Taking Part: England's Survey of Leisure, Culture and Sport (Year 4, 2008-09)

Technical Report (Final)

Prepared for: DCMS

Prepared by: Joel Williams/Angela Charlton BMRB Social Research
Telephone: 020 8433 4417 Email: joel.williams@bmrb.co.uk
Part of BMRB Limited (British Market Research Bureau) BMRB/JW/45107321 -324

BMRB is ISO9001:2008 and ISO 20252 accredited.

Printed on 100% recycled paper

BMRB is also a member of the London Remade environment scheme.

Table of Contents

Appendices

Introduction

This report outlines the methods used for Year 4 of the *Taking Part* survey. The sample for this survey was issued on a quarterly basis, starting with July 2008. The Year 4 sample comprises the July 08, Oct 08, Jan 09 and Apr 09 samples.

However, part way through the work, the data year was switched to cover the Apr 08, July 08, Oct 08 and Jan 09 samples. The Apr 08 sample design is described in the technical report covering Year 3 since it was initially considered the fourth quarter of Year 3.

This report covers the July 08, Oct 08, Jan 09 and Apr 09 samples in terms of design, fieldwork, fieldwork outcomes and coding. However, the section covering *adult* weights and design factors refers to the 'new' Year 4 dataset, covering the Apr 08, July 08, Oct 08 and Jan 09 samples. The section covering *child* weights and design factors refers to the 'old' Year 4 dataset, covering the July 08, Oct 08, Jan 09 and Apr 09 samples.

This report has been written by the BMRB project team at BMRB – Joel Williams (Project Consultant), Angela Charlton (Project Manager), Michael Potter (Senior Research Executive) and Stephen McGinigal (Research Executive).

1 Sample Design

1.1 Survey population and sample frame

The survey was initially designed to yield a representative sample of 14,000 adults aged 16+ and of ~1,200 children aged 11-15 who are normally resident in England. Part way through (from Sept 08 onwards), the child population was extended to cover those aged 5-10 with data supplied by parental proxy.

For practical purposes, residents of institutional accommodation (armed forces barracks, student halls of residence, hospitals, care homes, prisons etc.) were excluded.

BMRB utilised the residential Postal Address File (PAF) as the sample frame. This provides a list of almost all private residential addresses in the UK and is the most comprehensive frame available. Because it lists addresses, not individuals, interviewers were required to randomly select respondents from among those eligible.

1.2 Selection of Primary Sampling Units

1.2.1 Different Primary Sampling Units based on population density

A new sample design was adopted for Year 4 in which the primary sampling unit *varied* based on the local area's population density. The objective was to loosen the level of clustering in the most densely populated areas while tightening it in the least densely populated areas. The theory was that the statistical cost due to clustering would be reduced significantly in the most densely populated areas but not increased a great deal in the least densely populated areas[1].

A formula was developed to ensure that:

- approximately one third of PSUs would be 'double' Medium Layer Super Output Areas (MSOAs);
- approximately one third of PSUs would be single MSOAs
- approximately one third of PSUs would be paired Lower Layer Super Output Areas (LSOAs).

'Double' MSOAs were formed by linking MSOAs with adjacent ONS codes. For the most part, those with adjacent ONS codes were also geographically adjacent, although this did not hold in every case. Two further rules were applied: no

[1] An analysis of the results will follow in a later paper.

doubles could be formed that crossed local authority boundaries and no single MSOAs could be left unpaired. These singles were attached to adjacent doubles to form trio MSOAs. Census-derived data used for stratification was computed for each of these new PSUs.

Once this stage was complete, a measure of address density was formed. Following previous convention with this survey, 30 addresses were to be issued per PSU everywhere except London where 42 were to be issued per PSU. From this a 'selected addresses per square kilometre' value was computed for every 'double' MSOA. Where this value was greater than 1.50, the double MSOA would be used as the PSU. Where this value fell between 0.35 and 1.49, the standard single MSOA would be used as the PSU. Where this value fell below 0.35, standard single MSOAs would be sampled but a second sampling stage would take place: two LSOAs would be sampled from those within the sampled MSOA.

This design ensured an even division between the three PSU types and an expected average of 8.91 selected addresses per square kilometre, a little less than in previous editions but not by a large enough margin to make a major impact on costs.

1.2.2 PSU stratification

Before BMRB sampled the PSUs, the list of PSUs was stratified into 19 geographic areas: the 9 English regions (minus any local authorities in the Top Ten in terms of population) plus the 10 most populous local authorities which counted as separate strata: Birmingham, Leeds, Sheffield, Manchester, Bradford, Liverpool, Kirklees, Bristol, East Riding of Yorkshire, and the Wirral. These ten were selected in order to generate representative – if highly clustered – samples in each and compare the survey estimates[2] with those generated by Sport England's concurrent *Active People* survey where question items were the same or similar.

Other levels of stratification were employed within each of these geographic areas. BMRB used forward stepwise regression techniques to analyse the individual-level associations between demographic values and the relevant performance indicators in the 2005-07 survey data. Where associations were strong, BMRB selected the equivalent PSU-level aggregate value as a stratification variable.

The precise stratification design varied based on the total number of PAF addresses in each geographic area. It was important to ensure that every six month data release contained data from at least two PSUs in every stratum,

[2] Both direct estimates and modelled estimates from RAE Consulting. RAE Consulting 'borrowed' data from the rest of the survey to refine the direct local authority-level estimates.

allowing proper standard errors to be computed. Consequently, the design had to ensure that at least *four* PSUs were sampled per stratum on an annual basis. More recently, DCMS has requested reversion to quarterly datasets so, in future, the degree of stratification will need to be reduced to ensure a minimum of eight PSUs per stratum per year.

Table 1.1 shows the design.

MANPROF = Proportion of residents aged 16+ classified as managerial/professional according to Census 2001

AGED1635 = Proportion of residents aged 16+ aged 16-35 according to Census 2001

ACAT1-3 = Proportion of households classified in ACORN categories 1 to 3 ['wealthy achievers', 'urban prosperity' and 'comfortably off']

Table 1.1 PSU stratification design for Taking Part 2008-9

	Stratification levels		
Region / local authority	**Level 1**	**Level 2**	**Level 3**
North East England	5 percentiles of MANPROF	2 percentiles of AGED1635	Sorted by ACAT1-3 (Implicit)
North West England	6 percentiles of MANPROF	4 percentiles of AGED1635	Sorted by ACAT1-3 (Implicit)
Yorkshire / Humber	5 percentiles of MANPROF	2 percentiles of AGED1635	Sorted by ACAT1-3 (Implicit)
East Midlands	5 percentiles of MANPROF	3 percentiles of AGED1635	Sorted by ACAT1-3 (Implicit)
West Midlands	6 percentiles of MANPROF	3 percentiles of AGED1635	Sorted by ACAT1-3 (Implicit)
East of England	6 percentiles of MANPROF	4 percentiles of AGED1635	Sorted by ACAT1-3 (Implicit)
London	7 percentiles of MANPROF	4 percentiles of AGED1635	Sorted by ACAT1-3 (Implicit)

South East England	8 percentiles of MANPROF	4 percentiles of AGED1635	Sorted by ACAT1-3 (Implicit)
South West England	6 percentiles of MANPROF	3 percentiles of AGED1635	Sorted by ACAT1-3 (Implicit)
Birmingham	4 percentiles of MANPROF	Sorted by AGED1635 (Implicit)	-
Leeds	3 percentiles of MANPROF	Sorted by AGED1635 (Implicit)	-
Sheffield	2 percentiles of MANPROF	Sorted by AGED1635 (Implicit)	-
Manchester	2 percentiles of MANPROF	Sorted by AGED1635 (Implicit)	-
Bradford	2 percentiles of MANPROF	Sorted by AGED1635 (Implicit)	-
Liverpool	Sorted by MANPROF (Implicit)	-	-
Kirklees	Sorted by MANPROF (Implicit)	-	-
Bristol	Sorted by MANPROF (Implicit)	-	-
East Riding of Yorkshire	Sorted by MANPROF (Implicit)	-	-
Wirral	Sorted by MANPROF (Implicit)	-	-

This design produced 197 explicit strata in total plus additional implicit stratification.

1.2.3 PSU sampling method

BMRB sampled the PSUs with a probability proportionate to size (number of delivery points/addresses) using the method of random start and fixed interval.

The design called for a proportionate sample but sampling fractions varied slightly by region to take account of historically different conversion rates (interviews per sampled address) in different regions. BMRB calculated regional conversion rates based on the 2005-06 and 2006-07 editions of *Taking Part*.

In total, BMRB sampled 1,176 PSUs:

- 298 'double' MSOAs,
- 296 single MSOAs, and
- 291 single MSOAs in which two LSOAs were sampled with a probability proportionate to size (= 582 LSOAs).

In each PSU, BMRB sampled 30 addresses after sorting addresses by postcode and house number to maximise the spatial dispersion of the sample. In London 42 addresses were sampled because of its historically relatively low response rate.

BMRB actually drew a sample of PSUs equal to *110%* of the required size. One in eleven of the PSUs was systematically assigned to the reserve sample using the final stratification 'order' and a random start. However, the reserve sample was not required.

Indeed, by January 2009 it became clear that the interview target would be exceeded. Consequently, a small number of PSUs was (randomly) deleted from the Jan 09 sample. A slightly different method was used to reduce the number of interviews from the Apr 09 sample: a small number of addresses was removed from each sampled PSU (two from each PSU outside of London and four from each London PSU). These deletions have been accounted for when computing the data weights[3].

1.2.4 Quarter and Month assignment

The sampled PSUs were sorted using their original stratification values and tagged with a 'fieldwork quarter' label via the 'snaked' allocation system: Q1-Q2-Q3-Q4-Q4-Q3-Q2-Q1-Q1-Q2 etc. but with a random start (e.g. 'Q3').

[3] Strictly speaking, given that a PSU's allocation to any one of the four quarters was random the overall probability of address deletion ought to be the same for all addresses. However, because we observe a fair degree of seasonal variation in the estimates it is important to ensure an (approximately) equal sum of weights for each quarter. Because this adjustment compensates for a design change rather than for non-response it makes sense to deal with it when computing the basic design weight rather than when computing the non-response weight. It is also worth remembering that the adult dataset actually comprises the last quarter of Year 3 and the first three quarters of Year 4 which complicates calculations of the probability of allocation to a specified quarter.

The primary objective was to achieve a representative sample for the twelve month period viewed as a whole but a secondary objective was to achieve a representative sub-sample of the whole in each quarter.

1.2.5 Sampling of individuals

At each sampled address, the interviewer would randomly sample one dwelling unit (if more than one), then randomly sample one household (if more than one) within the sampled dwelling unit. Interviewers used unique Kish Grids assigned to each address to assist them in this process.

The same Kish Grid was also used to randomly sample individuals within the household. Interviews were sought with:

- 1 adult aged 16+
- 1 child aged 11-15 (if resident)

The child interviews did not begin until Sept 08 (the July 08 sample was split into two parts for this purpose). Also, from Sept 08 any parents of 5-10 year olds who were interviewed for the adult survey were asked to provide information about one randomly sampled child in this age range.

2 Fieldwork

All fieldwork was conducted on behalf of BMRB by interviewers trained and supervised by Kantar Operations[4]. In total, approximately 330 interviewers worked on the survey in Year 4.

2.1 Briefings

Before starting work on Taking Part all interviewers attend a face-to-face briefing. These are presented by BMRB researchers and Kantar Operations field staff. The initial briefings held in July and August 2005 of Year 1 were also attended by representatives from the DCMS and/or representatives of the Non-Departmental Public Bodies.

After doing 37 briefings during Year 1, 5 briefings in Year 2 and 2 briefings in Year 3, there were 3 full-day briefings for Year 4. Around 12-15 interviewers attended each briefing.

Each briefing included the following topics:

1. Background and information on the Taking Part Survey and its use by the DCMS.

2. Information about sampling procedures; contact procedures and dwelling/respondent selection; the importance of high response rates, with methods of ensuring contact and encouraging co-operation; and the use of incentives.

3. Description of the questionnaire, and interview procedures, including explanations of the more complex questions and question sequences. Particular attention was paid to the questions used to measure the PSA targets.

4. Group exercise to get interviewers to think of ways to respond to potential refusals on the doorstep.

In addition to attending the face to face briefing, interviewers were also required to read the written Interviewer Instructions and carry out at least two practice interviews before starting their first assignment.

[4] Kantar is the information, insight and consultancy arm of WPP. Kantar Operations manage all aspects of the operational side of research on behalf of BMRB. Further details can be found at http://www.kantaroperations.com.

Interviewers briefed in Year 1, 2 or 3 were not required to attend a further face-to-face briefing. Interviewers were sent an updated set of Interviewer Instructions, highlighting the changes to their assignment in Year 4.

2.2 Supervision and quality control

Several methods were used to ensure the quality and validity of the data collection operation.

A proportion of interviewers, particularly those less experienced, were accompanied in the field by supervisors. All interviewers who were new to random probability sample surveys were accompanied on the first day of a Taking Part assignment by a supervisor.

A proportion of respondents were re-contacted to verify that an interview had taken place. In total, 14% of respondents were re-contacted in Year 4 to verify that the interviewer had contacted someone and whether or not an interview was completed. Addresses for back checking were selected on the basis of Kantar Operations overall field quality procedures, whereby all interviewers have their work checked at least twice a year.

These back checking procedures were mainly carried out by telephone. Where no telephone number was available a short postal questionnaire was sent to the address to collect the same information.

2.3 Fieldwork dates and fieldwork management

Fieldwork was conducted between 28th July 2008 and 9th August 2009.

Fieldwork was managed on a quarterly basis and assignments were issued to interviewers prior to each quarter starting. Table 2.1 shows the number of assignments and core sample addresses per quarter. Quarter 1 was separated into two (a and b), as the year was initially launched without the child survey. This was due to the late confirmation of the child survey and the lack of time available for piloting the new 5-10 and 11-15 child questionnaires before the start of the Quarter 1. Quarter 1b signalled the launch of the new child surveys.

Table 2.1 Assignments and core sample addresses per quarter

Quarter	Number of assignments issued	Number of core sample addresses
Quarter 1a	109	3450
Quarter 1b	111	3510
Quarter 2	221	6990

Quarter 3	212	6720
Quarter 4	222	6516

Interviewers had about 4-5 weeks to cover all the addresses in their assignment and report final outcomes. Interviewers were encouraged to start their assignment as early as possible in fieldwork to try to maximise the time available for making contact at the addresses.

Once all the issued addresses had been covered the Address Contact Sheets were returned to Kantar Operations and a decision was taken about re-issuing non-productive outcomes. As a general rule all non-productive addresses (non-contacts, refusals, broken appointments, etc.) were considered for re-issue unless there was a specific reason not to or it was not considered cost effective (e.g. response rate and interview projections were on track or if only one or two addresses in an assignment were available for reissue). Once the first re-issue period had been completed a decision was taken about whether to re-issue addresses that were still non-productive for a second or third time.

Table 2.2 shows the fieldwork dates for each sample month.

Table 2.2 Fieldwork dates for each sample month

Quarter	Fieldwork start	Fieldwork end (includes re-issue period)
Quarter 1a	28/7/08	16/11/08
Quarter 1b	1/9/08	21/12/08
Quarter 2	20/10/08	22/2/09
Quarter 3	19/1/09	10/5/09
Quarter 4	20/4/09	9/8/09

2.4 Fieldwork procedures and documents

2.4.1 Introductory letters and leaflet

All the core sample addresses were sent an advance letter and a Taking Part respondent leaflet. The letters and leaflets were sent by interviewers a couple of days before starting their assignment.

The letter and leaflet were designed to answer respondents' questions and encourage them to take part. No changes were made to these documents in Year 4 of the survey (having already been revised in the previous years of the

survey to make them more respondent friendly and persuasive, these documents seemed to be working well).

The letters outlined the background to the survey, stressed the importance of the respondent taking part, the confidential nature of the survey and the financial 'thank you' for taking part. The letters were despatched on DCMS headed paper and signed by the project manager at the DCMS to authenticate the survey.

There were also 2 'reissue' letters – one for those addresses where the initial interviewer was unable to make contact at the address and one for those where a refusal had occurred. Both were despatched on BMRB headed paper and signed by the project manager at BMRB.

All letters provided a telephone number and an email address so that individuals could find out more about the survey, make an appointment for an interviewer to call, or opt out of the survey. Over the course of the year, **577** people, representing **2.12%** of addresses issued, opted out of the survey by contacting BMRB, Kantar Operations or the DCMS.

Copies of the letters and the leaflet can be found in Appendix B and Appendix D respectively.

2.4.2 Limited/non speakers of English

In cases where the selected person had limited or no English, interviewers were permitted to use another person to interpret, provided such a person was appropriate (e.g. a close relative). The minimum age for an interpreter was set at 12 years old.

2.4.3 Address Contact Sheets

Each address was issued to the interviewer on a document called the Address Contact Sheet (ACS). The ACS used for the main sample served six main functions:

it contained full address details for the sampled address;

interviewers used it to make random selections of dwelling units and eligible adults;

interviewers used it to complete the screening for the child interview, make the selection of the child and record parental permission to approach the child for interview;

interviewers used it to record the outcome of their attempts to make contact and conduct an interview at the address;

it included the signed receipt of the incentive.

Interviewers made a minimum of eight calls at each address before regarding it as a non-contact, recording details of these on the ACS. Calls had to be made on different days of the week and at different times of day: at least two of the calls had to be made on a weekday evening (after 7.00 p.m.) and at least one call at a weekend (10.00 a.m. – 9.00 p.m.), in order to make contact with households where everyone was working.

An example ACS is included in Appendix C.

2.5 The Child Surveys

The new Taking Part child surveys were launched in Quarter 1b of Year 4, missing the initial launch due to the pilot testing of both questionnaires. Full details of the pilot findings for both the 5 to 10 and 11 to 15 questionnaires can be found in Appendix H.

The 5 to 10 interview was carried out by proxy with the adult respondent if they were the parent of a 5-10 year old. The 11 to 15 interview was carried out with the child, following parental consent being granted.

2.5.1 Screening procedures for the child sample

The child screening was carried out at all addresses in the sample. Where an eligible 5-10 year old and an eligible 11-15 year old were identified the interviewer was instructed to attempt to carry out both extra interviews (a "child interview by proxy" for 5-10 year olds and a "child interview" for 11-15 year olds) at that household.

There were screening instructions for both the 5-10 proxy interview and the 11 to 15 interview on the main address contact sheet, but in order not to jeopardise the adult survey the child screening was left until after the adult interview unless brought up by the respondent. Once the selection of any children aged 11 to 15 had been made, the interviewer was required to obtain written parental permission before proceeding with the interview. The adult was shown the Parental Permission Card (see Appendix C2) to indicate what the interviewer would be asking the child, and asked to sign the "parental/guardian permission" section of the address contact sheet. This was not required with the 5 to 10 proxy interview as this was completed by the parent on behalf of the child.

2.5.2 Attempting interviews with the children

For the 5 to 10 proxy interview, the interviewer was instructed to continue straight into the child survey after the adult interview if possible. For the 11 to 15 interview, the interviewer was only permitted to approach the child to attempt

an interview once parent/guardian permission had been obtained. It was recommended that the 11 to 15 interview should be conducted during the same visit as the adult interview if possible, though appointments for a re-visit could be made for the 11 to 15 interview if necessary.

2.6 Interview length

The mean adult sample interview length for the questionnaire in quarters 1 to 3 was 41.4 minutes (median 38.6 minutes). In quarter 4 the interview length was cut and only key participation and demographic information was collected. For this shorter adult questionnaire, the mean length of interview was 22.6 minutes (median 21.8 minutes).

The 5 to 10 child interview mean length was 11.2 minutes (median 10.3 minutes), while the 11 to 15 child interview had a mean length of 24.5 minutes (median 22.5 minutes)[5].

2.7 Respondent incentives

Discussions were held regarding the incentives to be used, but following the success of the incentives used in Year 3, the incentives were not changed for Year 4. As with the previous year, each household received a book of stamps with the advance letter. In addition, each household that completed the interview(s) received a £5 high street voucher.

This combination of incentives was found to be the most effective incentive to decrease refusal rates and increase response rates from an experiment of various incentives in Year 2 of the survey.

No additional incentive was provided for the child surveys.

[5] All figures have been calculated after capping the lower and upper extreme values - adult Q1-Q3 interview – lower 0.2%, upper 1.1%; adult Q4 interview – lower 0.1%, upper 0.6%; 11-15 interview – lower 0.7%, upper 1.1% and 5-10 interview – lower 1.4%, upper 0.3%. Extreme lower (including negative) and upper values are likely to have arisen from interviews being split into two or more sessions, since the computation is not date-sensitive (e.g. if an interview was concluded on a subsequent day but earlier in the day, the difference between relative start and end times could be negative, or unexpectedly small).

3 Coding open ended questions

Code frames for open-ended (and 'partially open-ended[6]) questions for the survey were originally developed by BMRB and signed-off by DCMS in Year 1 of the survey, based on the verbatim answers of such questions. Code frames were reviewed quarterly in Year 4 and new codes were added if the questionnaire had changed or, occasionally, on existing questions where respondents had given answers which they had not given in the previous years of the survey.

The coding of open-ended questions was carried out using a web-based package called Ascribe by an experienced team of coders in Kantar Operations. Five per cent of open-ended answers were checked by senior coders. New coders had 100% of their work checked until the required standard was reached and thereafter their work was systematically spot-checked. On questions where the "Other" answer category exceeded 10%, answers were also reviewed.

The coding team also code socio-economic data for this survey to produce Standard Occupational Classification (SOC) and National Statistics Socio-economic Classification (NS-SEC) categorisation, from a series of standard questions which were designed for NS-SEC and SOC categorisation.

BMRB researchers briefed the coding team whenever new or revised code frames were adopted and kept in close contact with the coding team throughout fieldwork to ensure that coding was carried out at regular intervals. At least every quarter of the survey year the coding was accessed by the BMRB research team to check the quality of the coders' work in terms of what had been back-coded to each answer category, and to see what sort of answers had been left in "Other". The research team advised the coding team of any additional back-coding or any new codes which were necessary. New codes which were added were suggested by BMRB and signed-off by DCMS.

A list of all of the code frames used on open-ended and partially open-ended questions in Year 4 can be found in Appendix G.

[6] Questions with a response list but with an "Other – specify" response option, for the respondent to give an answer that the response list did not cover, are known as "partially open-ended".

4 Fieldwork outcomes

4.1 Adult sample

Table 4.1 shows the fieldwork outcomes for the adult sample issued in Year 4 of Taking Part. The final contact rate was 91.7%[7] and the final co-operation rate was 63.4%[8]. The (unadjusted) response rate was **58.1%**.

It is standard practice to assume that a proportion of the outcomes classified as 'Residential address but no contact with anyone at address' is actually deadwood. This proportion is equal to the proportion of other outcomes that is classified as deadwood.

27,186 (total number of outcomes) minus 1,639 (total residential non-contacts) = 25,547 outcomes, of which 2,517 are deadwood (9.85%). 2,517 * 9.85% = 248 assumed deadwood addresses among the residential non-contacts. This increases the total deadwood count to 2,765 (2,517 + 248) and the total non-deadwood outcomes is reduced to 24,421 (27,186- 2,765). The *adjusted* response rate = **58.7%.**

Table 4.1 Fieldwork outcomes (adult sample)

OUTCOME		**OUTCOME GROUPING**		**% of total issues**	**% of non-deadwood**
Not yet built/under construction	60	Deadwood	2,517	9.3%	-
Derelict/demolished	101				
Vacant/empty housing	1435				
Non-residential address	337				
Communal establishment	64				
Address residential & occupied but not main residence	219				
Other ineligible	90				
Inaccessible	31				
Unable to locate address	180				

[7] (Interviews + Refusals + Other unproductive)/ Total non-deadwood

[8] Interviews / (Interviews + Refusals + Other unproductive)

Residential address but no contact with anyone at address	1639	Non contact	2,057	7.6%	8.3%
Person selected but no contact with selected person	418				
No contact with parent to get parental permission	0				
Information about occupants refused	2733	Refusal	6,292	23.1%	25.5%
Office refusal	577				
Parent refused permission to interview	18				
Refusal by selected person	2441				
Proxy refusal	486				
Refusal during the interview	37				
Broken appointment	599	Other unproductive	1,977	7.3%	8.0%
Selected person ill at home during survey period	145				
Selected person away or in hospital throughout survey period	236				
Selected person physically or mentally unable	328				
Selected person has inadequate English	167				
Other unproductive	491				
Interview reported but no data received	11				
Full interview	14,343	Interview	14,343	52.8%	58.1%
TOTAL			27,186		

4.2 Child Sample

4.2.1 5-10 sample

Table 4.2 shows the fieldwork outcomes for the 5-10 child sample. The final contact rate should be **100%** as screening for the 5-10 child interview by proxy should only take place with households co-operating with the main (adult) survey. However in 4 households the screening was completed and no contact (or re-contact) was made with the parent or guardian of the selected 5-10 year old.

The final co-operation rate was **90.8%**[9] and response rate was **90.6%**. Occasionally, BMRB achieved a child interview without the adult interview (due to lost data) but, as a general formula, the *cumulative* response rate for the 5-10 survey is adult response rate * child response rate = 58.1%*90.6% = **52.6%.**

Table 4.2 Fieldwork outcomes (5-10 sample)

OUTCOME		**OUTCOME GROUPING**		**% of total issues**	**% of non-deadwood**
No child aged 5-10 in household or main interview not with parent of 5-10 year old	13,456	Deadwood	22,106	93.1%	-
Information for child screening refused	95				
Unable to complete child screening (non-response/deadwood in adult survey)	8,555				
Residential address but no contact with anyone at address (when seeking child interview)	-	Non contact	4	0.02%	0.25%
Child selected but no contact (or re-contact) with parent of child	4				

[9] (Interviews / (Interviews + Refusals + Other Unproductive)

Selection information refused	-	Refusals	95	0.4%	5.8%
Office refusal	2				
Refusal during interview	30				
Refusal by selected person	48				
Proxy refusal	15				
Broken appointment	14	Other unproductive	55	0.2%	3.4%
Selected person ill at home during survey period	3				
Selected person away or in hospital throughout survey period	4				
Selected person physically or mentally unable	-				
Selected person has inadequate English	1				
Other unproductive	26				
Interview reported but no data received	7				
Full interview	1,475	Interview	1,476	6.2%	90.6%
Partial interview	1				
TOTAL			23,736		

4.2.2 11-15 sample

Table 4.3 shows the fieldwork outcomes for the 11-15 child sample. The final contact rate was **97.6%**[10] and the final co-operation rate was **71.8%**[11]. The response rate was **70.1%**. It should be borne in mind that the request for an interview with an 11-15 year old could only be made in households co-operating with the main (adult) survey request. Occasionally, BMRB achieved a child interview without the adult interview (due to broken appointments with the adult or lost data) but, as a general formula, the *cumulative* response rate for the child survey is adult response rate * child response rate = 58.1%*70.1% = 40.7%.

Table 4.3 Fieldwork outcomes (child sample)

OUTCOME		**OUTCOME GROUPING**		**% of total issues**	**% of non-deadwood**
No child aged 11-15 in household	13,435	Deadwood	22,101	93.1%	-
Information for child screening refused	82				
Unable to complete child screening (non-response /deadwood in adult survey)	8,584				
Child selected but no contact with selected child	34	Non-contacts	40	0.2%	2.4%
No contact with parent to get parental permission	6				
Selection information refused	2	Refusal	305	1.3%	18.7%
Office refusal	2				
Refusal during interview	15				
Parent refused permission to interview	169				
Refusal by selected child	89				
Proxy refusal	28				

[10] (Interviews + Refusals + Other unproductive)/Total non-deadwood

[11] (Interviews / (Interviews + Refusals + Other Unproductive)

Broken appointment	32	Other unproductive	144	0.6%	8.8%
Selected child ill at home during survey period	3				
Selected child away or in hospital throughout survey period	19				
Selected child physically or mentally unable	10				
Selected child has inadequate English	1				
Other unproductive	62				
Interview reported but no data received	17				
Full interview	1,146	Interview	1,146	4.8%	70.1%
Partial interview	0				
TOTAL			23,736		

5 Weighting

5.1 Adult data design weights

The adult data weights were computed for the quarters Apr 08, July 08, Oct 08 and Jan 09. The first of these quarters was also the fourth quarter of Year 3.

The design weight is equal to the inverse of the individual's selection probability.

The individual's selection probability was computed as follows:

> Address selection probability * (1/ number of dwelling units at address) * (1/ number of individuals aged 16+ in selected dwelling unit).

However, the April-June 2008 quarter had two roles: 1) to provide data to support the PSA requirements covering 2005-8, and 2) to provide data to support the new PSA requirements covering 2008-11. The questionnaire program randomly allocated respondents to the 'old' or the 'new' questionnaire with a .5 probability of each. In order to maximise representativeness, this allocation was carried out separately within strata formed by some of the preceding questionnaire items.

Consequently, as far as the 'new' questionnaire is concerned, the individual's selection probability in the Q1 (Apr 08) sample was computed as follows:

Address selection probability * (1/ number of dwelling units at address) * (1/ number of individuals aged 16+ in selected dwelling unit) * (1/2).

A small number of PSUs were deleted from the Q4 (Jan 09) sample so the individual's selection probability in the Jan 09 sample was computed as follows:

Address selection probability * (1/ number of dwelling units at address) * (1/ number of individuals aged 16+ in selected dwelling unit) * (6720/7032).

5.2 Adult data non-response weights

Non-response weights were computed in two stages:

1) Area-type non-response weights;

2) Target population weights.

Area-based information was attached to each issued address in the core sample. This included a mix of Census data, Census-derived data (such as the ACORN geo-demographic classification) and administrative data (e.g. population density, deprivation indices and government boundaries).

The CHAID procedure was used to classify addresses in terms of mean response rate. Three area-based variables were used for the six month dataset covering January to June 2008. The second of these two quarters also stands as the first quarter of the new dataset. The distinguishing variables were:

- Region
- % of PSU population classified as White British according to Census
- PSU population density

This procedure produced 7 classes with mean response rates ranging from 47% to 67%. An area-type non-response weight was computed that was equal to the inverse of the weighted mean response rate for the class. The 7 classes are detailed in Table 5.1 below:

Table 5.1 Area-type non-response weights (Q1 (Apr 08 sample))

REGION	% White British	Population density (individuals per square mile)	MEAN RR	WEIGHT (1/RR)
All	>97.2%	All	66.5%	1.50
All	>93.3<97.2%	<430	66.0%	1.52
All	>93.3<97.2%	>430<5972	63.3%	1.58
All	>93.3<97.2%	>5972	60.4%	1.65
NW, Yorkshire & Humber, West Midlands, SE, SW	<85.9%	All	57.4%	1.74
All	>85.9<93.3	All	56.7%	1.76
NE, East Midlands, East of England, London	<85.9%	All	46.8%	2.14

A slightly different set of input variables was used for Q2-4 (July 2008 to March 2009) and the CHAID procedure was run separately for Q2 and for Q3-4 combined as these datasets were required at different times. Nine classes were produced for Q2 and five for Q3-4[12].

The CHAID procedure found only two variables were significant: region and the proportion aged 16-35.

[12] The procedure was altered slightly for Q3-4 to produce slightly fewer classes.

Table 5.2 Area-type non-response weights (Q2 (July 08 sample))

REGION	% aged 16-35	MEAN RR	WEIGHT (1/RR)
NE, SW	All	68.0%	1.47
Yorkshire & Humber, East of England, SE	<26.6%<30.1%	65.4%	1.53
Yorkshire & Humber, East of England, SE	<21.9%	65.4%	1.53
NW, East Midlands	All	63.7%	1.57
Yorkshire & Humber, East of England, SE	<21.9%<26.6%	61.0%	1.64
West Midlands	All	55.6%	1.80
Yorkshire & Humber, East of England, SE	>34.2%	53.2%	1.88
Yorkshire & Humber, East of England, SE	<30.1%<34.2%	52.4%	1.91
London	All	45.5%	2.20

Table 5.3 Area-type non-response weights (Q3 (Oct 08) and Q4 (Jan 09) samples)

REGION	% aged 16-35	MEAN RR	WEIGHT (1/RR)
NE, NW, SW	All	63.7%	1.57
Yorkshire & Humber, East Midlands, West Midlands, East of England, SE	>25.0%<29.0%	63.3%	1.58
Yorkshire & Humber, East Midlands, West Midlands, East of England, SE	<25.0%	58.5%	1.71
Yorkshire & Humber, East Midlands, West Midlands, East of England, SE	>29.0%	54.9%	1.82
London	All	46.1%	2.17

In the second and final non-response weighting stage, BMRB applied rim weights to match targets provided by the 2007 ONS mid-year population estimates. The targets were based on sex, age, region, and (LFS estimate) ethnic group.

One other target was used for the Oct 08/Jan 09 data: interview date. This was designed to counteract the slightly uneven 'flow' of interviews over the course of the fieldwork period and address any bias in estimates due to seasonality. The targets were calculated by taking the distribution of interviews for each issue month (e.g. *x*% of interviews completed in the issue month, *y*% completed in the

following month, *z*% completed in the third month etc.) and calculating the total distribution across all issue months. This average was then applied to the whole fieldwork period to get 'expected' numbers of interviews for each month of the fieldwork period. Note that this stage was not included when the Q1 (Apr 08) and Q2 (July 08) datasets were produced because of the relatively small number of cases in each.

This second stage of non-response weighting was quite minor as the first stage brought most of the key distributions close to these targets.

Table 5.4 Targets used for second stage non-response weighting

Age	**Male**	**Female**
16 - 19	3.36%	3.15%
20 - 24	4.29%	4.06%
25 - 29	4.07%	4.04%
30 - 34	4.00%	3.99%
35 - 39	4.60%	4.64%
40 - 44	4.75%	4.81%
45 - 49	4.23%	4.31%
50 - 54	3.69%	3.78%
55 - 59	3.71%	3.82%
60 - 64	3.42%	3.58%
65 - 69	2.60%	2.79%
70 - 74	2.22%	2.51%
75+	3.74%	5.84%
TOTAL	48.68%	51.32%

Region	
North East	5.07%
North West	13.41%
Yorkshire & Humberside	10.14%
East Midlands	8.65%
West Midlands	10.45%
East of England	11.06%
London	14.72%
South East	16.23%
South West	10.27%
TOTAL	100.00%

Ethnic group	
White	90.14%
Indian	2.23%
Pakistani / Bangladeshi / Other Asian	2.53%
Black	2.43%
Mixed/Chinese/Other ethnic group	2.67%
TOTAL	100.00%

Interview date	
Issued Q1-2	50.0%
Issued Q3-4 / Interview in October 2008	5.37%
Issued Q3-4 / Interview in November 2008	10.67%
Issued Q3-4 / Interview in December 2008	5.76%
Issued Q3-4 / Interview in January 2009	8.16%
Issued Q3-4 / Interview in February 2009	11.08%
Issued Q3-4 / Interview in March 2009	5.76%
Issued Q3-4 / Interview in April/May 2009	3.21%
TOTAL	100.00%

5.3 Final adult weights

Weights were 'capped' to avoid inflating the variance of the survey estimates. Although such caps may introduce extra bias to the estimates, this is likely to be minor when only the top 1-2% of weights are capped (as here)[13].

The three datasets (Q1, Q2 and Q3-4) were then combined, with the sum of weights scaled respectively to 25%, 25% and 50% of the total.

[13] Most of the largest weights are due to unexpectedly large numbers of dwelling units at a single address. These are likely either to be interviewer errors or to be very unusual cases that happen to fall into this particular sample. The addresses used in a survey represent a random sample of the PAF which means that the proportion that turn out to contain multiple dwelling units is only an *estimate* of the proportion in the full PAF. This estimate is subject to natural sampling error so should not be taken as absolute.

5.4 Child data design weights

The first thing to note is that the child dataset covers a different but overlapping period from the adult dataset. It covers samples issued in July 08, Oct 08, Jan 09 and Apr 09. Consequently, Q1 for the child dataset is Q2 for the adult dataset.

Although interviews with 11-15 year olds had been a feature of the survey design since Jan 06, these were suspended for Y4 until the last month of Q1 (Sept 09) to allow for questionnaire design changes to be finalised. At the same time, any adult respondent who was identified as having a parental/caregiver relationship with at least one resident child aged 5-10 was asked to provide information by proxy about one (randomly selected) child.

As before, the design weight is equal to the inverse of the sampling probability.

11-15s

The basic sampling probability for 11-15s was computed as follows:

Address selection probability * (1/ number of dwelling units at address) * (1/ number of individuals aged 11-15 in selected dwelling unit).

This basic value was adjusted in various quarters:

Q1 (to reflect late start of child survey; only 3511 of the sampled 6962 addresses were allocated to the child survey):

Address selection probability * (1/ number of dwelling units at address) * (1/ number of individuals aged 11-15 in selected dwelling unit)*(3511/6962).

Q3 (to reflect deletion of a small number of PSUs):

Address selection probability * (1/ number of dwelling units at address) * (1/ number of individuals aged 11-15 in selected dwelling unit)*(6720/7032).

Q4 (to reflect deletion of a small number of addresses in sampled PSUs and at a proportionately higher rate in London):

London:

Address selection probability * (1/ number of dwelling units at address) * (1/ number of individuals aged 11-15 in selected dwelling unit)*(1140/1260).

Outside of London:

Address selection probability * (1/ number of dwelling units at address) * (1/ number of individuals aged 11-15 in selected dwelling unit)*(5376/5760).

5-10s

For the 5-10s, the sampling probability was dependent upon the adult respondent sampling probability. However, it needs to take into account the fact that in two parent households, the child may have been sampled via either of the parents/caregivers.

Unfortunately, the questionnaire did not collect information about the child's relationship with every adult in the household, only the adult respondent[14]. Consequently, the number of resident parent/caregivers of the sampled child has had to be imputed. If the adult respondent reported a live-in partner (whether married or not) that live-in partner was considered to have a parental/caregiver relationship with the child. Consequently, the basic sampling probability is computed as:

Address selection probability * (1/ number of dwelling units at address) * (1/ number of adults aged 16+) * (1/ number of individuals aged 5-10 with a dependent relationship with adult respondent) * 2.

Where there was no live-in partner, the final term (*2) is dropped from the equation.

All of the additional adjustments due to the late survey start (Sept 08 instead of July 08) and address deletions that are noted in the section on 11-15s are also relevant here.

5.5 Child data non-response weights

The PSU-level non-response weight computed for the adult dataset was also used for the child dataset since both the 11-15 interview and the 5-10 proxy interview were dependent upon the initial co-operation of the sampled adult.

However, nothing had been computed for the child dataset Q4 since this goes beyond the end of the adult dataset. A new CHAID analysis was carried out based on the *whole* Jun 08 – Oct 08 – Jan 09 - Apr 09 adult dataset to fit with some other revisions that were being carried out at the time. Consequently, the area non-response weights that might be applied to adult dataset Q2-4 are slightly different.

Nevertheless, this revised procedure generated the following five response classes.

[14] This questionnaire error has since been corrected.

Table 5.5 Area non-response weights (child data)

REGION	% aged 65+	MEAN RR	WEIGHT (1/RR)
SW	All	66.3%	1.51
NE, NW, East Midlands	All	62.7%	1.59
Yorkshire & Humber, West Midlands, East of England, SE	>16.0%	60.5%	1.65
Yorkshire & Humber, West Midlands, East of England, SE	<16.0%	56.9%	1.71
London	All	48.2%	2.07

In the second and final non-response weighting stage, BMRB applied rim weights to match targets provided by the 2008 ONS mid-year population estimates. The targets were based on sex, age, region, and ethnic group[15]. At the same time, each quarter was given an equal weight.

Table 5.6 Targets used for second stage non-response weighting

Age	Male	Female
5	4.52%	4.28%
6	4.37%	4.17%
7	4.34%	4.18%
8	4.47%	4.26%
9	4.58%	4.38%
10	4.66%	4.47%
11	4.79%	4.57%
12	4.76%	4.55%
13	4.80%	4.56%
14	4.91%	4.65%
15	5.00%	4.73%

Region	5-10	11-15
North East	2.51%	2.35%
North West	7.06%	6.50%
Yorkshire & Humberside	5.26%	4.87%
East Midlands	4.47%	4.12%
West Midlands	5.74%	5.18%
East of England	6.00%	5.35%

[15] Based on a projection from the 2007 estimates. the 2008 ethnic estimates were not yet available.

London	7.79%	6.31%
South East	8.77%	7.89%
South West	5.09%	4.74%

Ethnic group	5-10	11-15
White	44.14%	39.63%
Not white	8.55%	7.68%

Sample quarter	
Issued July 08	25.0%
Issued Oct 08	25.0%
Issued Jan 09	25.0%
Issued Apr 09	25.0%

6 Final design effects for key variables

6.1 Adult data

Significance tests assume that the achieved sample is a simple random sample from the survey population. The design effect takes into account the actual complexity of the sample design, reflecting the compromises necessary for real world survey practice.

The actual sample size divided by the design effect equals the *effective* sample size. The effective sample size - rather than the actual sample size - is used for tests of significance.

Table 6.1 below shows a selection of key (weighted) Y4 results, the attendant design effects and the 95% confidence intervals for each result[16]. The design effects range from 1.6 to 2.5 but all of the estimates are accurate to +/-1.3 percentage points or less.

Table 6.1 Design effects for key PSA variables (Year 4)

RESULT	Weighted result	Design effect	95% confidence intervals [range]
% Using a library service at least once in the last 12 months	39.63%	1.56	38.6% - 40.6% [2.0pp]
% Visiting a museum/gallery/archive at least once in last 12 months	44.51%	1.85	43.4% - 45.6% [2.2pp]
% Visiting 2+ historic environment sites in last 12 months	56.95%	2.49	55.7% - 58.2% [2.5pp]
% Engaging in at least three arts activities in the last 12 months	59.83%	1.99	58.7% - 60.9% [2.2pp]
% Doing at least 12 'thirty minute plus' sessions of moderate intensity sports / recreational physical activity in last 4 weeks	22.67%	1.70	21.8% - 23.6% [1.8pp]
% meeting DCMS target (2/5)	65.86%	2.12	64.7% - 67.0% [2.3pp]

[16] Computed using the SPSS Complex Samples Module.

Table 6.2 details the design effects for a number of key sub-groups. The design effects tend to be lower, reflecting the fact that these sub-groups will be more thinly distributed between PSUs leading to a smaller cluster effect[17].

Table 6.2 Key sub-group design effects (adult data Y4)

	PSA variable					
	Library use	Museum / gallery / archive visits	Historic site visits	Arts activity	Sport activity	2+/5
All	1.56	1.85	2.49	1.99	1.70	2.12
SEX						
• males	1.39	1.58	1.84	1.57	1.57	1.69
• females	1.48	1.58	2.05	1.59	1.47	1.69
DISABILITY STATUS						
• longstanding illness / disability / infirmity	1.53	1.80	2.33	1.83	1.61	2.01
• no longstanding illness / disability / infirmity	1.28	1.49	1.56	1.43	1.43	1.48
ETHNIC GROUP						
• BME	1.49	1.74	2.32	1.78	1.63	1.87
• White	1.61	2.04	1.88	2.30	1.90	2.38
NS-SEC						
• NS-SEC 1-4	1.42	1.49	1.79	1.56	1.49	1.61
• NS-SEC 5-8	1.39	1.46	1.87	1.58	1.54	1.59
AGE GROUP						
• 16-24	1.30	1.38	1.52	1.30	1.23	1.31
• 25-44	1.31	1.52	1.90	1.56	1.38	1.71
• 45-64	1.40	1.43	1.48	1.40	1.42	1.35
• 65-74	1.35	1.19	1.48	1.48	1.27	1.41
• 75+	1.27	1.41	1.34	1.39	1.25	1.38

[17] There are two versions of these design effects that can be calculated. One uses a notional simple random sample of the full population as the benchmark and one uses a notional random sample of the sub-group population as the benchmark. The former is a more realistic assessment of the impact of complex sample design but the latter makes calculation of standard errors simpler as these are derived simply by multiplying the standard error of the simple random sample by the square root of the design effect (also known as the 'design factor').

Average (nationally distributed groups)	1.38	1.48	1.74	1.52	1.42	1.57

For other measures, an average overall design effect of 1.52 may be used for calculating the effective sample size

6.2 Child data

Table 6.3 below shows a selection of key (weighted) Y4 results, the attendant design effects and the 95% confidence intervals for each result[18]. The design effects range from 0.8 to 1.6. They are lower than the equivalent adult design effects because the child cluster sizes are smaller.

Table 6.3 Design effects for key child dataset variables (Year 4)

RESULT	Weighted result	Design effect	95% confidence intervals [range]
Hours of in school cultural participation (11-15s only)	4.8	1.36	4.5 – 5.2 [0.7]
Hours of in or out of school cultural participation (11-15s only)	9.8	1.35	9.2 – 10.5 [1.3]
Hours of out of school cultural participation (5-15s only)	4.2	1.46	4.0 – 4.5 [0.5]
Hours of in school sport participation (11-15s only)	1.5	0.84	1.3 – 1.7 [0.4]
Hours of out of school sports participation	2.7	1.67	2.5 – 2.9 [0.4]
Hours of in or out of school sports participation (11-15s only)	4.6	1.15	4.2 – 5.0 [0.8]
Meet 5-hour cultural target for in or out of school participation (11-15s only)	66.3%	1.55	62.8% - 69.8% [6.9%]
Meet 5-hour cultural target for out of school participation (5-10s only)	27.1%	1.38	24.4% - 29.8% [5.4%]
Whether meet 5-hour sport target for in or out of school participation (11-15s only)	32.2%	1.47	28.9% - 35.4% [6.6%]
Whether meet 3-hour sport target for out of school participation - (5-10s only)	28.5%	1.39	25.8% - 31.3% [5.5%]

[18] Computed using both STATA and the SPSS Complex Samples Module (to check consistency). The design effects reflect only the highest level of stratification (region) due to a significant number of PSUs with one or fewer completed child interviews.

Table 6.4 details the design effects for a number of key sub-groups. The design effects tend to be very slightly lower than for the full sample. They are *very* low for disabled children because there is no cluster effect to consider.

Table 6.4 Key sub-group design effects (child data Y4)

	ALL	DISABLED	BME	WHITE	MALES	FEMALES	5 TO 10	11 TO 15
Hours of in school cultural participation (11-15s only)	1.36	1.03	1.43	1.31	1.36	1.27		1.36
Hours of in or out of school cultural participation (11-15s only)	1.35	0.95	1.30	1.30	1.23	1.26		1.35
Hours of out of school cultural participation (5-15s only)	1.46	0.95	1.80	1.35	1.45	1.28	1.47	1.23
Hours of in school sport participation (11-15s only)	0.84	0.98	1.55	0.83	0.77	1.66		0.84
Hours of out of school sports participation	1.67	1.23	1.23	1.66	1.54	1.43	1.41	1.43
Hours of in or out of school sports participation (11-15s only)	1.15	1.09	1.34	1.12	1.06	1.35		1.15
Meet 5-hour cultural target for in or out of school participation (11-15s only)	1.55	1.32	1.48	1.51	1.39	1.56		1.55
Meet 5-hour cultural target for out of school participation (5-10s only)	1.38	0.97	1.19	1.40	1.21	1.46	1.38	
Whether meet 5-hour sport target for in or out of school participation (11-15s only)	1.47	1.15	1.51	1.42	1.40	1.43		1.47
Whether meet 3-hour sport target for out of school participation - (5-10s only)	1.39	1.39	1.55	1.42	1.37	1.64	1.39	
AVERAGE	1.36	1.10	1.44	1.33	1.28	1.43	1.41	1.30