per responding address = Total number of respondents / number of responding addresses

It should be noted that in most cases the address corresponds to the household. However, there are a small number of cases where there are multiple households at the address to which the invitation letter is sent.

5.2 Response rates

The survey year for year 1 ran from the 16 November 2015 – 15 November 2016. Some responses for both waves 11 and 12 were received outside of this time frame, and so these were not included in the dataset or used in the year 1 analysis. However, they have been included in the response rate figures in order to present a full picture for waves 1 -12. Data will be included in the year 2 dataset.

The response rate overall was 18.9% across the survey year with 50.2% of all respondents taking part online and 49.8% filling in a paper questionnaire⁷. The response rate varied very little from phase 1 (18.8%) to phase 2 (18.9%), but did vary greatly between each individual wave. The response rate was as low as 15% for wave 1 and rose as high as 20.1% for wave 10. There are several possible explanations, which are outlined in the section which follows (on reasons for variations in response rates).

⁷ Note that the percentage of online cases in the final data set is 52% rather than 50%. This is because in the case of duplicate cases being received from a household, the online response was retained in preference to the paper response. Also the final data includes those aged 14-15 years old and they were more likely to respond online than adults were.

Table 5.1: Response rate by wave and phase

Wave	Response Rate by Wave	Phase	Response rate by phase	Proportion of returns by mode	
				Online	Postal
1	15.0%	1	18.9%	49.8%	50.2%
2	18.3%				
3	19.9%				
4	19.7%				
5	19.8%				
6	19.6%				
7	17.8%	2	18.8%	50.6%	49.4%
8	18.3%				
9	19.5%				
10	20.1%				
11	18.4%				
12	19.0%				

5.2.1 Response rate by mailing

Each wave of the Active Lives survey is comprised of four mailings with the paper questionnaire introduced at the third mailing (M3). Looking at response rate at each individual mailing does not give a full picture of the impact of the mailing on response. Therefore, it is best to look at cumulative response rate as the survey progresses in order to gauge the true effect of each mailing

M3 as the greatest overall impact on response rate (see Table 5.2). The paper questionnaire is issued to households at this stage. A minority of households may not have access to the internet so this may represent the only way they can complete the survey. Furthermore, there are fewer barriers to completing the physical questionnaire (respondents do not have to access the internet or follow instructions to log into the survey online) and the fact that it is quite a substantial pack that is sent (ie with two questionnaires and a letter in a C4 envelope), may draw respondents to it.

Table 5.2: Showing the cumulative response rate by mailing for waves 7-12 (Phase 2)

Mailing	Return Rate	Response Rate	Returns per HH	Proportion of mailing returns online	Proportion of mailing returns postal	Returns per HH
M1	4.1%	4.4%	1.21	100.0%	-	1.212673
M2	7.4%	8%	1.24	100.0%	-	1.239375
M3	14.5%	15.8%	1.34	14.0%	86.0%	1.342786
M4	17.3%	18.8%	1.37	38.9%	61.1%	1.368173
Total	17.3%	18.8%	1.37	51.5%	48.5%	1.368173

Of the two online only mailings (M1 and M2, where the participant has not yet received a physical questionnaire), the first mailing appears to have the greatest impact on response rate. The fourth mailing (M4) has the lowest overall impact on response rate but is beneficial as it helps to improve the overall demographic profile of the achieved sample.

5.2.2 Reasons for variations of response rate

It is important to note that there are a wide range of external factors, beyond Ipsos MORI's control, which may account for the variation seen in year 1 in response rate. For instance, the day of the week a survey lands and whether the survey lands during a holiday period or close to a bank holiday may have influenced the number of households responding. Although overall deadwood was assumed at 8%, this may have varied between each wave.

The response rate for wave 1 (15%) is a notable anomaly, it was much lower than for any following waves. It was decided that action would be taken to try to improve the response rate. During wave 2 the logos were added to the envelope of the first and final reminders. This appeared to have a positive impact on response rate. Logos were added to the envelope of the second reminder at wave 3 and to the third reminder, later in the survey year. More details of this can be found in chapter 8

During year 1 of the survey the target number of responses was 198,250. The number of target returns varied between 250 and 2,000 for the 326 LAs in England (for 267 of the 326 LAs the target number of returns was 500). In order to achieve these targets, the sample design varied between waves. For instance, if a response rate was low in a particular LA a greater number of households received invitations in following waves, than originally planned, and vice versa. The net impact was that there was an increased sample size in areas less likely to respond, which would have negatively impacted on the response rate overall. This is likely to account for some of the variation seen in the response rate across the survey year.

5.2.3 Other evidence for variation in response rate

The push-to-web method is comparatively new in the field of survey research. So while there is a wide range of literature on survey response rates generally there is a comparatively small amount of literature available on the factors that impact response rates in push-to-web surveys. However, it has been suggested that the response rate for push-to-web surveys varies by the survey subject matter ⁸

⁸ Williams, J. (2016/ 2017). An introduction to address-based online surveying. Social Research Practice 3, 23-36.

Response rates are known to be lower in online surveys in comparison to postal only surveys⁹, this is likely to extend to push-to-web surveys where the barriers to accessing the survey are greater (not sent a survey link directly to an email address) than traditional online surveys.

While an ever-increasing proportion of the population have access to the internet¹⁰, it is known that a significant proportion of the population (just over 10% of households in Great Britain)¹¹ do not have access to it. Therefore, a paper questionnaire is sent to respondents at the third reminder stage in the hope to combat some of the non-response from this group. However, the fact the survey is introduced initially as an online survey is likely to put some people off responding to it entirely.

It is worth mentioning that other experiments were carried out in Year 1 to improve the response rates. While none of these had a significant impact on the response rates they did impact it slightly. This may account for some of the variation seen in response rate.

5.3 Break offs

A break off occurs when a participant enters the online questionnaire but does not complete it. Software allows this abandoned survey data to be captured. These data can be analysed and used to identify problems with the survey, formatting issues on devices (which can arise on an ad-hoc basis due to device updates), indicate questions that respondents find difficult to answer or that there may be technical issues with. It is possible to quantify an overall break off rate by dividing the number who abandoned the survey by the number who started the questionnaire. It should be noted that cases which had a break off near the end of the online questionnaire could be included in the dataset. Once the question on number of adults in the household had been answered in the online questionnaire the case was treated as being productive.

The break-off rate for the Active Lives survey is comparatively low. It stood at 7.34% for phase 1 of year 1 and 8.35% for phase 2 of the survey. The survey increased in length from phase 1 to phase 2 and so it was not unexpected to see an increase in the proportion of break offs from the first to the second phase. While this represents a slight increase in break offs it is still relatively low for an online survey. This highlights the effectiveness of the design and format of the device agnostic questionnaire.

As mentioned in chapter 2 (under section 2,3) there were slightly different versions of the questionnaire during phase 2. For group 1 the break off rate was 8.03% and for group 2 it was 8.66%. In order to understand this, it is important to look at the break offs in more detail for both phase 1 and phase 2.

⁹ Manfreda K., Lozar, M. Bosnjak, J. Berzelak, IH., and Vasja V. 2008. Web Surveys Versus Other Survey Modes: A Meta-Analysis Comparing Response Rates. *International Journal of Market Research* 50:79–104.

¹⁰ https://www.ipsos-mori.com/Assets/Docs/Publications/Ipsos_Connect_TechTracker_Q3_2016.pdf

¹¹ ONS, Internet access – households and individuals:2016

<https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2016> (

5.3.1 Break offs by section

The survey was split into the following seven broad groups for the break off analysis:

- 1 **Pre-survey:** Any pages which you see before the survey actually begins
- 2 **Art & Culture:** Questions relating to whether or not someone had taken part in a particular art or cultural activity, including when they have done it over the course of the year and how often
- 3 **Sports, fitness and Recreation:** Questions relating to activities that the participant has taken part in within the last year. This includes when during the year they had undertaken them; participants are also asked about the frequency, duration and intensity of the activity, if they have done it within the last 28 days.
- 4 **Engagement:** This questions relate to a respondent's habits, motivations, opportunities and volunteering in sport or exercise (only present from Phase 2)
- 5 **About you:** The penultimate group of questions. Ask respondents questions relating to their health and for demographic information
- 6 **Community & Life:** This includes questions on a participant's wellbeing, their goals and their community
- 7 **Final questions:** Questions based around re-contact and participants are asked if they wish to receive the incentive or not.

Table 5.3: Break off by section for the two phases in year 1. Showing the proportion of all break offs that broke off at particular points in the survey

Section	Phase 1	Phase 2
1) Pre-Survey	6.9%	4.9%
2) Arts & Culture	12.0%	5.6%
3) Sport, Fitness & Recreation *	55.7%	53.4%
4) Engagement *	-	6.3%
5) Community & Life *	-	2.1%
6) About you	18.6%	17.7%
7) Final questions	6.8%	10.1%

** some questions in sections 3, 4 and 5 are received only by 50% of participants*

For both phases the area the part of the questionnaire where the most break offs occurred, is during the Sport, fitness and recreation section. This includes questions on which activities respondents have taken part in, within the last 12 months, with a range of follow-up questions depending on when they last took part in the activity. At this stage the questionnaire becomes more complex with a range of drop down menus, grids and questions which require higher levels of cognition. Thus, it is perhaps of no surprise that respondents are more likely to break off here than at other points in the questionnaire.

In the Sport, Fitness and Recreation section, the highest break offs were at questions asking about how often, for how long and how intense the activity was in the past 28 days (ACTIV3). This was the case for both phase 1 and phase 2 of the questionnaire.

Table 5.4: Showing the proportion of all break offs that are ACTIV3 or ACTYR

Question	Phase 1	Phase 2
ACTIV3 (how often, for how long and how intense)	30.4%	32.0%
ACTYR (which periods during the year)	12.2%	8.7%
ACTIV1 (which activities done in the last year)	6.2%	5.1%

During the cognitive interviews for phase 1 some participants found it difficult to remember how often during the last 28 days and how much time they spent doing these activities. It is not unreasonable to expect this to also be the case during fieldwork and respondents dropped out of the survey because they struggled to answer the ACTIV3 question.

ACTYR is the question with the second highest break off rate. Respondents may have difficulty recalling when in the last year they did the activity – which some participants struggled with during the cognitive interviews. This was also a question which respondents found difficult to understand – instead they wanted to report when they last did an activity, and did not expect to be asked all the periods in which they did it.

It is important to note that during phase 1 there were also technical issues with this question for some mobile users, which meant that the response grid was not fully visible when the respondent had undertaken a large number of activities. Due to a late change before going live for phase 1 this was not rectified before the survey went live, but was spotted at a later date. This was corrected during phase 1 fieldwork, but may account, in part, for the slightly higher drop off rate seen here during phase 1. This issue was rectified ahead of wave 6 on 7 April 2017.

ACTIV1 is the question which presented respondents with a list of sports, recreational activities and exercises that they may have undertaken, in the past year. This was presented as a list of types of activities. Each header can be clicked on and beneath it is a list of activities the respondent can select or alternatively they can search for an activity. A total of 6.2% and 5.1% of all break offs occurred here (during phase 1 and phase 2) respectively.

During phase 1, the question with the next highest proportion of break offs was CULYR (5.3%). This question asks respondents during what phase of the year they have taken part in cultural activities, for the cultural activities they say that they have been involved with in the past year. This question mirrors ACTYR (falling before it in the questionnaire and also asking which periods the activities were done in) and is likely to present participants with similar problems. CULYR was dropped between phase 1 and phase 2.

5.3.2 Break off by device

Metadata is collected on a range of different metrics including browser, operating systems, device type and the brand of the device (where applicable). With a device agnostic survey, it is important to look at not only the break offs by section (and question within that) but by type of device as well. Collecting the metadata allows us to do this.

Table 5.5: Break off rate by phase and device.

Device	Phase 1	Phase 2	Overall
Desktop / Laptop	6.2%	7.4%	6.9%
Smartphone	12.6%	12.6%	12.6%
Tablet	5.6%	7.1%	6.4%

It is important to look at not only the proportion of break offs by particularly devices but also what devices respondents complete the questionnaire on. A comparison of the two gives an insight into the difficulties that may be faced with the survey on particular devices.

The break off rate is higher on mobiles than it is on desktop / laptop computers or on tablets. Callegaro et al (2015) have suggested that break off rates tend to be two or three times higher for smartphone respondents compared to desktop / laptop respondents¹². Smartphone users make up a far higher proportion of break offs than they do completes and this is the case for both phase 1 and phase 2.

Table 5.6: Proportion of questionnaires on each device type by completion status of the questionnaire)

Device	Phase 1		Phase 2		Overall	
	Completes	Break offs	Completes	Break offs	Completes	Break offs
Desktop / Laptop	61.2%	51.4%	59.5%	52.1%	60.3%	51.8%
Smartphone	14.4%	26.1%	16.9%	26.7%	15.8%	26.5%
Tablet	21.5%	16.0%	20.8%	17.4%	21.1%	16.8%

* The proportions above do not add to a 100%, as in some cases device type is not captured. Additionally, a small proportion of respondents (less than 0.1% of all respondents completed the survey via consoles).

Overall 15.8% of those who complete the survey do so via smartphone but they represent 26.5% of cases that drop out of the survey. It appears that break offs for smart phone users are higher throughout the questionnaire, not that there is any specific question which is more problematic for smart phone users than users of other device types.

5.4 Profile of achieved sample

Table 5.7 shows the profile of the achieved sample broken down by mode. Note that this is the profiles for responses received in year 1 of the survey (ie 16 November 2015 to 15 November 2016). The number received for the whole of Waves 1-12 was larger.

¹² Callegaro, M. Manfredo, KL. Vehovar, V. (2015). Web survey methodology.

Table 5.7: Profile of the achieved adults sample (unweighted) in year one by mode

	CAWI	PAPER	Total
<i>Base</i>	<i>102,410</i>	<i>96,355</i>	<i>198,765</i>
Gender			
Male	45.7%	42.4%	44.1%
Female	54.3%	57.6%	55.9%
Other	0.0%	-	0.0%
Age			
16-24	8.1%	5.0%	6.6%
25-34	15.7%	8.9%	12.4%
35-44	19.1%	12.0%	15.7%
45-54	19.0%	15.7%	17.4%
55-64	18.5%	19.8%	19.1%
65-74	15.1%	23.0%	18.9%
75+	4.3%	15.6%	9.7%
Disability (whether limiting)			
Limiting disability	12.3%	23.0%	17.4%
Non-limiting disability	14.1%	20.1%	17.0%
No disability	73.6%	56.8%	65.6%
NS-SEC groups			
Managerial, administrative and professional occupations (NS-SEC 1-2)	55.5%	40.6%	48.3%
Intermediate occupations (NS-SEC 3)	11.0%	9.1%	10.1%
Small employers and own account workers (NS-SEC 4)	6.0%	6.1%	6.0%
Lower supervisory and technical occupations (NS-SEC 5)	6.1%	7.5%	6.8%
Semi-routine and routine occupations (NS-SEC 6-7)	8.3%	10.8%	9.5%
Long term unemployed or never worked (NS-SEC 8)	1.7%	2.3%	2.0%
Full-time student	4.7%	2.4%	3.5%
Aged 16-74 and unknown NSSEC	2.3%	5.5%	3.9%
Aged <16 or 75+	4.4%	15.7%	9.9%

Table 5.8: Profile of the achieved adults sample (unweighted) in year one by mode (continued)

Working Status			
Working full or part time	60.3%	47.3%	54.1%
Unemployed	2.7%	2.7%	2.7%
Not working (e.g. retired, looking after children)	28.9%	45.0%	36.6%
Student full or part time	5.0%	2.6%	3.9%
Other working status	3.0%	2.4%	2.7%
Ethnicity			
White-British	84.4%	89.5%	86.8%
White-Other	6.2%	4.0%	5.1%
Asian	5.0%	3.2%	4.2%
Black	1.6%	1.4%	1.5%
Chinese	.6%	.4%	.5%
Mixed	1.3%	.8%	1.1%
Other	.8%	.7%	.8%

The profile of the achieved sample varies from the overall profile of the population. There are several factors that contribute to this:

Survey topic: It has been shown in recent literature¹³ that the profile of push-to-web survey respondents varies depending on survey topics, this is likely to have had some impact on the profile of responders. Non-responders may not have a significant interest in the areas that the survey covers and so will not be inclined to complete the questionnaire.

Modern changes in household structure: It has been shown that there is a growing proportion of young adults (between the age of 20-34) that live at home with their parents¹⁴. These are unlikely to be the individual in the households who opens the mail (as the letter is sent to the household as a whole, and not an individual) and hence who completes the survey.

Online population: While the proportion of those online in England is overall very high, there is a significant proportion of certain sub-groups who are not online. They will not be able to complete the survey online and survey fatigue may have set in by the time the second reminder arrives (with the paper questionnaire). This may have some, if only a small, impact on the overall profile and the response rate.

Work to improve the profile of the overall achieved sample and increase return rates is a key part of the Active Lives study in year 1 and will continue to be so going forward.

5.5 Response rate for young people

Across the survey year the response rate for those aged 14-15 years old was 57%. Since the letters are sent to a named individual and deadwood does not apply, a response rate can be calculated by dividing the number of responses received by the number of invitations sent. This is higher than the adults survey. This is likely because those surveyed are from

¹³ Williams, J. (2016/2017). [An introduction to address-based online surveying](#). *Social Research Practice* 3, 23-36.

¹⁴ <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/bulletins/familiesandhouseholds/2016>

households already engaged with the survey, as at least one parent or guardian within the household has already participated in the survey.

Table 5.9: Profile of the achieved 14-15s sample (unweighted) in year one by mode

	CAWI	PAPER	Total
<i>Base</i>	2,275	539	2,814
Gender			
Male	45.7%	42.4%	44.1%
Female	54.3%	57.6%	55.9%
Other	0.0%	-	0.0%
Age			
14	47.6%	45.8%	47.3%
15	47.3%	46.4%	47.1%
16	4.9%	7.8%	5.4%
Disability (whether limiting)			
Limiting disability	6.7%	5.8%	6.5%
Non-limiting disability	7.8%	13.6%	8.9%
No disability	85.4%	80.6%	84.5%
Ethnicity			
White-British	82.2%	83.0%	82.4%
White-Other	3.9%	4.0%	3.9%
Asian	7.4%	6.5%	7.2%
Black	1.9%	2.3%	2.0%
Chinese	.4%	-	.4%
Mixed	3.1%	3.4%	3.2%
Other	1.0%	.8%	.9%

Weighting

6 Weighting

6.1 Design and purpose

6.1.1 Annual sample weights

Weighting is required to reduce the bias in survey estimates and weights are produced to make the weighted achieved sample match the population as closely as possible. For the Active Lives survey the weights correct for the unequal selection of addresses across local authorities and for the selection of adults and youths within households. They also adjust the achieved sample by month to control for seasonality. In addition, the sample is weighted to match ONS mid-year population estimates and national estimates derived from the Annual Population Survey.

The data needs more than one set of weights to take account of the following elements of the design: postal and online data and the need to analyse data in a time series

The final data includes four weights¹⁵:

- wt_final: for data collected in both the online and postal questionnaires;
- wt_final_online: for data collected in the online questionnaire only;
- wt_time_all: for time series analysis of data collected in both the online and postal questionnaires; and
- wt_time_online: for time series analysis of data collected in the online questionnaire only.

The same weighting strategy was used for each set of weights (this strategy is described below). Note that an approach that weighted each module separately and then combined them would have been less complicated. However, this approach was not an option as the postal sample is issued after the online. The postal sample is therefore not nationally representative and so cannot be treated a separate nationally representative module.

6.1.2 Quarterly and monthly weights

We defined months and quarters by when the questionnaire was completed/returned (rather than issued) and we used survey year months and quarters rather than calendar quarters (see the data validation chapter for more detail on this). The four quarters were therefore defined as follows:

Table 6.1: Date range of the four quarters

Quarter	Months included	Date range
Q1	M1, M2, M3	16/11/15 to 15/2/16
Q2	M4, M5, M6	16/2/16 to 15/5/16
Q3	M7, M8, M9	16/5/16 to 15/8/16
Q4	M10, M11, M12	16/8/16 to 15/11/16

¹⁵ Note that further weights were created to allow for analysis of phase 2 only data but since these data have not been archived, only the weights which apply to the full year of data have been archived.

6.2 Methods for creating weights

There were five stages to the weighting strategy:

1. Calculation of an individual (within household) selection weight;
2. Initial calibration to mid-year population counts for local authority and age/sex and to month counts assuming a proportionate sample;
3. A second stage of calibration to the same set of measures (age/sex, local authority and month) as well as to national estimates from the Annual Population Survey;
4. Trimming of the second stage of calibration; and
5. A final adjustment to regional counts.

We will describe each stage of the weighting for annual weights in detail and then state the difference to produce the time series weights.

Although the published tables were for adults aged 16 or older only, the youth sample members (aged 14 and 15) were included in the weighting. This was so that they could be included in any additional analyses. However, there were no youth sample members for the first month of the survey, and relatively few for the second month. Therefore, the youth sample had to be treated slightly differently when considering month of return in the weighting. The approach used was consistent for the annual and time series weights and is described in the relevant footnotes.

6.2.1 W1: Individual (within household) selection weight

For adults (16 or older), calculated as $W_1 = M_{hh} / m_{hh}$ where m_{hh} is number of adult participants within the household ($m_{hh} = 1$ or 2) and M_{hh} is the total number of adults in the household.

For the 14/15 year olds, calculated as $W_1 = Y_{hh} / y_{hh}$ where y_{hh} is number of 14/15 year old (youth) participants within the household and Y_{hh} is the total number of 14/15 year olds in the household.

In order to reduce the variance of the weights, and hence increase the efficiency of the sample, the individual selection weight was trimmed at 3.

6.2.2 W2: Initial calibration

The next stage was to calibrate to population counts for local authority and age/sex groups, as well as to month for the two sets of weights that covered the full survey year to ensure that the weighted distribution by month was proportionate. This was done separately for each region with the individual selection weights (W_1) used as the starting weights. This weighting stage corrected for the unequal selection probabilities across the local authorities, as well as disproportionate non-response; it also put each age group into its correct proportion. Adjusting for month also balanced the sample to adjust for seasonality.

Note that this initial calibration stage was not trimmed to ensure that the distributions by local authority, age/sex groups and month¹⁶ were as close to the true figures as possible. In fact, trimming at this stage would not have been sensible as it would merely have changed the weights for the very large and very small local authorities and the first few months that

¹⁶ Note that the calibration was set up so that the adjustment for month was only made for participants aged 16 and older. This was done because there were very few 14 and 15 year old participants in the first two months of the survey.

had lower returns. These were features of the design that we wanted to correct for, and so trimming the weights would be detrimental.

6.2.3 W3: Second stage of calibration

For the second stage of calibration, national estimates from the Annual Population Survey were added to the set of control totals for: white / non-white, working status by gender, household size, long-term health problem, socio-economic classification (NS-SEC), and highest educational qualification. The calibration was done separately for each region and the starting weights were the weights from the previous initial calibration stage (W_2).

6.2.4 W4: Trimming of second stage of calibration

In order to reduce the variance of weights, this calibration stage was trimmed. The weights from this calibration stage (W_3) were divided by the initial weights (W_2) to give the adjustment factor (f_{adj}). This adjustment factor was trimmed at the 5th and 95th percentiles within region to give the trimmed adjustment factor (f_{trim}) which when multiplied by the initial weights gave the trimmed calibration weights: $W_4 = W_2 \times f_{trim}$.

6.2.5 W5: Re-scale for region

The final stage was an adjustment so that the weighted counts by region were grossed up to the mid-year population counts. This stage simply multiplied all the weights by a constant within each region to give the final grossed weights (W_5).

6.2.6 Time-series weights

The time series weights were produced using a parallel approach. However, rather than calibrating separately by region, the calibration was done separately for each month¹⁷. We did not need to add region as a control total as it was implicitly included from having local authority as a control total – local authorities nest within region. Note that the youth sample (aged 14 and 15) was included when the weights were being produced, although were excluded from the final weights as no cases were collected for the first two months of the survey. This was done to be consistent with the annual weights, and also with time series analyses going forwards.

At the last stage, we re-scaled the weights so that the weighted distribution by month matched that of a proportionate sample.

6.3 Outline of weight variants

For most analyses, `wt_final` and `wt_final_online` will be the required weights, the former (`wt_final`) for analyses of the full dataset (online and postal) and the latter (`wt_final_online`) for measures that are only collected online.

There are also two sets of time series weights (`wt_time_all` and `wt_time_online`) that need to be used for any analysis by month or quarter across the full survey year. The former set (`wt_time_all`) for the full sample (online and postal) and the latter (`wt_time_online`) for the sample completing online only. The weights which are needed for different levels of analysis are outlined below

¹⁷ Because of the low numbers of 14 and 15 year old participants in the first two months, it was not possible to allocate the 14 and 15 year olds to months. So that they could be included in the calibration weighting, all 14 and 15 year olds were included for every month. This approach is consistent with that used for the annual weights.

Table 6.2: Weights needed for the different types of analyses.

Weight name	Weight type	Measures	Break variables
wt_final	All: Online and Postal	Participation, Frequency, Duration, BMI, Fruit and Veg	Any non-time related
wt_final_online	All: Online	<i>As above, but for activity composites only provided online and education status</i>	Any non-time related
wt_time_all	All: Online and Postal: time series (16 and older)	Participation, Frequency, Duration, BMI, Fruit and Veg, Volunteering	Month, quarter as breaks or when analysis is being carried out for one month or quarter by other breaks
wt_time_online	All: Online: time series (16 and older)	<i>As above, but for activity composites only provided online and education status</i>	Month, quarter as breaks or when analysis is being carried out for one month or quarter by other breaks

6.4 Technical information

6.4.1 Statistical efficiency

The table below shows the estimated effective sample size and corresponding design effect (deft) for the weights. On average the design effect due to weighting is about 1.5 – in other words the width of the confidence intervals around analyses of the corresponding data would be about 50% wider than would be the case if there were no weights. However, some of the variance of the weights is due to the design of the study because smaller local authorities were over-sampled to obtain a minimum sample size. Also, the sample obtained in each month varied during the course of the year. If we take the disproportionate sampling by local authority and differences in the sample size by month into account, then the average design effect is about 1.3.

Table 6.3: Design effect

Weight name	n	unadjusted		adjusted	
		neff	deft	neff	deft
wt_final_all	201,579	84,568	1.544	124,621	1.272
wt_final_online	104,685	44,137	1.540	61,848	1.301

6.4.2 Control totals

The control totals for the calibration consisted of 2015 ONS mid-year population estimates and national estimates from the Annual Population Survey Jan-Dec 2015. The following tables show the estimates for the control totals used for the weighting. Note that we have not included the control totals for local authority, given the number of data points; however, we have included region.

Table 6.4: Control totals used for the weighting

HIGHEST EDUCATION	NE	NW	Y&TH	EM	WM	EE	Lon	SE	SW	ALL
Degree or equivalent	15.3%	17.5%	16.2%	15.4%	16.3%	17.7%	33.8%	21.6%	19.2%	20.4%
Higher education	6.2%	5.5%	5.4%	6.2%	5.4%	5.4%	5.2%	6.1%	6.2%	5.7%
GCE A level or equivalent	13.6%	13.0%	13.0%	13.3%	11.1%	12.3%	9.2%	12.5%	13.2%	12.1%
GCSE grades A*-C or equivalent	14.2%	13.1%	12.9%	14.0%	13.1%	14.6%	8.1%	12.3%	12.5%	12.4%
Other qualification	4.8%	5.9%	6.6%	6.6%	5.9%	6.4%	8.4%	5.6%	5.2%	6.3%
No qual / DK	7.3%	7.0%	7.4%	5.9%	9.7%	5.8%	5.3%	4.2%	3.7%	6.0%
Not 25 to 64	38.6%	37.9%	38.4%	38.5%	38.6%	37.7%	29.9%	37.7%	40.1%	37.0%

HOUSEHOLD SIZE	NE	NW	Y&TH	EM	WM	EE	Lon	SE	SW	ALL
One-person h/hold (14+)	16.9%	15.0%	14.5%	13.1%	13.5%	12.7%	12.5%	12.2%	13.4%	13.5%
Other households (14+)	83.1%	85.0%	85.5%	86.9%	86.5%	87.3%	87.5%	87.8%	86.6%	86.5%

ETHNIC GROUP	NE	NW	Y&TH	EM	WM	EE	Lon	SE	SW	ALL
White (14+)	95.7%	90.9%	90.6%	90.9%	83.9%	91.6%	61.7%	91.9%	96.0%	86.5%
Non-white (14+)	4.3%	9.1%	9.4%	9.1%	16.1%	8.4%	38.3%	8.1%	4.0%	13.5%

WORKING STATUS / GENDER	NE	NW	Y&TH	EM	WM	EE	Lon	SE	SW	ALL
Males 16-74 FT	24.2%	25.7%	26.2%	27.4%	25.9%	28.2%	29.5%	28.0%	26.6%	27.2%
Other males 14+	24.5%	23.3%	22.8%	21.8%	23.3%	20.8%	19.8%	20.8%	22.1%	21.8%
Females 16-74 FT	14.5%	15.8%	14.4%	15.2%	14.4%	15.4%	18.6%	15.5%	14.2%	15.6%
Other females 14+	36.8%	35.2%	36.6%	35.6%	36.4%	35.6%	32.1%	35.7%	37.1%	35.4%

NS-SEC	NE	NW	Y&TH	EM	WM	EE	Lon	SE	SW	ALL
Managerial/professional	20.8%	23.7%	22.2%	23.5%	21.7%	26.6%	32.4%	29.2%	25.9%	26.0%
Intermediate	9.1%	9.1%	8.0%	7.8%	8.5%	8.5%	8.0%	8.9%	7.7%	8.4%
Self-employed and own account workers	4.4%	5.3%	5.8%	5.8%	5.1%	6.2%	7.3%	6.2%	6.5%	6.0%
Lower supervisory/technical	5.5%	4.6%	5.2%	5.3%	4.7%	4.8%	3.6%	4.6%	4.9%	4.7%
Semi routine and routine	17.0%	15.5%	16.7%	16.6%	16.3%	14.6%	11.9%	12.3%	14.2%	14.6%
Never worked, LTE and students	15.8%	14.6%	15.0%	13.2%	15.9%	11.2%	16.4%	11.4%	10.8%	13.7%
Unclassified	3.1%	3.2%	3.1%	3.2%	3.3%	2.8%	3.9%	2.6%	2.6%	3.0%
65+ & 14-15	24.4%	24.0%	23.9%	24.6%	24.5%	25.3%	16.5%	24.8%	27.4%	23.6%

LONG TERM HEALTH PROBLEM	NE	NW	Y&TH	EM	WM	EE	Lon	SE	SW	ALL
Health problems	37.0%	35.8%	37.4%	37.6%	33.0%	35.9%	27.3%	35.1%	37.2%	34.6%
None/missing	63.0%	64.2%	62.6%	62.4%	67.0%	64.1%	72.7%	64.9%	62.8%	65.4%

AGE / GENDER	NE	NW	Y&TH	EM	WM	EE	Lon	SE	SW	ALL
males 14-15	1.2%	1.4%	1.3%	1.4%	1.5%	1.4%	1.2%	1.3%	1.2%	1.3%
males 16-24	7.3%	7.0%	7.4%	7.1%	7.2%	6.4%	6.7%	6.6%	6.5%	6.9%
males 25-34	7.4%	7.9%	7.8%	7.3%	7.9%	7.5%	12.3%	7.3%	7.0%	8.3%
males 35-44	6.8%	7.3%	7.4%	7.3%	7.5%	7.7%	9.7%	7.8%	6.9%	7.7%
males 45-54	8.4%	8.5%	8.4%	8.6%	8.3%	8.5%	7.8%	8.7%	8.4%	8.4%
males 55-64	7.4%	7.0%	7.0%	7.2%	6.8%	6.9%	5.3%	6.9%	7.2%	6.8%
males 65-74	6.0%	5.9%	5.8%	6.2%	5.9%	6.1%	3.6%	5.9%	6.7%	5.7%
males 75+	4.1%	3.9%	4.0%	4.1%	4.1%	4.4%	2.7%	4.3%	4.8%	4.0%
females 14-15	1.1%	1.3%	1.3%	1.2%	1.3%	1.2%	1.3%	1.3%	1.3%	1.3%
females 16-24	6.9%	6.8%	7.1%	6.8%	6.9%	6.1%	6.7%	6.3%	6.2%	6.6%
females 25-34	7.6%	7.9%	7.9%	7.5%	7.9%	7.7%	12.1%	7.5%	6.9%	8.3%
females 35-44	7.2%	7.5%	7.5%	7.5%	7.5%	7.9%	9.3%	8.1%	7.2%	7.9%
females 45-54	8.8%	8.8%	8.6%	8.8%	8.5%	8.8%	8.0%	8.9%	8.7%	8.6%
females 55-64	7.7%	7.2%	7.1%	7.3%	7.0%	7.2%	5.6%	7.2%	7.6%	7.0%
females 65-74	6.5%	6.3%	6.2%	6.5%	6.3%	6.5%	4.1%	6.4%	7.2%	6.1%
females 75+	5.5%	5.2%	5.3%	5.3%	5.4%	5.7%	3.6%	5.6%	6.2%	5.2%

MONTH	NE	NW	Y&TH	EM	WM	EE	Lon	SE	SW	ALL
Jan	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%
Feb	7.5%	7.5%	7.5%	7.5%	7.5%	7.5%	7.5%	7.5%	7.5%	7.5%
Mar	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%
Apr	8.0%	8.0%	8.0%	8.0%	8.0%	8.0%	8.0%	8.0%	8.0%	8.0%
May	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%
Jun	8.0%	8.0%	8.0%	8.0%	8.0%	8.0%	8.0%	8.0%	8.0%	8.0%
Jul	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%
Aug	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%
Sep	8.0%	8.0%	8.0%	8.0%	8.0%	8.0%	8.0%	8.0%	8.0%	8.0%
Oct	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%
Nov	8.0%	8.0%	8.0%	8.0%	8.0%	8.0%	8.0%	8.0%	8.0%	8.0%
Dec	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%
14-15 (age)	2.3%	2.7%	2.6%	2.6%	2.8%	2.6%	2.5%	2.6%	2.5%	2.6%

REGION	Count	%
North East	2,214,952	4.9%
North West	5,979,822	13.1%
Yorkshire and The Humber	4,489,982	9.8%
East Midlands	3,918,592	8.6%
West Midlands	4,761,343	10.4%
East	5,054,961	11.1%
London	7,093,505	15.6%
South East	7,446,732	16.3%
South West	4,630,693	10.2%
All	45,590,582	100.0%

Data validation and management

7 Data validation and management

7.1 Overview

7.1.1 Questionnaire versions

As described in earlier sections, the data have been collected from two sources: an online questionnaire and a paper questionnaire. The online questionnaire includes some built-in routing and checks within it, whereas the paper questionnaire relies on correct navigation by respondents and there is no constraint on the answers they can give.

In addition, the online data are available immediately in the office in their raw form, however the paper questionnaire data have to be scanned and keyed as part of a separate process. Tick box answers are captured by scanning and numbers and other verbatim answers are captured by keying. These data are delivered in weekly batches and there is potential for errors to be introduced during the process of scanning and keying.

All this means that in the early stages of the data processing, the paper questionnaire data have to be handled separately from the online data and are also subject to a higher level of checking and editing than the online data.

As described in the questionnaire section, a new questionnaire was introduced after six months (Phase 2) which included some additional questions not asked from the beginning of the year, as well as the removal of a small number of questions. Within the online questionnaire there are two routes through the questionnaire, each taken by half the online sample, with some questions only asked by group 1 and some by group 2.

In addition, there are paper and online questionnaires for the young person questionnaire which also changed at phase 2. This means that there are nine different versions of the questionnaire and the data from all of these had to be brought together. This section describes the processes by which the data were cleaned and edited, merged together and duplicates removed. These different questionnaire versions also have implications for how missing data are handled and described and how derived variables are created.

7.1.2 Month of data collection

The fieldwork section has described how the survey sample is divided into waves and issued once a month. The fieldwork cycle for a single wave lasts about five to six weeks. The data analysis involves cutting off cases by month of response to the questionnaire. For data analysis there are 12 months, starting on the 16th of the month and finishing on the 15th of the following month (November 2015 to November 2016).

Within the data for one month, there may be cases issued over several different waves. For example, in month 3 (January 16–February 15 2016) data could come from Waves 1, 2 or 3 of the survey. The date of completion of the questionnaire is important for the way in which data are handled and weighted. Online data come with an automatic date stamp and it is known exactly when the questionnaire was completed. In the first phase of the survey, date of completion was not asked on the paper questionnaire.

The date of completion was assumed to be 2 days in advance of the date the questionnaire was processed by the scanning agency (unless the day was a Monday, in which case it was assumed to be 3 days before). In phase 2 a date of completion question was added to the paper questionnaire. Where a valid date was put on the questionnaire, consistent

with the timing of that case being issued and the questionnaire being returned, the date given by the respondent is taken as the date of return. In some cases, respondents give no date, give an incomplete date, or give a date which is impossible (e.g. after the date the questionnaire was received in the office or before the questionnaire was sent to the respondent). In these cases, the date of receipt is assumed to be 2-3 days before the date the questionnaire was processed by the scanning agency.

At the very start of phase 1, the questionnaires were not processed daily. For these cases we do not have a reliable data of completion, but we have derived their month of completion based on the order in which they arrived in with the scanning agency and the number received in the first month of the survey. Since month 2 the scanners have processed and batched the questionnaires daily in order to have a reliable date of receipt, even where respondents do not provide a date of completion.

7.2 Data editing

7.2.1 Postal data – Forced edits

The postal data are subject to errors introduced by respondents, as well as errors resulting from scanning or keying errors. Many of these errors can be dealt with through standard edit rules. For example, if a single code question has more than one category ticked, it is set to 'missing – incorrectly multi-coded'. If a routed question is asked when it should not be, then it can be set to not applicable and the original answer over-written. If a respondent says there are no adults in the household (including them), then the answer can be set to be 1 since we assume that they excluded themselves (although of course if someone wrote 1, when they should have given an answer of 2, this is less easy to identify).

A large number of respondents ticked all the qualifications they had, rather than the highest one, a forced edit was used to retain the highest qualification. For the sport and activity data, if someone did not tick they did it in the last year, but then provided data on their participation in the last 28 days, the data were edited to record that they had done the activity in the last year. A full record is kept of the forced edits done on the data. These edits are done to improve the quality of the data and to make them more consistent and easier to analyse.

7.2.2 Postal data – Manual edits

In examining the early data from the paper questionnaire and comparing it with the scanned data, it was apparent that there were certain common errors in the data which would either not be corrected by forced edits, or where a forced edit might lead to unnecessary missing data. For example, respondents sometimes start ticking in one row on the activity grid, but then mistakenly move up or down a line, further along the row. This means they may say they have walked for travel in the last year, then have no further data on that row, but on walking for leisure have information on participation in the last month, but no data about the last year.





It was common for people to tick no to the question about disability and then change their answer to yes. In a forced edit this would be treated as missing because multiple answers were given but in looking at the questionnaire it was possible to see the no answer had been crossed out but both were picked up by the scanner. Therefore, alongside the forced edits a series of manual edits were specified. These were for errors which might be corrected better by viewing the questionnaire.

A bespoke computer software was developed which allows the scanned page to be viewed against the data captured and for an operator to make edits to the data. These corrections can then be output and applied to the data. Screen shots are

shown on the next page. The edits identified through the automated editing as needing manual input are fed into the program and operators review all the manual edits and make a decision about whether to make a change. A cover page lists each error and by selecting that error the operator is taken to the correct page.

Figure 7.1: Example of the paper questionnaire editing tool

Batch / Serial 160808010040 160671364580082 Showing edited data See Scanned Flag Form Mark page reviewed Cancel changes Save changes Swap Page

Active Lives Survey

The survey asks about healthy lifestyles and leisure, recreational, and cultural activities. It is being conducted on behalf of government agencies including Sport England, Arts Council England and Public Health England by Ipsos MORI, an independent social research organisation.

Example Below is an example of the type of question you may be asked in the questionnaire...

If you see the **e** symbol it means there is extra information available to help you answer the question

	Have you done this activity in the past 12 months? <i>If you haven't please leave the box blank</i>		Have you done this activity in the past 4 weeks? <i>If you haven't please leave the box blank</i>		If you have done this activity during the past 4 weeks could you tell us...				
	Yes	No	Yes	No	on how many days you did it?	the time you usually spent on the activity per day	if the effort you usually put into the activity was enough to...		
					Days (in past 4 weeks)	Hours (per day)	Mins (per day)	raise your breathing rate	make you out of breath or sweat
Team Sports									
Football	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0 8	0 1	3 0	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Cricket	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>				<input type="checkbox"/>	<input checked="" type="checkbox"/>

160671364580082

Page 08: HOURS 'ACTO1H1_new ACTO1H1_2-ACTO1H1_ACTO1H1_2_other_sport' Forced to -96 in loop syntax & MANUAL CHECK > 23
Page 09: Q7 FRUIT over 10 MANUAL CHECK

Batch / Serial 160630010001 160650321952581 Flag Form Mark page reviewed Cancel changes Save changes

Q2 Listed below are types of walking, cycling, gardening and dance which you may have done within the past 12 months

Please ✓ all activities that you have done within the past 12 months; answering the follow up questions for each activity as appropriate

	Have you done this activity in the past 12 months? <i>If you haven't please leave the box blank</i>		Have you done this activity in the past 4 weeks? <i>If you haven't please leave the box blank</i>		If you have done this activity during the past 4 weeks could you tell us...				
	Yes	No	Yes	No	on how many days you did it?	the time you usually spent on the activity per day	if the effort you usually put into the activity was enough to...		
					Days (in past 4 weeks)	Hours (per day)	Mins (per day)	raise your breathing rate	make you out of breath or sweat
Walking e									
Walking for travel	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>				<input type="checkbox"/>	<input type="checkbox"/>
Walking for leisure (incl. dog walking and rambling)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	2 8	0 1	0 0	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Cycling									
Cycling for travel (including commuting)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>				<input type="checkbox"/>	<input type="checkbox"/>
Cycling for leisure and all other cycling	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>				<input type="checkbox"/>	<input type="checkbox"/>
Gardening									
Gardening	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0 7	0 2	0 0	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Dancing									
Creative or artistic dance e	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>				<input type="checkbox"/>	<input type="checkbox"/>
Other types of dance e	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>				<input type="checkbox"/>	<input type="checkbox"/>

I have not done any of these activities in the past 12 months

e Walking: Include all continuous walks of at least 10 minutes without stopping. If you stop for short breaks, such as waiting to cross a road this still counts as continuous. Exclude walking around the shops. Include walking a dog as leisure walking. Mountain/hill walking and hiking should be included at the next question on page 5.

e Creative or artistic dances: For example ballet, ballroom, belly dancing, contemporary, contact improvisation, Flamenco, folk, hip-hop, historical/period, Irish, jazz, jive, Latin American, line or square dancing, salsa, street dance, South Asian, tango or tap.

e Other types of dances: Dance based fitness classes/Zumba should be included under the fitness section later in the questionnaire.

Walking for travel	<input type="checkbox"/>	<input type="checkbox"/>						<input type="checkbox"/>	<input type="checkbox"/>
Walking for leisure	<input checked="" type="checkbox"/>	<input type="checkbox"/>	2 8		0 0			<input checked="" type="checkbox"/>	<input type="checkbox"/>
Cycling for travel	<input type="checkbox"/>	<input type="checkbox"/>						<input type="checkbox"/>	<input type="checkbox"/>
Cycling for leisure	<input type="checkbox"/>	<input type="checkbox"/>						<input type="checkbox"/>	<input type="checkbox"/>
Gardening	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0 7		0 2			<input checked="" type="checkbox"/>	<input type="checkbox"/>
Creative or artistic dance	<input type="checkbox"/>	<input type="checkbox"/>						<input type="checkbox"/>	<input type="checkbox"/>
Other types of dance	<input type="checkbox"/>	<input type="checkbox"/>						<input type="checkbox"/>	<input type="checkbox"/>

Written instructions are provided to coders to explain what should be done and the kinds of problems to be rectified at each question. Options include:

- No change – data are as captured and associated forced edit will deal with any problems,
- Change because captured data are not what was written or intended by respondent (e.g. multicode actually had one answer crossed out and so the crossed-out answer can be removed, a 7 read as 1, etc.),
- Change because data are what was written but an obvious error was made which can be rectified (e.g. answers to breathing questions ticked in wrong row, when it is clear which activity they should be applied to, it is clear that ‘no’ to out of breath and sweaty was actually used to mean ‘I didn’t do the activity’).

Any complex cases which the coder has queries about are flagged and reviewed by a researcher. The data are then output into an error file which is applied to the data and the manually edited data are thus captured, alongside any remaining forced edits.

7.2.3 Online data

The online data need less editing as the checks and edits are found within the questionnaire. For example, extreme high values of time spent doing activities, or durations of less than 10 minutes are checked. Where multiple answers are selected on single code answer questions, respondents are asked to correct their answers. After the data were received in the office, rules were set for defining missing values and a small number of further edits were possible.

7.2.4 Missing values

In the survey data there are various reasons why a question may not have been answered. On the postal data, a question which should have been answered may have been missed by respondents. On the online survey, in order to allow respondents to proceed past questions which they may not know the answer to or do not wish to answer, codes are used for the answers which allow them to say ‘don’t know’ or ‘prefer not to say’. There are also questions which may not be applicable because they were not asked for respondents in that mode, phase or group. Different missing values have been used to cover these reasons.

Table 7.1: Missing values and codes used

Code used	Description
-99	Missing, should have been answered
-98	Not applicable: Survey routing
-97	Incorrectly multi-coded
-96	Out of range
-95	Don't know/Cannot give estimate
-94	Prefer not to say
-93	Not applicable: Postal response
-92	Not applicable: Phase 1 case
-91	Not applicable: Phase 2 case
-90	Not applicable: not asked to this group

Wherever possible, the base for questions has been set to all participants. However, for questions not asked at all for one group, missing values have to be used. For example, information on phase 2 place of study is only available on the online questionnaire for adults (so postal cases are coded with -93). Information on occupational and working status is not available from the young person questionnaire and so -90 is used for these cases.

For the main activity measures (e.g. whether participated at least twice in the last 28 days, or at least 150 minutes of moderate intensity activity in the last 28 days) the base is all participants. If there are missing data on one of the activities, this is just treated as not having done the activity. This is because there are so many different activities asked about and so many different variables which feed in (number of sessions, minutes and two intensity questions) that if anyone with missing data on one or more of these variables were excluded, there would be a huge number of respondents for whom these key measures could not be calculated. Furthermore, the questionnaire was designed that the absence of a tick for having done the activity in the last year is treated as not having it done the activity, so there are no missing data on whether the activity was done in the last year.

7.3 Back-coding of other sports

7.3.1 Capturing open data on other sports and activities

In the online questionnaire, respondents were offered a list of about 180 different activities to choose from, many of which included several activities within them (e.g. squash and racketball) and some of which appeared multiple times to aid respondents in finding them. Nonetheless, there were still activities which people did over the last year which were not in our list of activities or which they had not found in the list. Also, on the paper questionnaire, only about 50 activities are offered on the list.

Therefore, both the online and paper questionnaire offered a space for respondents to record other activities which they did in the last year and last 28 days and to provide details on the frequency, duration and intensity. In order for these data to feed into the main data set these needed to be coded into the categories of activities. These categories included those provided on the questionnaire as well as a few additional categories for activities which were mentioned but were not in the original list (e.g. high ropes courses).

7.3.2 Coding the answers

A coding scheme was created which included all the answers from the online and postal questionnaires, some additional generic responses (e.g. football where the type was not specified) and new activities not included in the original scheme. In addition, a code was created for instances where multiple relevant activities were included in one other answer so they could feed into overall composites, and codes were created for arts activities and for activities which were not relevant at all (to ensure they did not feed into the composites).

All the answers were brought together in Excel. A VLOOKUP was used to automatically code any answers which were worded exactly as the code was labelled. All remaining other answers were manually coded against the code list. All manually coded answers were added to the master look up list which was used for a VLOOKUP in later rounds of back coding. Over time the master list became longer and the most common activities could be automatically coded using the VLOOKUP, but any which could not were manually coded.

At the end of the process all other answers had been assigned a code which indicated which type of activity they were, which could then be used in the derivation of the participation and composite variables. Once codes had been assigned

to all open-ended responses, the coded data was then merged back into the main dataset. To do this, the data was pulled into an SPSS file, then matched back onto the core data by serial and mode. This combination of matching variables ensured that each case was unique in the coding and the data, thereby making certain cases were matched back correctly.

Once the coded data had been matched back onto the core data, a great deal of care had to be taken to ensure that derived variables correctly captured response data.

If the data were added into raw variables too soon, the derived variables would have been calculated incorrectly. To mitigate against this the other back coded data were treated separately, with derived variables being created for these other variables, as for the standard activity variables. Data was then back-coded as a very last step in the data-processing. This ensure that the duration measure, for example, was created as the sum of the durations of the pre-coded activity and the back-coded activity.

Any capping and imputation of activity lengths (as described below) was also applied to the 'other' codes. This was done by creating the values for each activity, then applying these to the 'other specify' codes, according to the specific activity mentioned.

7.4 Deduplication

As documented in the fieldwork section, reminder letters were sent to all households in which two adults had not responded to the survey. In any survey involving a postal element (whether invitation and/or completion), there is a delay between a survey being completed and returned by the respondent and its being processed on receipt. This introduces the possibility of reminder letters being sent to households where the survey has already been completed by one or two adults. In turn, this can cause duplicate responses, where either a single respondent completes the survey a second time, or where more than two people in a household complete the survey.

As such, there were two potential forms of duplicate response within the final Active Lives survey data: individual level duplicates and household level duplicates. Where this occurred, it was necessary to identify and remove duplicate responses from the final dataset. Because each household was provided with two online logins and up to two paper questionnaires it was possible for a single household to complete four questionnaires. Individual duplicates occurred where the same person completed two questionnaires. Household duplicates occurred when more than two different people in the household completed a questionnaire.

As a first step, all cases were identified where the same serial number had been used more than once. These cases were compared, and, where the same age and gender was provided, were assumed to have been received from the same person. In order to ensure that as much data as possible was kept, online responses were prioritised over paper responses (since the online questionnaire was more comprehensive and allowed for more sophisticated routing).

After individual duplicates had been removed, responses were checked to ensure that there was a maximum of two responses from adults in a single household. Where more than two responses had been received, cases were again prioritised by mode and date, and 'duplicate' responses removed from the data.

Following this process of deduplication, it was assumed that all responses in the dataset were from unique individuals. Even after this process, it was possible that unique responses had been provided using the same survey serial number (where two different people both completed the survey using the same serial – one online and one on paper). For the

sake of clarity, these ‘legitimate’ duplicate responses were sorted by mode and date, and renumbered to ensure that each case in the data had a unique ID. These cases have a final digit in the serial number of 3 or 4 rather than 1 or 2.

7.5 Selection of cases to be included

Sometimes online participants break off before the end of the questionnaire. If this is close to the end, their data can still be useful, since the survey routing will dictate that they must have completed all questions for which they were eligible until that point. Similarly, postal participants sometimes break off, or in other cases miss questions within the questionnaire. This can result in significant quantities of missing data, since routing cannot be enforced within a postal questionnaire. We can include cases which don't have complete data but if all demographic or all activity data are missing then they need to be excluded. Sometimes completely, or almost completely blank postal questionnaires are returned, or questionnaires which have been written across but not filled in.

Survey responses were checked at several stages to ensure that only cases with useful data were included. At the initial data collation stage, the scanning agency removed obviously blank paper questionnaires. Following this, during data processing, rules were enforced for the paper and online surveys to ensure that participants had provided sufficient data to make their data useful. For the online survey, this meant participant had to reach a certain point in the questionnaire for their data to count as valid. Paper data was judged by the completion of specific key questions that were vital for survey weighting and analysis.

7.6 Checking of combined final file

7.6.1 Checking raw data

Once the final data had been collated and all illegitimate responses removed, further checks were undertaken. Firstly, all variables were checked to ensure that the number and patterns of response tallied with what was expected based on the survey routing. Once this was confirmed, further sense-checks were conducted to ensure that the broad pattern of responses made sense against what might be expected.

When processing postal responses, questionnaire serials are occasionally incorrectly entered. This can be caused by the barcode being smudged, causing the scanning software to read it incorrectly, or by the barcode being completely obscured (sometimes by the participant) such that manual entry of the serial is required, with a mistake then being made by the data entry clerk. It was necessary to ensure that all cases had a valid serial, so that important demographic information could be matched back into the final datafile. This was achieved by firstly comparing the final datafile against the original sample to identify cases with invalid serials. These cases were then rechecked against the sample in order to find the closest matching ‘genuine’ serial. Following that, a final sense check was carried out to ensure that the suggested matching serial was sufficiently similar to the scanned serial. In all cases, either digits had been transposed, or a typographic error had been made (e.g. a visually similar number was inserted). There was one case which could not be matched to a sample serial: this case was dropped from the data.

7.6.2 Checking final survey weights

After the survey weights had been created the following checks were carried out:

- that the weighted profile of respondents matches the weighting targets as closely as possible;

- that the range of weights is not excessive (note that the sample design for this study means that there are large weights);
- that the weights correspond correctly to the mode, phase and group for that case;
- that every case has a weight >0 for each of the weighting variables where relevant; and,
- that the weighted analysis is different from the unweighted analysis but that the difference is of the scale and direction expected.

7.7 Imputation and cleaning

Early analysis of the data from the first few months of the year highlighted some common errors made by participants. These included:

- Putting time spent on activities for the whole month rather than a session;
- Providing a figure greater than 28 days for number of days participated in an activity in the last four weeks;
- Putting *minutes* in the *hours* box (e.g. 20 hours instead of 20 minutes);
- Not answering all the questions needed to create some of the key participation variables (e.g. providing days and missing hours and minutes);
- Double counting activities, especially in the gym (e.g. recording that they did a combined session as well as recording each individual activity within it);
- Recording the same activity in “other” and under the provided code (which may be correct) which means even in online data, when combined it can generate high numbers of days in a month spent on one activity when backcodes are put back into the main code for an activity;
- Recording a height and weight combination which results in an improbable BMI.

These errors were handled in various points of data processing.

Some of these errors were corrected in initial edits (e.g. answers of >23 hours per session are not permitted, and in manual checking some obvious mistakes can be corrected). However, some of these errors cannot be rectified by manual checks of raw data since they only manifest themselves in creating derived variables, or because there is no clear way to correct them. At the same time, because the activities build across composite variables, it was necessary to take action to manage extreme values. It was also important to compensate for missing data at key variables, because if every case with missing or extreme data on one of the components feeding into the key sports participation variables was excluded there would have been too much missing data. Nonetheless there are some missing data which could not be compensated for (e.g. missing days): where days are missing it was assumed that time spent doing the activity in the month was 0.

After a period of testing against early data and discussion between Sport England and Ipsos MORI, a protocol was agreed for dealing with these issues:

- Where people provided information about the frequency with which they undertook an activity, but did not provide sufficient information to calculate the duration of session (time spent on an activity per day) the session value was imputed as the median session duration for that activity.
- Extreme session values were capped to remove outliers. The upper limit was set as the 95th percentile for durations for each activity except where there were insufficient cases to calculate the percentile. In this instance, the 95th percentile for a similar activity was used as the cap limit.
- In order to handle the same gym or fitness activities being entered twice (under a combined gym session, and under each individual activity done within the session), a series of flags was created which suggested duplication of gym activities:
 - If all session lengths add up to the sum of the combined session
 - If total days across all fitness activities sum to more than 28
 - If the number of days equal the number of days for the combined session
 - Where both combined sessions and individual activities are reported and the individual activities are short (say less than 15 minutes)

Where an individual reported a combined gym session *and* also reported completing at least two individual activities *and* triggered at least two rules from the list above, then it was assumed that they had duplicated their responses. In these cases, individual sessions data were removed in creating the derived variables so that just the combined gym session remains (for all measures involving data in last 28 days – weekly participation, twice monthly, frequency, duration, MEM28 and MEM7 – but not for participation in the last 12 months). These corrections were applied to all paper questionnaire data and to online data from phase 1. At phase 2 the online questionnaire was changed slightly to minimise the risk of duplication and so this process was not necessary

7.8 Creation of derived variables

7.8.1 Activity derived variables

During the processing of the final data, 'derived variables' were created. These variables combine data from multiple questions to create activity level measures of participation. These variables were created using SPSS syntax to calculate the duration, frequency and intensity with which people participate in activities. These variables were then used to create headline measures of activity, such as twice monthly participation,

7.8.2 Creation of composite sports

Once derived variables had been created at activity level, they were then aggregated up to create measures of participation spanning whole sports or groups of sports. This was done by summing the measures for each individual activity (or, for binary Yes/No measures, by counting a mention of Yes at any individual activity). For example, data for the following activities was combined into a single code of 'Racket sports':

- Tennis;

- Badminton;
- Squash & Racketball; and
- Table Tennis.

Overall measures of activity were also created. These measures reflected the Chief Medical Officer's recommendations on activity levels, and aggregated minutes, frequency, and intensity of activity per week across *all* activities, grouping individuals into those who are active, fairly active and inactive.

7.8.3 Demographic derived variables

In addition to the activity and composite measures, demographic and geographic variables were created from the raw questionnaire data. These included, for example, variables grouping age, ethnicity, physical impairment, socio-economic classification and educational status. Other individual level health measures were also created, including BMI score and portions of fruit and vegetables eaten the day before the participant completed the questionnaire.

7.9 The checks on the derived variables

Once the derived variables had been created, a variety of checks were performed to ensure that they had been calculated correctly.

7.9.1 Checking activity and sports measures

The main activity related derived variables are created for multiple composite activities (as described above). Initially, checks were carried out at the activity level. Firstly, the hard logic of the syntax used to derive each measure was checked against the specification. Next, selected activities were tested by cross-tabulating the raw (source) variables against the derived variables to confirm that the data matched as it should. Finally, multiple scenarios were tested to ensure that more complex questionnaire responses did not 'break' the routing. This included checking cases where data for specific input variables was missing, out of range or had been imputed or capped. In addition, checks were conducted to ensure that other answers were feeding into participation measures correctly.

The composite sports variables were only created once it was confirmed that the individual activity variables had been derived correctly. Checks were also needed to ensure that the correct activities fed into each composite, which would then be used for multiple participation variables. Primarily, the SPSS syntax was checked against the specification (which was itself checked and signed off by the Sport England team) to ensure that composite variables were defined correctly. It was not possible to check every single participation measure for each composite. Instead, for one participation measure (twelve monthly participation), all the composites were checked to ensure that all cases where a participant had mentioned an individual activity were counted towards the composite variable. Then for a single composite, all the participation measures were checked to ensure they had been created correctly.

Comparisons were made between different participation measures to check that the way in which they related was consistent with how they had been defined. Where inconsistencies were found, these were investigated and corrections made.

Where problems were found, the syntax was corrected, the variables recreated and the checks repeated to ensure that the final data were correct.

7.9.2 Checking demographic variables

Demographic variables were checked primarily by cross tabulation of the raw variables against the derived variables. A sense check was applied to variables to ensure that the frequencies ‘looked’ right – for example by checking IMD quartiles against Local Authority. Finally, the demographic variables were checked against each other to ensure that they were internally consistent. This included checking that age bands tallied across variables and that derived variables which used the same source data contained the same number of valid responses.

Experiments in the design

8 Experiments in the design

8.1 Response related experiments

Various experiments were conducted throughout year 1 of the Active Lives survey to try and improve the response rate. The design, methodology and outcome of these experiments are discussed over the next chapter.

8.1.1 Strapline on envelopes and the addition of logos to the envelopes

During phase 1 cognitive testing of the questionnaire and survey materials it was noted by a number of participants that blank envelopes made the invitation look like a circular or something that had '*come from an estate agent...*'. Participants stated that they would not open an envelope and would simply dispose of it. There was a suggestion of using a tag line on the envelope to offer an element of interest and enticement, so the participant would open the envelope. It is not uncommon to see a tag line on direct mail and it is a method that has been tried in other surveys administered by Ipsos MORI.

Various messages were proposed which could be categorised into four main groups:

- 1) Those that had a general appeal
- 2) Those that appealed for help which were thought to be persuasive
- 3) Those offering respondents a chance to voice their opinion, coupled with an appeal for help
- 4) Those with a clear reference to an opportunity for the recipient. This drew on the theory that there may be a perceived sense of regret and a lost opportunity if the participant did not open the envelope.

The messages were first tested internally and then refined. After this they were tested during the second round of cognitive interviews for phase 1.

Some of the messages offered pointers to the content of the letter referencing local projects, leisure activities or an opportunity to benefit the local area or community. For some participants, these types of messages were too broad and were not easily linked to the content of the survey. Furthermore, referring to leisure activities was associated with leisure centres or swimming pools – which would put some people off opening the survey, as it was not perceived to be relevant to them.

There were no messages which were clearly more favourable to participants, rather messages seemed to put participants off opening the envelope. '*Have your say...*' did offer a sense of interest and did not put people off opening the envelope. It was concluded that this message would be used on the initial invitation and all three reminders.

8.1.2 Addition of logos to the envelopes

Following the disappointing return rate in Month 1 of the survey, it was decided to print the logos of Sport England, Public Health England and Arts Council England on the front of the envelopes of the first and third reminder letters (mailings two and four). It was hypothesised that the logos would make the letters look more authoritative and encourage recipients to open the envelope and read the content.

The logos had a significant impact on the final response rate; 18.3% in wave 2 compared to 15.0% in wave 1. In wave 3, the logos were printed on the envelope of the initial invitation, as well as on the first and thirds reminders (they were added to the envelope of the second reminder at wave 7). The response rate went up further to 19.9% in wave 3.

Envelopes were pre-printed with the tag line above the envelope window with the logos printed in the top left hand corner of the envelope, above this. We ran out of envelopes that were pre-printed with the ‘*Have your say...*’ tag line prior to Wave 6 launching. It was deemed that in combination with the logos (which appeared to benefit response rate) the tag line gave a crowded feel to the envelope and that it was perhaps superfluous. Hence, from wave 6 the envelopes were printed with only the logos.

8.1.3 Named sample

There is ample evidence showing that personalisation increases the response rate of postal surveys.¹⁸ We therefore conducted a split-sample experiment in waves 4 and 5 to test the impact of matching names from the Edited Electoral Register to the sampled addresses. Names were successfully matched to 38% of sampled addresses.

The matched names were used to personalise the letters for those addresses that had been randomly allocated to the treatment group. Allocation was based on simply whether the 13th digit of the household ID was odd (treatment) or even (control). The control group included only cases where a name had been successfully matched.

Table 8.1: Named sample experiment results – address return rate

	Wave 4	Wave 5
Control	17.9%	18.2%
Treatment	18.3%	18.2%

As the table above shows, contrary to expectations there was no significant difference in address response rates, between the treatment group and the control group. There was reason to believe that the logos on the envelopes had the same impact personalisation would have had without logos; i.e. encourage the recipients to open the envelope. In other words, matching names did not have any added impact when using logos.

8.1.4 Web address

One of the barriers to completion is the amount of effort that potential respondents have to use in order to access the survey. The URL for the survey is www.activelivessurvey.org. It was thought this might be too long and that if the number of key strokes, and hence effort, to access the survey is reduced then this may increase the number of people that go online to take part in it.

The number of URLs available that were shorter than the one already in use was very limited. Of the URLs available it was agreed that www.activesurvey.org should be used in a split sample experiment in order to test if shortening the URL had any benefit on address response rate.

In this experiment, 50% of the sample received letters showing the original URL (www.activelivessurvey.org) and the other 50% of the sample received letters containing the shortened URL (www.activesurvey.org). This experiment was conducted

¹⁸ Don A. Dillman, Virginia Lesser, Robert Mason, John Carlson, Fern Willits, Rob Robertson & Bryan Burke. (2007). Personalization of Mail Surveys for General Public and Populations with a Group Identity: Results from Nine Studies. *Rural Sociology*, 72(4): 632–646.

from wave 7 to wave 9 of the survey. The households which had an odd digit at the 14th number of the ID received letters containing the original URL and those containing an even number, at the same place, received the shortened URL in their invitation and reminder letters. Both URLs were setup to automatically forward to the main survey website

Table 8.2: Online address return rate by wave and URL type

	Normal URL	Short URL	Overall
Wave 7	8.6%	8.1%	8.3%
Wave 8	9.0%	8.6%	8.8%
Wave 9	9.4%	9.2%	9.3%
Total	9.0%	8.7%	8.8%

Table 8.2 shows the online address return rate for those that received letters with the normal URL and those that received letters with the shortened URL. This shows that survey participation did not increase among those who were sent the short URL, with it in fact being slightly higher for those who received the normal URL. The decision was taken to revert to the original URL across the whole of the sample.

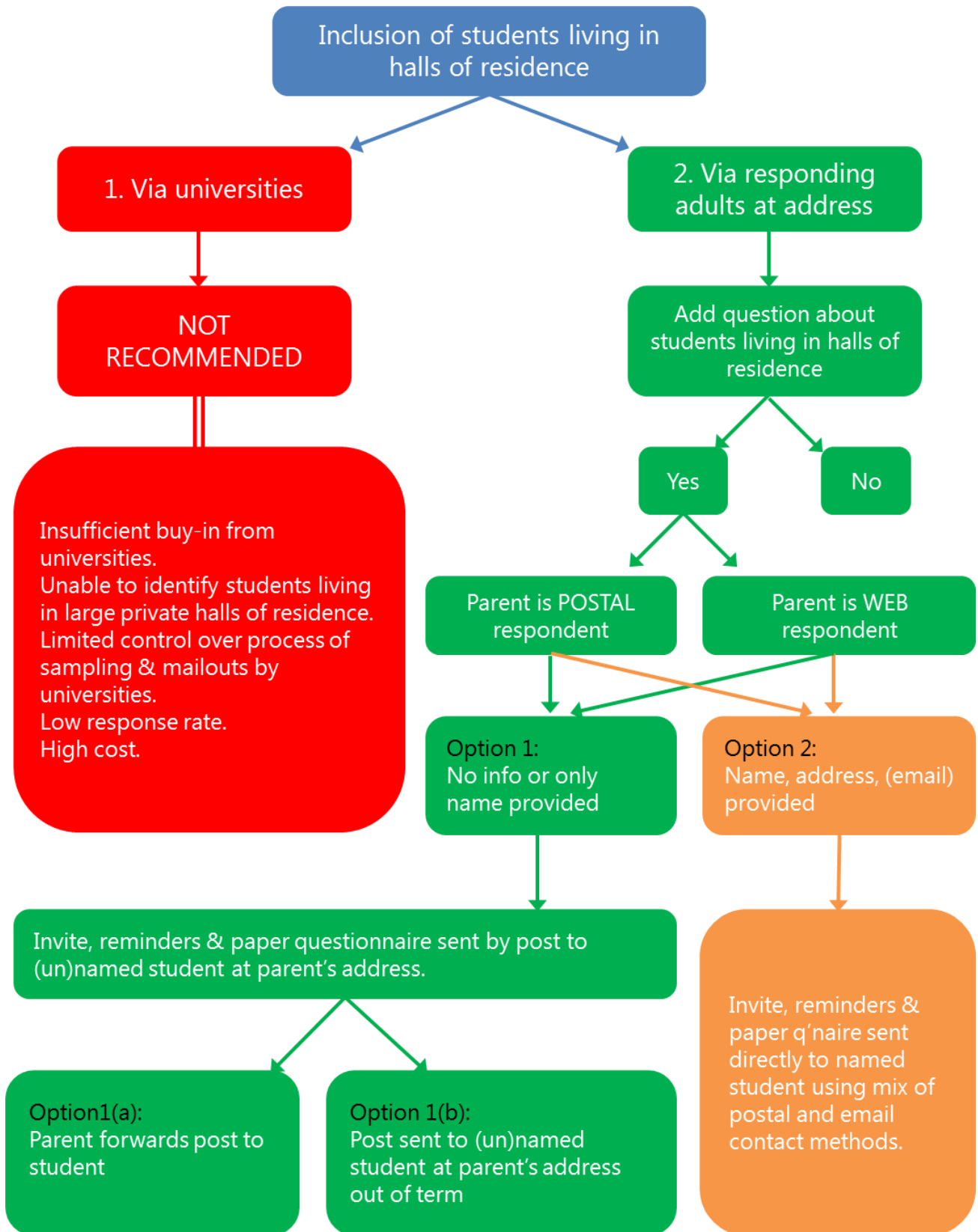
8.2 Student survey experiment

Halls of residence tend not to be listed in the small user Postcode Address File (PAF). Consequently, students living in halls of residence during term time would only have a chance of being included in the survey if they were at home at the time the survey request was made. The resulting under-representation of students was likely to bias the survey estimates of sport participation. This section outlines the design and delivery of the student element of the Active Lives survey. It is presented with technical details followed by findings and the recommendation that was given for the future of the survey.

8.2.1 Design

During initial setup meetings Sport England raised the issue of needing to collect data for the student population. Ipsos MORI understood this need but noted the inherent difficulties with contacting students in halls of residence as they are not included in the Postal Address File (PAF), the main sampling frame for Active Lives. The different approaches considered for contacting this population are outlined in figure 8.1

Figure 8.1: Approaches considered for contacting students



After considering options for administering the survey it was decided the most practical method of contacting students was through parents who had completed the main survey. In order to achieve this the final section of the online adult survey included questions to identify students living in halls and asking for the following details of any children living away from home in halls of residence:

- Consent to contact student
- Student name
- Student address
- Student email

8.2.2 Sampling

In some cases, consent to contact a student living in a hall of residence was given but no details for the student was provided. This dictated that rather than contacting students directly we had to send survey materials via the parent postal address and include stamps to send on to students. Ultimately this meant that for sampling purposes students had to be categorised into four groups:

Group 1: No contact details (parent address would be used for contact)

Group 2: Student postal address only

Group 3: Student email address only

Group 4: Student email address and student postal address

Fieldwork for the student survey was conducted over 3 waves from 1 March 2016 – 28 June 2016. During this period, we acquired consent to contact 924 students, the majority of whom we had no specific details for and had to go via the parents' home address. The full sampling breakdown by group is provided in table 8.4

Table 8.3: Showing the breakdown of the student survey

Group	Details	Total
Group 1	No contact details	715
Group 2	Student postal address only	12
Group 3	Student email address only	76
Group 4	Student email address and student postal address	121
Total		924

Given the variation in information available, different methods for contacting each group were adopted. This being more complex than in the adult survey due to the inclusion of extra envelopes, stamps and instructions for parents/students. The various sampling combinations and the materials sent to each group are outlined in the table below.

Table 8.4: Materials provided to each group in the student survey

Study group	Information	Mailing 1	Mailing 2	Mailing 3	Mailing 4
1	No contact details: The same mailing strategy as for the main sample (up to 4 reminders, with a paper questionnaire sent in the 3 rd mailing) with request to forward to named student.	DL for student inside C5 addressed to parents.	DL for student inside C5 addressed to parents.	Questionnaire mailing. C3 for students inside slightly larger envelope for parents.	DL for student inside C5 addressed to parents.
	Posted to - Parent address	<ul style="list-style-type: none"> • C5 • M1 parent slip • DL envelope with stamp on • Student letter 	<ul style="list-style-type: none"> • C5 • M2 parent slip • DL envelope with stamp on • Student letter • Stamp 	<ul style="list-style-type: none"> • C4 • M3 parent slip + Questionnaire • Folded C4 • Student letter • Questionnaire • Folded C4 BRE • Stamp 	<ul style="list-style-type: none"> • C5 • M4 parent slip • DL envelope • Student letter • Stamp
2	Postal address only: The same mailing strategy as for the main sample will be adopted (up to four reminders, with a paper questionnaire sent in the 3 rd mailing), sent to the halls of residence.	Posted to student. DL for student	Posted to student. DL for student	Questionnaire posted to student C3 for students	Posted to student. DL for student
	Posted to - Student address	<ul style="list-style-type: none"> • DL • M1 letter 	<ul style="list-style-type: none"> • DL • M2 letter 	<ul style="list-style-type: none"> • C4 • M3 letter + Questionnaire • Folded C4 BRE • Questionnaire 	<ul style="list-style-type: none"> • DL • M4 letter
3	Email address only: Up to 2 emails will be sent to the named student followed by two postal reminders sent to the address of the main respondent with request to forward to named student. A paper questionnaire will be sent with the final postal reminder.	Email	Email	Posted to parents DL for student inside C5 addressed to parents.	Questionnaire Posted to parents. C3 for students inside slightly larger envelope for parents.
	Posted to - Student email and parent address	<ul style="list-style-type: none"> • M1 email 	<ul style="list-style-type: none"> • M2 email 	<ul style="list-style-type: none"> • C5 • M3 parent slip (invite) • DL envelope with stamp • Student letter 	<ul style="list-style-type: none"> • C4 • M4 parent slip + Q • Folded C4 BRE • Student letter • Questionnaire • Folded C4 envelope • Stamp
4	Email address and postal address: Up to two emails followed by two postal reminders will be sent to the named student. A paper questionnaire will be sent with the final postal reminder.	Email	Email	Posted to student (letter only). DL for student	Questionnaire posted to student. C3 for student
	Posted to - Student email and student address	<ul style="list-style-type: none"> • M1 email 	<ul style="list-style-type: none"> • M2 email 	<ul style="list-style-type: none"> • DL • M3 letter (invite) 	<ul style="list-style-type: none"> • C4 • M4 letter + Questionnaire • Folded C4 BRE • Questionnaire

8.2.3 Findings and recommendations

Overall we achieved 195 returns from a mailout of 924 (160 online vs 35 offline) representing a response rate of 21.1%. While the response rate was broadly in line with the adult survey, the cost per complete was more than ten times higher for the student survey.

Given the small number of returns of approximately 60 per student survey wave, the base sizes for the student population were very small and the opportunity for any meaningful analysis was very limited.

The high cost-per-return was largely due to a minimum charge for small print runs, the lack of opportunity for cost-efficiencies, the higher administrative burden and complexity of matching multiple materials as well as the additional cost of extra stamps required for mail to reach the target audience.

Logistically the survey also presented other challenges, particularly in terms of the resource required to sign off materials and manually create sample for each mailing. On this basis we believed the cost for the survey to be prohibitively high for the number of returns achieved and it was not recommended that it be continued. The survey was discontinued after the third wave.

8.3 Progress bar/ progress text experiment

At the request of Sport England, a progress bar was added to the online survey during wave 2 of the survey on the 22 December 2015. A split sample experiment was conducted in order to see if there is any negative impact of the progress bar on break off or response rate. There was found to be none and a progress bar was rolled out across the entire sample from wave 3.

The progress bar was not an interactive progress bar. Rather it was four images which showed progress through the survey by addition of sections.

Figure 8.2: Example of the progress bar indicator



A further split sample experiment was conducted from wave 7 to wave 9.

As the routing for the survey is complex and the length of the survey is largely dependent on questions answered previously, it was reasoned that the progress bar may not always reflect the true length of the survey and this may be a partial cause of participants dropping out of the survey. Internal professional designers were involved in designing a new progress indicator and together with the survey team developed the progress text indicator.

In this experiment half the sample received the standard progress bar and half received the progress text. An experiment was conducted to see if this had any bearing on the response rate or the break off rate.

The progress text indicator simply stated which section of the questionnaire a participant was on and did not show any picture or graphic. While not a pure progress bar / indicator it perhaps gave participants a better indicator of where they are in the survey, compared to the progress bar.

Figure 8.3: Example of the progress text indicator

Section 2 of 4

Sport, fitness and recreation

–

Table 8.5: Online return rate (number of online questionnaires completed / number of households surveyed) by progress indicator

	Wave 7	Wave 8	Wave 9	Overall
Progress bar	10.99%	11.43%	12.06%	11.49%
Progress text indicator	10.75%	11.57%	12.10%	11.47%
Overall	10.87%	11.50%	12.08%	11.48%

Table 8.5 shows that there was no difference in participation rates between those who had the progress bar and progress bar text indicator. Analysis was also conducted on the break off rates and there was found to be no significant difference between the two. As such, it was decided to revert to using one progress indicator across the whole sample; with the progress text indicator used.

Appendices

Appendices

Geographical breakdown of Year 1 survey responses

Local Authority	Targets	Achieved
Adur	500	514
Allerdale	500	517
Amber Valley	500	536
Arun	500	508
Ashfield	500	495
Ashford	500	505
Aylesbury Vale	500	520
Babergh	500	498
Barking and Dagenham	1,000	1,036
Barnet	1,000	1,009
Barnsley	500	525
Barrow-in-Furness	500	504
Basildon	500	525
Basingstoke and Deane	500	504
Bassetlaw	500	514
Bath and North East Somerset	1,000	991
Bedford	500	493
Bexley	500	517
Birmingham	2,000	2,020
Blaby	500	495
Blackburn with Darwen	500	506
Blackpool	500	501
Bolsover	500	515
Bolton	1,000	1,009
Boston	500	484
Bournemouth	500	505
Bracknell Forest	500	507
Bradford	1,000	1,069
Braintree	500	498
Breckland	500	506
Brent	500	516
Brentwood	500	508
Brighton and Hove	500	500
Bristol, City of	2,000	2,045
Broadland	500	504
Bromley	500	519
Bromsgrove	500	508
Broxbourne	500	511
Broxtowe	500	537
Burnley	500	532
Bury	1,000	1,026
Calderdale	500	520
Cambridge	750	763
Camden	500	519
Cannock Chase	500	498
Canterbury	500	506

Local Authority	Targets	Achieved
Carlisle	500	497
Castle Point	500	514
Central Bedfordshire	500	522
Charnwood	500	504
Chelmsford	500	492
Cheltenham	500	510
Cherwell	500	503
Cheshire East	500	505
Cheshire West and Chester	1,000	1,037
Chesterfield	500	510
Chichester	500	489
Chiltern	500	507
Chorley	500	516
Christchurch	500	515
City of London	250	267
Colchester	500	499
Copeland	500	518
Corby	500	512
Cornwall	1,000	1,029
Cotswold	500	507
County Durham	1,000	972
Coventry	1,000	980
Craven	500	501
Crawley	500	520
Croydon	500	517
Dacorum	500	491
Darlington	500	525
Dartford	500	491
Daventry	500	504
Derby	1,000	1,044
Derbyshire Dales	500	514
Doncaster	500	492
Dover	500	520
Dudley	1,000	998
Ealing	500	502
East Cambridgeshire	500	488
East Devon	500	520
East Dorset	500	513
East Hampshire	500	506
East Hertfordshire	500	510
East Lindsey	500	496
East Northamptonshire	500	526
East Riding of Yorkshire	500	504
East Staffordshire	500	514
Eastbourne	500	505
Eastleigh	500	511

Local Authority	Targets	Achieved
Eden	500	519
Elmbridge	500	524
Enfield	1,000	1,025
Epping Forest	500	503
Epsom and Ewell	500	537
Erewash	500	513
Exeter	500	497
Fareham	500	534
Fenland	500	510
Forest Heath	500	491
Forest of Dean	500	499
Fylde	500	505
Gateshead	500	509
Gedling	500	499
Gloucester	500	506
Gosport	500	509
Gravesham	500	488
Great Yarmouth	500	490
Greenwich	1,000	1,035
Guildford	500	507
Hackney	500	512
Halton	500	500
Hambleton	500	490
Hammersmith and Fulham	500	522
Harborough	500	506
Haringey	1,000	1,043
Harlow	500	510
Harrogate	500	539
Harrow	500	519
Hart	500	500
Hartlepool	500	504
Hastings	500	509
Havant	500	507
Havering	500	499
Herefordshire, County of	500	501
Hertsmere	500	491
High Peak	500	512
Hillingdon	500	490
Hinckley and Bosworth	500	516
Horsham	500	560
Hounslow	500	527
Huntingdonshire	500	513
Hyndburn	500	497
Ipswich	500	496
Isle of Wight	500	490
Isles of Scilly	250	247
Islington	500	512
Kensington and Chelsea	500	505
Kettering	500	533
Kings Lynn and West Norfolk	500	510
Kingston upon Hull, City of	500	524
Kingston upon Thames	500	516
Kirklees	1,000	1,005

Local Authority	Targets	Achieved
Knowsley	500	511
Lambeth	1,000	1,001
Lancaster	500	508
Leeds	2000	2,041
Leicester	1,000	1,005
Lewes	500	522
Lewisham	500	500
Lichfield	500	510
Lincoln	500	492
Liverpool	2,000	2,022
Luton	1,000	1,002
Maidstone	500	477
Maldon	500	524
Malvern Hills	500	495
Manchester	2,000	2,035
Mansfield	500	525
Medway	500	506
Melton	500	495
Mendip	500	512
Merton	500	530
Mid Devon	500	513
Mid Suffolk	500	511
Mid Sussex	500	512
Middlesbrough	500	504
Milton Keynes	1,000	1,024
Mole Valley	500	511
New Forest	500	515
Newark and Sherwood	500	471
Newcastle upon Tyne	2,000	2,064
Newcastle-under-Lyme	500	516
Newham	500	492
North Devon	500	496
North Dorset	500	526
North East Derbyshire	500	513
North East Lincolnshire	500	509
North Hertfordshire	500	503
North Kesteven	500	529
North Lincolnshire	500	514
North Norfolk	500	499
North Somerset	500	497
North Tyneside	500	507
North Warwickshire	500	513
North West Leicestershire	500	527
Northampton	500	508
Northumberland	1,000	1,021
Norwich	750	764
Nottingham	2,000	2,016
Nuneaton and Bedworth	500	512
Oadby and Wigston	500	508
Oldham	1,000	1,000
Oxford	750	776
Pendle	500	526
Peterborough	1,000	1,047

Local Authority	Targets	Achieved
Plymouth	1,000	1,046
Poole	500	503
Portsmouth	1,000	1,015
Preston	500	511
Purbeck	500	524
Reading	1,000	1,003
Redbridge	1,000	992
Redcar and Cleveland	500	522
Redditch	500	499
Reigate and Banstead	500	494
Ribble Valley	500	517
Richmond upon Thames	500	514
Richmondshire	500	507
Rochdale	1,000	1,010
Rochford	500	519
Rossendale	500	505
Rother	500	492
Rotherham	500	493
Rugby	500	484
Runnymede	500	514
Rushcliffe	500	502
Rushmoor	500	522
Rutland	500	497
Ryedale	500	508
Salford	1,000	1,011
Sandwell	1,000	975
Scarborough	500	517
Sedgemoor	500	524
Sefton	500	517
Selby	500	516
Sevenoaks	500	507
Sheffield	2,000	2,057
Shepway	500	492
Shropshire	500	522
Slough	1,000	1,065
Solihull	500	495
South Bucks	500	487
South Cambridgeshire	500	506
South Derbyshire	500	510
South Gloucestershire	1,000	1,018
South Hams	500	510
South Holland	500	516
South Kesteven	500	508
South Lakeland	500	526
South Norfolk	500	518
South Northamptonshire	500	487
South Oxfordshire	500	499
South Ribble	500	510
South Somerset	500	510
South Staffordshire	500	497
South Tyneside	500	501
Southampton	1,000	1,021
Southend-on-Sea	1,000	996

Local Authority	Targets	Achieved
Southwark	500	514
Spelthorne	500	507
St Albans	500	508
St Edmundsbury	500	508
St. Helens	500	518
Stafford	500	498
Staffordshire Moorlands	500	510
Stevenage	500	532
Stockport	1,000	985
Stockton-on-Tees	500	496
Stoke-on-Trent	1,000	1,016
Stratford-on-Avon	500	507
Stroud	500	518
Suffolk Coastal	500	503
Sunderland	1000	1,053
Surrey Heath	500	498
Sutton	500	513
Swale	500	497
Swindon	1,000	1,031
Tameside	1,000	998
Tamworth	500	504
Tandridge	500	493
Taunton Deane	500	508
Teignbridge	500	507
Telford and Wrekin	500	503
Tendring	500	509
Test Valley	500	516
Tewkesbury	500	517
Thanet	500	475
Three Rivers	500	490
Thurrock	1,000	1,023
Tonbridge and Malling	500	501
Torbay	500	515
Torridge	500	507
Tower Hamlets	500	499
Trafford	1,000	1,024
Tunbridge Wells	500	512
Uttlesford	500	527
Vale of White Horse	500	501
Wakefield	1,000	1,005
Walsall	1,000	1,031
Waltham Forest	500	504
Wandsworth	500	494
Warrington	1,000	1,014
Warwick	500	495
Watford	500	502
Waveney	500	499
Waverley	500	490
Wealden	500	518
Wellingborough	500	489
Welwyn Hatfield	500	509
West Berkshire	500	521
West Devon	500	491

Local Authority	Targets	Achieved
West Dorset	500	518
West Lancashire	500	502
West Lindsey	500	496
West Oxfordshire	500	503
West Somerset	500	499
Westminster	500	498
Weymouth and Portland	500	506
Wigan	1,000	1,081
Wiltshire	500	515
Winchester	500	500
Windsor and Maidenhead	500	500
Wirral	500	534
Woking	500	506
Wokingham	500	514
Wolverhampton	1,000	989
Worcester	500	515
Worthing	500	500
Wychavon	500	512
Wycombe	500	528
Wyre	500	525
Wyre Forest	500	492
York	500	515
Total	198,250	201,579

