An analysis of mode effects using data from the Health Survey for England 2006 and the Boost Survey for London

Sarah Tipping, Steven Hope, Kevin Pickering, Jennifer Mindell and Bob Erens

An analysis of mode effects using data from the Health Survey for England 2006 and the Boost Survey for London

Sarah Tipping, Steven Hope, Kevin Pickering, Jennifer Mindell and Bob Erens

Prepared for The NHS Information Centre for health and social care

July 2008

P2795

Contents

List of tables

INTRODUCTION

The NHS Information Centre for health and social care (IC) is currently reviewing all the health-related surveys it commissions in order to ensure they are fit for purpose, cost-effective, and provide value for money.

As part of its review, the IC commissioned NatCen to examine the extent to which results from the Health Survey for England (HSE) 2006, which is carried out using face-to-face interviews, and the Boost Survey for London, carried out largely using a self-completion questionnaire, are comparable, and to report on the magnitude and direction of any differences. The HSE is an annual, nationally-representative sample of the English population living in private households. Selected households are surveyed using computer-assisted personal interviews (CAPI) to gain household- and individual-level data. Respondents are then visited by a nurse for further measurements and the collection of a small blood sample. Some information is collected using self-completion questionnaires, but this is a very small component of the HSE. For the London Boost, an interviewer collected brief household-level information and then left a longer self-completion questionnaire for each eligible respondent; these were then collected later by the interviewer or posted back by the respondent directly to NatCen's office. There was no nurse visit on the London Boost. The purpose of the Boost was to increase the sample size to enable analyses at Primary Care Trust (PCT) level for all of London's 31 PCTs.

The samples for both HSE and the London Boost were selected using a two-stage, stratified sampling procedure, with addresses selected from the small-user Postcode Address File (PAF). All adults (aged 16 and over) and two children (aged 0 to 15 years) in each selected address were eligible to participate.

All London Boost Survey respondents and all HSE respondents who were part of the core sample and resident in one of London's PCTs were included in the analysis carried out for this report. Throughout this report we will refer to HSE Core sample members as the 'Core' and respondents from the Boost survey of London as the 'Boost'. The analysis only covers adults aged 16 or over; children have not been included.

There were two main strands to the comparisons carried out in this report:

1. an analysis of the effects of differential response rates, and
2. an analysis of the effects of differential measurement error.

Response differences between the two surveys were examined in terms of the overall response levels of households and individuals, and the amount of missing data items: where respondents had refused or skipped individual questions. The effects of differential response rates on sample composition were also examined by comparing the socio-demographic characteristics of the two samples. The effects of differential response on the weighted estimates were also compared in order to look for any residual bias once the non-response weights had been applied. (The two samples were weighted separately using the same weighting scheme, as described in detail in APPENDIX B.)

The second strand of the analysis was to investigate the effects of measurement error attributable to the different modes of questioning. To do this the socio-demographic profile of the Core sample was adjusted to make it match that of the Boost sample. The comparison was then carried out on the matched sample, so that any remaining differences would be attributable to measurement error rather than sample composition.

This was done by adjusting the sample composition of the Core sample to match that of the *unweighted* Boost sample, so the samples would be comparable to one another; it should be noted, however, that these matched samples are no longer representative of a wider population, and the results must not be used outside the context of this report in order to generalise to the London population.

SUMMARY AND RECOMMENDATIONS

The aim of this report was to look for differences in key measures of health and health-related behaviours by mode of data collection (i.e. a face-to-face personal interview versus a self-completion questionnaire). Comparisons were made between key health estimates from London participants in the 'Core' Health Survey for England (HSE) 2006 and the 'Boost' Survey for London. The 'Core' HSE involved a face-to-face computer assisted interview (CAPI), while the London 'Boost' obtained most of its data using a self-completion questionnaire. Despite the differences in mode, the Boost survey was designed to be as comparable with the Core survey as possible, as it was intended from the outset that the two achieved samples should be combined during analyses to maximise the sample size within each Primary Care Trust (PCT) in London.

Before the analyses of health measures were carried out, the socio-demographic profiles of the two achieved samples were compared. For the majority of characteristics, there were no differences between the Core and Boost samples. The exceptions were ethnicity and educational qualifications. The difference in the ethnic profile of the two samples, however, was reduced after the non-response weights were applied. The difference in educational qualifications persisted after weighting, and seemed to be attributable to differences in question format between the two surveys.

For many of the key health measures, there were no significant differences between the two surveys. These included estimates for long-term illness, limiting long-term illness, rates of current smoking, whether respondents drank alcohol, and how often they usually drank. Estimates produced from combined Boost and Core data will not be biased for these variables.

However, there were a number of statistically significant differences between the estimates for some key measures including: general health, GHQ12 score, number of cigarettes smoked, number of alcoholic units consumed on the heaviest drinking day, portions of fruit and vegetables consumed, levels of moderate physical activity, and (for women) perceived social support. For these measures, it is difficult to determine the specific causes of the differences as they are likely to be due to a combination of mode and other effects. Large differences between the two variables imply some degree of bias in one or both of the estimates. For some of these variables, we are on fairly safe ground in making assumptions about which estimate is likely to be the 'better' one, but this is not the case for all the variables.

There is evidence that self-completion data collection methods are better for recording sensitive behaviours, such as levels of smoking and drinking. This would suggest the Boost sample estimates may be more accurate for these behaviours.

For fruit and vegetable consumption and the physical activity questions, the Core data are likely to be the more accurate, because in the self-completion format the lack of guidance for respondents completing these cognitively demanding and complex modules means there is a high level of missing data for these questions.

The true estimates for self-assessed general health, GHQ12 score and levels of perceived social support are unclear.

Height, weight and BMI are directly comparable when the interviewer measurements alone are used. When self-reported height and weight is included in the Boost sample, then bias is introduced.

The decision to combine data depends not only on the variables to be combined, but also on the purpose of the particular analysis that is being carried out. If the aim is to make comparisons between PCTs in London, then the combined data are useful: i.e. since the interest is in the relative differences between London PCTs, even though the estimates may be biased, this bias will be found for all PCTs and results will still be comparable. However, caution must still be exercised even for comparing PCTs, since the proportion of the Boost sample to the Core sample varies per PCT - within each PCT the proportion of combined sample that is from the Core ranges between 16-35%, with only 6% and 8% in Camden and Islington, respectively. This means the bias to the estimates caused by combining the Core and Boost samples will also vary slightly by PCT.

To allow for comparisons to be made between PCTs an additional adjustment has been made to the combined weights for the survey analysis. This adjustment makes the ratio of Boost sample to Core sample within each PCT the same as the overall ratio of Boost sample to Core sample across London. Overall, 77% of the combined adult sample is from the London Boost. The proportion of Boost sample within each PCT is adjusted to match this. This enables direct comparisons to be made between health estimates in different PCTs. There would still be unknown bias caused by using two modes but this would be the same within each PCT. If there is doubt concerning the results of any particular comparison between a pair of PCTs the results could be checked again using Boost data only.

Combining the Core and Boost samples is more problematic if comparisons are being made between PCTs and national estimates, since it would be impossible to tell whether any differences found were true or caused by bias. Where the sample size permits, the preference is to compare the PCTs with national HSE estimates using Core data only. Where this is impossible because of small sample sizes, the estimates from the combined sample may be used but must be treated with caution.

1 BACKGROUND TO THE PROJECT

1.1 Introduction

There is an increasing need for robust, valid data on health and health-related behaviours at Primary Care Trust (PCT) level, which has lead to a growing demand for PCT-level health surveys. Alternatives, such as synthetic estimates, are not always appropriate as levels of precision are often low and they cannot be used to monitor changes in response to local interventions.

The Health Survey for England (HSE) is an annual survey, commissioned by the IC, that provides regular and reliable survey data on a broad range of health topics, lifestyle and health determinants for residents in England. The HSE was not designed to provide data for local areas, such as PCTs, as the sample sizes are too small to provide reliable estimates at this level. The HSE sample in a PCT would need to be supplemented before PCT-level analyses could be carried out. The optimum design for a boost sample would be one which exactly matched the main HSE design; however, this is not always possible for cost reasons.

The London Health Observatory commissioned a boost to the HSE in the London area in order to obtain survey results for London as a whole as well as for each of the 31 PCTs within the city. Given the size of the sample required, the London Boost used a less costly methodology of self-completion questionnaires rather than a personal interview as used on HSE. The sample is described in more detail in Section 1.3.

1.2 Overview of HSE 2006 methodology

The core HSE 2006 sample was designed to be representative of the population living in private households in England. People living in institutions were not included. The addresses were selected from the small user Postcode Address File (PAF). This is a list compiled by the Royal Mail of all the addresses in England to which mail is delivered. The PAF has very good coverage of private addresses, and only a very small proportion of households (less than 1%) are excluded. In order to increase the precision of the sample, prior to selection the PAF was sorted by local authority (PCT within London) and the percentage of households with a head of household in a non-manual occupation (Socio-Economic Groups 1-6, 13), taken from the 2001 Census.

The sample was drawn in two stages; at the first stage a sample of 720 Primary Sampling Units (PSUs)[1] was drawn with probability proportional to the total number of addresses within them. At the second stage, a fixed number of addresses was selected per PSU. This design gives each address an equal chance of being selected, making the sample of addresses representative of all addresses in England. 720 PSUs were drawn in total, 102 of which fell within London. Once selected, the PSUs were randomly allocated to the 12 months of the year (60 per month) so that each quarter provided a nationally representative sample.

[1] PSUs were postcode sectors or groups of postcode sectors. Postcode sectors with fewer than 500 addresses were combined with neighbouring sectors before selection to avoid undue clustering of sampled addresses.

Most addresses selected from the PAF contain a single household. However, a small proportion of addresses (about 1%) contain more than one household. If an address contained more than one household, interviewers randomly selected a maximum of three.

The HSE is a face-to-face survey; interviewers conduct a short household questionnaire followed by individual interviews with all adults (household members aged 16 or over) and up to two children (aged 0-15). Individual interviews can be run concurrently using CAPI (Computer Assisted Personal Interviewing); up to four household members can interviewed together at one time. The household and individual interviews take about an hour on average. In addition there are self-completion questionnaires for all adults, which are filled in whilst the interviewer is still present. The interviewer then takes the completed self-completion questionnaires away with them. The content of the self-completion questionnaire varies according to the age of the sample member; for example, the smoking and drinking section for children, 16-17 year olds and some 18-24 year olds[2] is in the self-completion module, while for other adults it is in the main CAPI questionnaire. The interviewer also takes height and weight measurements from sample members. The HSE interview also incorporates a nurse visit where additional physical measurements (e.g. blood pressure, waist and hip circumferences) and a small blood sample are taken.

1.3 Overview of the London Boost Survey methodology

The sample for the London Boost was designed to obtain an achieved sample of about 200 adult respondents in each of the 31 London PCTs, with additional boosts in both Camden and Islington PCTs.

As with the Core sample, the addresses were selected from the small user PAF. Again the PAF was sorted by PCT and within each PCT, by the percentage of households with a head of household in a non-manual occupation. The sample was then drawn in two stages; at the first stage 202 PSUs were selected with unequal probability. This was to ensure 6 PSUs were selected per London PCT, except in both Camden and Islington, where 14 PSUs were selected in each. At the second stage a fixed number of addresses were selected per PSU: 40 addresses were selected within each inner London PSU, where response rates were expected to be lower, and 34 addresses were selected per outer London PSU. The unequal selection probabilities mean address selection weights are required at the analysis stage to make the sample of selected addresses representative of London. As with the Core HSE sample, interviewers randomly selected a maximum of three households at each address.

The methodology chosen for the London Boost was adapted from that used by the Welsh Health Survey, and involved administering self-completion questionnaires rather than face-to-face CAPI. Interviewers visited each selected household to carry out a short household interview (using a paper questionnaire), and to recruit household members. All adults (aged 16 or over) and up to two children (aged 0-15) were eligible for the survey and were given a self-completion questionnaire. Interviewers were instructed to return to the household to collect completed individual questionnaires, rather than leave respondents to mail them back. On their return visit, the interviewer would attempt to encourage non-responding individuals to complete their questionnaires. In addition, the interviewer measured the height and weight of all consenting respondents who were present at any of the visits. Unlike the HSE, there was no second stage nurse visit.

[2] Respondents aged between 18 and 24 have the option of CAPI or self-completion.

The self-completion questionnaire was kept to a reasonable length to encourage participation; hence only a sub-set of topics and questions from the HSE could be included. Questions were taken from the HSE self-completion questionnaires where these existed (e.g. smoking and alcohol questions, social support). Other HSE questions were converted to self-completion format with the aim of keeping the question wording as similar as possible to the HSE interview.

1.4 Combining data from mixed modes

One consequence of opting for the methodology used on the London Boost survey is that different modes of data collection have been used within each PCT for the two parts of the sample: face-to-face interviews for Core HSE respondents, and self-completion questionnaires for Boost respondents. Since using different modes may affect the comparability of the two samples, it is important to be aware of any potential problems that could arise when combining the Core and Boost samples in order to provide estimates both at the London and PCT levels.

Any such differences, or 'mode effects', may be expressed as differences in coverage error, non-response error or measurement error.

Coverage error is caused when the mode of data collection has an impact on the sampling frame used to select sample members. Certain groups of individuals may be excluded from one sampling frame and not the other, meaning they would be covered by one mode but not another. Coverage error is not applicable to this study, as the two samples were selected from the same sample frame (PAF) and were recruited in the same way.

Non-response error is caused by differential non-response bias, whereby different sample members are more likely to respond to one mode than another. As a result, the survey estimates could vary by mode[3]. Non-response error is examined in Section 3.

Measurement error occurs when the answer provided by a respondent is affected by the mode in which the question is asked. Again, this will have an impact on the survey estimates provided by each mode. Measurement error is examined in Sections 4 and 5.

[3] Such differences will remain if the two samples have different weighting schemes or if there is unknown bias that is not addressed by the weighting scheme.

2 METHODS USED IN THE ANALYSIS

There were two strands to the analysis, looking firstly at the effect of non-response and secondly at the effect of differential measurement error. These are outlined in more detail below.

The effects of non-response error were investigated by looking at differences in response rates between the two surveys. A comparison was made between the overall response levels of households and individuals to each survey to assess whether differences in mode had had an effect on response. The amount of missing data items by mode was also examined by comparing the number of respondents who had not answered (i.e. refused or skipped) individual questions.

The effects of differential response on sample composition were investigated further by comparing the socio-demographic characteristics of the two samples. Non-response weights were not applied, as the aim was to identify differences in the achieved samples. However, selection weights were required to make the two samples comparable, as the sample design of the London Boost meant it was not representative of the London population.

The effects of differential response on a number of key survey estimates were then examined by comparing weighted survey estimates from the two samples. This was to look for any evidence of residual bias after non-response weights has been applied. This residual bias could be due to differential response as well as to the use of different modes. For this analysis, the two samples were weighted separately using the same weighting procedures. The non-response weights were generated to correct for the effects of differential non-response, using the weighting strategy that was originally developed for the HSE 2003 general population sample and which has been used on HSE every year since. For each sample, two sets of non-response weights were generated: a set of household weights to adjust for non-contact and refusal of households, and a set of individual weights to adjust for any additional non-response among individuals in participating households. Further details of the weighting schemes can be found in APPENDIX B.

The final analysis was to investigate the effects of differential measurement error. To do this, the socio-demographic profile of the achieved Core sample was adjusted to match the achieved Boost sample using propensity score matching (PSM).[4] PSM is a method of matching two samples. At the first stage, the differences in socio-demographic profile of the two samples were modelled using a logistic regression model. The models showed which characteristics varied the most between the two samples. A range of socio-demographic variables was tested and those variables that were significantly different were included in the final model, which was used to generate the propensity score.

The propensity score is the probability, or propensity, of an individual to be in a specific sample. This score is used to match the two samples, since respondents with similar scores should be similar in terms of the characteristics in the model. The output of the matching process is a weight which, when applied to the Core data, makes the Core sample match the Boost sample in terms of the socio-demographic variables in the model. Boost respondents were given a weight of 1, while weights for Core respondents

[4] A list of the characteristics used for this matching is given in Table C1 in Appendix C

vary and ensure that the profiles of the two samples match. The matching was carried out separately for men and women.

The comparison of key estimates was then carried out on the matched samples, so that any differences would be attributable to measurement error rather than sample composition. The analysis uses the same methods as for looking at the effect of differential response, but the matched weight is applied rather than the non-response weight. The Core sample was adjusted to match the *unweighted* Boost sample, so the samples would be comparable with one another; it should be noted, however, that these matched samples are no longer representative of a wider population, and the results must not be used outside the context of this report in order to generalise to the London population.

Although non-response weighting adjusts the samples to the same population and makes them similar in terms of age and sex profiles, PSM results in two samples that are a closer match because survey variables are included in the modelling. These provide a much wider pool of characteristics that can be used to match the two samples. More details on the PSM model can be found in APPENDIX C.

The analyses were carried out using cross-tabulations, with differences tested using chi-square tests and two-tailed t-tests. It is possible to use simple statistical tests rather than more complex analyses, because the matching controls for differences in the same way that a regression model would. For each variable, the p-value is reported, i.e. the probability that the difference found would occur by chance had the two samples come from the same population. If the p-value is small then it is very unlikely that the differences between the two samples has occurred by chance. Variables were said to be significantly different if the p-value of the test was lower than 0.05.

Since both samples were clustered, stratified and weighted, the analysis was run in Stata version 10 using the ‘svy’ commands to ensure the complex sample design was taken into account when running significance tests. The analyses were carried out on all adults aged 16 and over (children were excluded). Men and women were analysed separately.

3 COMPARISON OF RESPONSE RATES

3.1 Household response

Response rates of the Core and Boost samples were compared, both at the level of selected households, and at the level of individuals within responding households. The comparison showed only small differences in the response rates for households, but larger differences in response at the individual level.

There were 2040 addresses issued in London PCTs for the Core sample, giving 1672 eligible households. Eligible addresses were those containing occupied, private households; ineligible addresses were those that the interviewer found to be vacant, yet to be built, demolished, commercial sites or institutions. Household interviews were carried out in 963 of the eligible households, giving a household response rate of 58%.

The figures for the Boost have been weighted by the address selection weights[5]. This removes any bias caused by selecting more addresses in certain PCTs and makes the sample of Boost addresses representative of London. 7432 addresses were issued for the Boost sample. Interviews were attempted at every eligible address (6321). Household interviews were carried out at 3882 of the 6321 eligible households identified in the Boost sample, giving a household response rate of 61%. The response rates for both samples are given in Table 1 below. The Core sample is unweighted as the sample of addresses was drawn with equal probability and therefore already representative of London.

Table 1 Household response rates by sample type

	HSE Core (unweighted)			London Boost (weighted)[a]		
	N	% all	% eligible	N	% all	% eligible
Selected addresses	2040	100		7432	100	
Ineligible addresses	318	16		704	9	
Unknown eligibility	65	3		424	6	
Households at which interview sought	1657	81		6305	85	
Extra households sampled at multi-household addresses	15			16		
Total eligible households	**1672**		**100**	**6321**		**100**
Productive	**963**		**58**	**3882**		**61**
Full	765		46	3765		60
Partial	198		12	117		2
Unproductive	**709**		**42**	**2439**		**39**
No contact	40		2	323		5
Refusal	544		33	1818		29
Other unproductive	125		7	298		5

[a] The Boost addresses have been weighted by the address selection weight.

[5] The unweighted household response for the Boost sample was 62% (3873 responding households from 6234 eligible households).

The household response rate was higher for the Boost sample than the Core sample ($p<0.05$). Likely reasons for this difference include the length of the household interview; the questionnaire was shorter for the Boost sample than for the Core (about 10 minutes, whereas the Core household CAPI interview leads on directly to the individual CAPI interview, which could last up to an hour), and this may have encouraged household response.

A comparison of household response rates was made between a number of area-level variables including: PCT spearhead status[6], whether PCT was in inner or outer London, index of multiple deprivation (IMD) quintiles[7], the proportion of persons in the PSU belonging to a non-white minority ethnic group and the proportion of household heads in the PSU from non-manual occupations (both based on data from the 2001 Census). The response rates by category are shown in Table 2, as before the Boost sample has been weighted by address selection weights.

Table 2 Household response rates by sample type and area characteristics

	Household response rates		
	Core (unweighted)	Boost (weighted)	P-value
	%	%	
PCT spearhead status			
Non-spearhead	55	62	***
Spearhead	63	61	
Inner London PCTs	60	60	
Outer London PCTs	57	62	**
Index of Multiple Deprivation (IMD) (quintiles)			
Least deprived (<11.9)	52	62	***
2nd least deprived - 11.9 to 19.2	58	62	
Middle quintile - 19.3 to 27.7	54	62	**
2nd most deprived - 27.7 to 37.9	65	59	*
Most deprived (>37.9)	61	62	
Proportion of minority ethnic population (quintiles)			
Least dense (<12.1%)	56	59	
2nd lowest density - 12.1 to 19.0	56	62	
Middle quintile - 19.1 to 27.2	58	63	
2nd highest density - 27.2 to 39.6	55	61	*
Most dense (>39.6%)	63	61	
Proportion of non-manual heads of households (quintiles)			
Lowest density (<53.0)	65	63	
2nd lowest density - 53.0 to 61.8	60	59	
Middle quintile - 61.9 to 69.7	57	59	
2nd highest density - 69.8 to 78.0	54	64	***
Highest density (>78.0)	53	61	**
Total	58	61	**

* $p<0.05$ ** $p<0.01$ *** $p<0.001$

[6] Spearhead PCTs get extra funding to tackle health inequalities. They contain the 20% most health-deprived local authorities.

[7] http://www.communities.gov.uk/communities/neighbourhoodrenewal/deprivation/

Differences in response between Core and Boost
The differences in household response rate were compared between different categories of area-level characteristics. The Boost was found to have a significantly higher response rate in outer London PCTs (62% compared to 57% for the Core). There was a lot of overlap between spearhead status and location of PCT; 16 of the 18 outer London PCTs were also non-spearhead PCTs, hence the response rate in non-spearhead PCTs was also significantly higher for the Boost sample (62% compared with 55% for the Core).

There were a number of significant differences between response for different deprivation categories. The Boost methodology appears to work better in less deprived areas, while the Core performs better in slightly more deprived areas (but there is no significant difference in the most deprived areas).

There were few differences in response by the proportion of minority ethnic residents. The general trend is that the Boost is better in areas with a lower proportion of ethnic minorities. There are some differences by proportion of non-manual heads of households. The Boost sample had a significantly better response in areas with a higher proportion of non-manual households. There was no significant difference in areas with lower proportions of non-manual households.

Differences in response within each sample type
Comparisons between responding and non-responding households within each sample type showed that there was more variation in the level of household response for the Core sample than for the Boost. For the Core, there were significant differences between responding and non-responding households by IMD, PCT spearhead status and the proportion of households with non-manual heads ($p<0.05$), For the Boost sample there were significant differences in response by proportion of households with a non-manual head only ($p<0.05$).

3.2 Individual response

There were larger differences in the participation rates of eligible individuals within responding households. Individuals in the Core sample were more likely to give a productive individual interview once the household had responded. As before, the response rates for the Boost sample have been weighted by the address selection weights in order to make the Boost sample representative of London and the two samples comparable. There were 7714 adults (aged 16 years or over) in the 3882 responding Boost households (2.0 adults per household) and 1841 adults in the 963 responding Core households (1.9 adults per household), all of whom were eligible for the individual interview. 65% of the eligible adults in the Boost sample gave a productive interview[8], compared with 85% of the eligible adults in the Core sample. This gives 5004 productive adult respondents in the Boost sample and 1569 in the Core. The response rates for adults are given in

[8] The unweighted individual response is also 65% (4942 adult interviews out of 7654 adult household members in responding households).

Table 3.

Table 3 Individual response rates by sample type

		Individual response rates [a]
	N	%
HSE Core		
Productive CAPI interview	1569	85
No contact	131	7
Refusal before interview (personal)	26	1
Refusal before interview (proxy)	47	3
Refusal during interview	6	0
Broken appointment, no re-contact	4	0
Ill at home during survey period	6	0
Away/in hospital during survey period	23	1
Physically or mentally unable/incompetent	2	0
Language difficulties	20	1
Other reason for no interview	7	0
Eligible adults (16+) in responding Core households	1841	100
London Boost		
Productive self-completion questionnaire	5004	65
Personal refusal by named person	297	4
Proxy refusal (on behalf of named person)	218	3
Ill at home during survey period	10	0
Away/ at college / in hospital during survey period	65	1
Questionnaire not returned/completed	1701	22
Questionnaire returned blank	199	3
Language difficulty	77	1
Other reason	144	2
Eligible adults (16+) in responding Boost households	7714	100

[a] The Boost addresses have been weighted by the address selection weight.
The samples have different outcome codes because of the different modes used.

It is likely that the higher Core response rate is attributable to the presence of the interviewer and their ability to motivate reluctant household members to participate. Whilst the London Boost was designed to elicit the highest response possible from individuals within responding households (e.g. by encouraging interviewers to return to the household to collect the completed questionnaires), an interviewer would have less influence and control over self-completion questionnaires than in face-to-face interviews. In addition, the design of the Boost sample makes it possible for a household interview to be conducted but no individual interviews completed. This is unlikely in the Core sample as the individual who carries out the household interview will usually go on to complete an individual interview, meaning there is usually at least one individual interview within each responding Core household.

The following two tables show individual response by age group and household type for each sample. It can be seen that older age groups are more likely to respond in both samples[9]. Again, the figures for the Boost sample are based on data weighted by the address selection weights.

[9] There are a number of missing ages for non-responding Boost sample members.

Table 4 Individual response rates by sample type and age group (based on the number of eligible adults)

	16-34	35-54	55+	Missing	Total
	%	%	%	%	%
HSE Core					
Productive CAPI interview	80	87	91	0	85
No contact	12	5	2	40	7
Refusal before interview (personal)	2	1	1	20	1
Refusal before interview (proxy)	3	3	1	30	3
Refusal during interview	0	0	1	0	0
Broken appointment, no re-contact	0	0	0	0	0
Ill at home during survey period	0	0	1	0	0
Away/in hosp during survey period	2	1	1	0	1
Physically or mentally unable/incompetent	0	0	0	0	0
Language difficulties	1	1	1	10	1
Other reason for no interview	0	0	1	0	0
Base: Eligible adults (16+) in responding Core households	666	673	492	10	1841
London Boost					
Productive self completion questionnaire	58	70	73	9	65
Personal refusal by named person	3	3	5	16	4
Proxy refusal (on behalf of named person)	3	3	2	17	3
Person ill at home during survey period	0	0	0	0	0
Person away/ at college / in hospital etc during survey period	1	0	1	6	1
Questionnaire not returned/completed	29	20	14	36	22
Questionnaire returned blank	3	2	2	7	3
Language difficulty	1	1	1	1	1
Other reason	2	1	2	9	2
Base: Eligible adults (16+) in responding Boost households (weighted)	2864	2683	1996	171	7714

[a] The Boost addresses have been weighted by the address selection weight.

The breakdown of individual response by household type shows that Individuals in large adult households are least likely to respond, although the difference in response was greater for the Boost than for the Core.

Table 5 Individual response rates by sample type and household type (based on the number of eligible adults)

	1 adult aged 16-59, no children	2 adults, both 16-59, no children	Small family [b]	Large family[b]	Large adult house-hold[b]	2 adults, 1 or both aged 60+, no children	1 adult, aged 60+, no children	Total
	%	%	%	%	%	%	%	%
HSE Core								
Productive CAPI interview	100	84	90	88	72	91	98	85
No contact	0	9	5	2	15	4	0	7
Refusal before interview (personal)	0	2	1	1	3	0	0	1
Refusal before interview (proxy)	0	2	3	3	4	1	0	3
Refusal during interview (includes partials)	0	0	0	0	0	1	2	0
Broken appointment, no re-contact	0	0	0	1	1	0	0	0
Ill at home during survey period	0	0	0	1	0	1	0	0
Away/in hospital during survey period	0	1	1	2	2	0	0	1
Physically or mentally unable/incompetent	0	0	0	0	0	0	0	0
Language difficulties	0	0	0	3	3	0	0	1
Other reason for no interview	0	1	0	0	0	1	0	0
Eligible adults (16+) in responding Core households	163	328	330	153	520	218	129	1841
London Boost								
Productive self completion questionnaire	78	63	72	64	54	74	79	65
Personal refusal by named person	4	4	3	3	4	3	6	4
Proxy refusal (on behalf of named person)	0	3	2	3	4	3	0	3
Person ill at home during survey period	0	0	0	0	0	1	0	0
Person away/ at college / in hospital etc during survey period	0	1	1	1	1	0	0	1
Questionnaire not returned/completed	16	25	19	23	29	14	9	22
Questionnaire returned blank	1	3	2	3	3	3	2	3
Language difficulty	0	1	1	2	1	1	0	1
Other reason	1	1	1	2	3	2	2	2
Eligible adults (16+) in responding Boost households	539	1250	1407	621	2507	891	499	7714

[a] The Boost addresses have been weighted by the address selection weight.

[b] Small family: Up to 2 adults and 2 children. Large family: 3+ children and 1+ adults or 2 children and 2+ adults. Large adult household: 3+ adults and up to 1 child.

3.3 Total response

The overall individual response rates were 40% for the Boost (61% household x 65% individual) and 49% for the Core (58% household x 85% individual). This was contrary to original expectations. In designing the two surveys, the response assumptions were for the Boost to achieve somewhat higher response than the Core:

- the assumed response for the Core was 56% (65% household x 86% individual), based on HSE 2003 response rates
- the assumed response for the Boost was 59% (70% household x 84% individual). Since the Boost was modelled on the Welsh Health Survey (WHS) methodology, the assumptions were based on the achieved response for the 2003 WHS (78% household and 86% individual), making downward adjustments for the typically lower household response rates achieved in London.

The reasons why the Boost performed more poorly than expected are unclear. It may be due to the use of very experienced interviewers on HSE, while the Boost often used newer interviewers since the interview itself was much less complex than the full HSE. It could also be that the Boost methodology is less suited to urban areas such as London, where two contacts are required to obtain a productive interview (firstly dropping off the questionnaire, and subsequently returning to collect it) rather than only one contact (the HSE interview can be completed on first contact with all eligible adults if they are all present). Although face-to-face interviews generally take longer than the self-completion questionnaires, the importance of the interviewer in encouraging response should not be overlooked. Participants who are not disposed to complete a self-completion questionnaire may be unwilling to refuse an interviewer who has bothered to visit their house and ask them questions personally. Other possible explanations could involve the interactive nature of the face-to-face interview, which may be more interesting / appealing to participants than answering ‘exam-type’ questions.

Furthermore, the high proportion of people from ethnic groups other than White British may mean that English language proficiency is more of a problem in London than elsewhere. (In Wales, the questionnaires were in both English and Welsh, but no translated questionnaires were available for London as more than 300 first languages are spoken in the capital). The proportion of individuals in London who have problems reading and writing English sufficiently well to answer a self-completion questionnaire may be higher than the proportion who are unable to understand and answer spoken questions in English.

It would thus appear that, in London at least, there is no benefit in terms of response rates in adopting the WHS methodology; if anything, this methodology is likely to lead to lower individual response rates than the HSE methodology. However, it cannot be concluded that this would apply throughout England, as it may be a finding specific to highly urbanised areas such as London.

3.4 Item non-response

Item non-response is the term given to missing information on individual questions for respondents who have otherwise completed the questionnaire. Item non-response is generally higher for paper self-completion methods than either electronic self-completion or interviewer-assisted methods.

A comparison was made of the amount of item non-response in the Core and Boost samples. For the purposes of this analysis, item non-response has been designated as

missed questions, missing information and refusals. An explicit 'don't know' answer category was not provided for any of the questions in the Boost self-completion questionnaire. This means the focus of the analysis is on refusals and skipped questions, where no answer category at all has been ticked. This analysis is carried out on unweighted data.

As expected, the Boost survey had consistently higher levels of item non-response than the Core, although the level of item non-response for the Boost was generally low (less than 5%) for the majority of questions in the survey. This is encouraging, since a high level of item non-response is an indicator of poor data quality. There was, however, a fair degree of variation in the amount of item non-response, with levels being higher for more sensitive or complex questions.

It is likely that the different levels of item non-response can be attributed to the different modes of data collection. The Core interview was carried out in CAPI, which reduces the levels of item non-response in two ways: firstly, the interviewer can encourage the respondent to answer; and secondly, inadvertently missed items are reduced because the electronic CAPI questionnaire automatically filters to the correct next question.

Missing items are generally lower for the household component of the Boost survey, where the interviewer carried out a short household interview on paper. Data items in this section include respondent sex, age, household type, activity of the Household Reference Person (HRP)[10], NS-SEC of the HRP and whether any household members smoked. The proportion of responses missing for these questions is given in Table 6. The level for NS-SEC is higher as this variable requires a greater degree of information, but it is still low.

Table 6 Levels of item non-response for key variables from the household questionnaire by sample type

	Respondents with item missing	
	Core	Boost
	%	%
Respondent sex	0.0	0.0
Respondent age	0.0	0.0
Household type	0.0	0.0
Activity of HRP	0.8	2.5
NS-SEC of HRP (8 groups)	0.4	2.1
Smoker in household	0.1	1.8
Base (unweighted)	*1569*	*4942*

The gap between the Boost and Core widened for individual questions from the self-completion survey. Table 7 shows the levels of item non-response for a number of key individual variables.

Table 7 Levels of item non-response for key variables from the individual questionnaire by sample type

	Respondents with item missing

[10] The person in the household responsible for the accommodation, or if more than one person, the person with the highest income.

	Core	Boost
	%	%
Respondent ethnicity	0.4	2.5
Marital status	0.1	2.1
Respondent economic activity	0.3	5.9
Highest education qualification	0.3	9.2
Cigarette smoking status	0.9	2.4
Frequency drunk alcohol in last 12 months	1.1	2.1
Grouped portions of fruit eaten yesterday	0.1	3.0
Self-assessed general health	0.0	1.2
Long-standing illness or disability	0.0	1.9
Limiting long-standing illness or disability	0.0	1.9
Summary physical activity level	0.2	16.1
Base (unweighted)	*1569*	*4942*

Levels of item non-response are lower for straightforward questions with simple answer categories, such as general health. They are slightly higher for more sensitive questions, such as ethnicity, and for complex sections, such as the physical activity questions. Levels are particularly high for respondent economic activity, education and physical activity. These are discussed in more detail below.

Questions are missed for different reasons: the respondent may find a question sensitive and refuse to answer; the answer may be too cognitively demanding for the respondent to work out; or the respondent may have skipped the question in error. It is likely that the qualification question is an example of the latter category. The question contained a very high degree of item non-response; 9% of the Boost sample did not complete it, compared with less than 1% of the Core. The distribution of responses for the education question is given in Table 8. It can be seen that the Boost sample contains a much lower proportion of respondents reporting no qualifications (there is a 10% difference between the Core and Boost) but the proportions reporting each level of qualifications is otherwise very similar.

Table 8 Highest education qualification by sample type

	Core	Boost	Total
	%	%	%
Missing	0.3	9.2	7.1
Higher degree/Degree/NVQ4/5	34.8	34.5	34.5
NVQ3/GCE A Level equiv - any grade	10.6	13.5	12.8
NVQ1/2 GCE O Level equiv - any grade	25.5	24.2	24.5
Other (e.g. City and Guilds, RSA/OCR, BTEC)	3.5	3.2	3.3
No qualification	25.2	15.5	17.8
Base (unweighted)	*1569*	*4942*	*6511*

It may have been that the missing cases were overwhelmingly respondents with no qualifications, who may have skipped the question because they felt it was not relevant. The answer category ‘no qualifications’ was last in the list of responses, so it would have been relatively easy for respondents without qualifications to glance at the question and dismiss it as not relevant. If this were the case then the proportion of Boost respondents with no qualifications would be close to the Core figure.

The proportion missing for respondent economic activity is also high. Its possible that respondents felt the question to be sensitive, had trouble selecting an appropriate answer category, or overlooked the correct category so left it blank.

Table 9 Respondent economic activity by sample type

	Core	Boost	Total
	%	%	%
Missing	0.3	5.9	4.6
Full time study	6.9	7.9	7.6
In paid work	55.9	52.4	53.2
Looking for work	3.5	2.9	3.0
Ill (Long and short term)	3.8	3.7	3.7
Retired	18.0	14.3	15.2
Looking after home	10.9	9.6	9.9
Doing something else	0.8	3.3	2.7
Base (unweighted)	1569	4942	6511

In Table 9, the missing category includes Boost respondents who had not ticked an answer category and Core respondents who had refused to give an answer. In addition to the 6% of Boost respondents who skipped the economic activity question, 3% ticked ‘doing something else’, compared with less than 1% of Core respondents. It could be that these Boost respondents had trouble fitting their economic activity into the given answer categories, something an interviewer could have assisted with.

There were also instances where a respondent had incorrectly followed the filtering and skipped a follow-up question that they should have answered. Around 13% of current smokers in the Boost sample had not recorded the number of cigarettes they smoked. The corresponding figure for Core respondents was less than 1%. Likewise, respondents who had said they had drunk alcohol in the last 7 days were asked to give the number of units drunk on the heaviest day. 9% of boost respondents did not give this information compared with only 1% of Core respondents.

It is likely that these responses are missing because respondents did not follow the proper question routing, rather than because they were reluctant to answer these questions. There were very few refusals in the face-to-face interview, which suggests that respondents did not find these questions to be particularly sensitive. Previous research has also shown that respondents tend to be more honest about smoking and drinking behaviours in self-completion questionnaires than in face-to-face interviews.

The higher levels of item non-response for questions in the Boost survey mean that there are also much higher levels of missing data in the summary variables used during analysis. This is because these ’derived variables‘ usually combine the responses made to more than one question, and if a respondent has a missing case in just one of the component questions then they will be missing for the derived variable. This is particularly noticeable the physical activity derived variable, which summarises the responses made to a long series of questions. 16% of Boost respondents were coded as missing for this derived variable because of the high levels of missing data for the different components within the physical activity section. The levels of missing items are shown in Table 10; it is likely that the high levels are caused by the complex question layout (for a self-completion questionnaire) of this section.

Table 10 Missing data for physical activity questions by sample type

	Respondents with item missing	
	Core	Boost
	%	%
Summary physical activity level	0.2	16.1
Components:		
- Manual work	0.0	1.5
- Heavy housework	0.0	5.8
- Walking	0.0	5.3
- Sports	0.0	11.7
Base (unweighted)	*1569*	*4942*

The Core survey also included a self-completion booklet, which was given to respondents to fill in during the interviewer's visit. This booklet included questions on perceived social support and the General Health Questionnaire (GHQ12) questions, and was completed by 1343 Core respondents.

Table 11 Levels of item non-response for key variables from the self-completion questionnaire by sample type

	Respondents with item missing	
	Core	Boost
	%	%
General Health Questionnaire (GHQ12) score	1.2	2.9
Perceived social support score	2.4	3.5
Base (unweighted)	*1343*	*4942*

Although the levels of item non-response were more similar for these questions, the Boost levels were still higher than the Core. This may be partially due to the shorter length of the Core booklet and that it is completed with the interviewer present, who can thus assist the respondent with any queries and encourage completion.

Item non-response can bias survey estimates if the levels are high and the variable of interest is strongly related to the respondents' behaviour (e.g. if heavy smokers are more likely to refuse to fill in the number of cigarettes smoked). However, since there is usually no means of knowing this relationship, it is generally assumed that they are the same as respondents for whom valid data are available.

The survey estimates in this report are based on valid estimates only; no attempt has been made to impute missing data. For each comparison, the missing responses were excluded.

4 COMPARISON OF DEMOGRAPHICS

A comparison was made of the demographic profile of the achieved samples, looking at characteristics of the individual respondent (age, ethnicity, marital status, education and current economic activity), the respondent's household (household type, household reference person's (HRP's) current economic activity, HRP's NS-SEC) and the area in which the respondent lived (area-level deprivation (IMD) quintiles and PCT spearhead status). The analysis was carried out on adults aged 16 and over; men and women were analysed separately. The detailed tables are shown in Appendix D.

For this analysis, the data were weighted by selection weights only, which means the Boost sample has been adjusted for the unequal selection probabilities, making it comparable to the Core sample. No adjustments have been made for differential non-response bias.

The profiles of the two samples are generally very close (Table 12). Both samples have a higher proportion of women to men, as men are generally less likely to respond to survey research. This pattern is more evident in the Boost sample, but the difference between the two is non-significant. The samples were both weighted to ensure the sex profile matched that of the London population (see APPENDIX B).

Table 12 Sex distribution by sample type (selection weighting only)

	Core	Boost	Population
	%	%	%
Women	53.1	55.5	50.5
Men	46.9	44.5	49.5
Base (unweighted)	*1569*	*4942*	*7413268*
Base (weighted)	*1638*	*5189*	-

Population estimates taken from the 2005 mid-year Population Estimates produced by the ONS. An adjustment has been made to the estimates to give population in private households only.

A comparison of the age profiles to the population is given in

Table 13. The age profiles are very similar and are both relatively close to the population figures. There is a slight age bias in each sample: older people are over-represented in both the Core and Boost samples when compared to the population figures, although the differences are small. By contrast, younger people, especially younger men, are somewhat under-represented. This is typical of most general population samples, as young men tend to be harder to contact and, when found, more reluctant to take part.

Table 13 Age distribution by sample type (selection weight only)

Age (grouped)	WOMEN Core	Boost	Population	MEN Core	Boost	Population
	%	%	%	%	%	%
16-24	12.2	13.0	14.2	12.2	12.2	14.7
25-34	22.0	21.4	23.8	22.6	20.1	25.1
35-44	23.6	21.6	20.3	21.9	21.4	22.2
45-54	15.3	15.9	14.6	15.4	17.9	14.7
55-64	11.2	12.3	11.3	13.2	13.3	10.7
65+	15.8	15.8	15.9	14.8	15.0	12.7
Base (unweighted)	833	2726	3745524	736	2202	3667744
Base (weighted)	870	2871		768	2302	

Population estimates taken from the 2005 mid-year Population Estimates produced by the ONS. An adjustment has been made to the estimates to give population in private households only.

There is a significant difference in the proportion of respondents who belong to a non-White minority ethnic group, with more non-White respondents in the Core sample than in the Boost (Table 14). The Boost sample, however, is the closer of the two to the ethnic distribution of the population of London.

Table 14 Ethnic profile by sample type (selection weight only)

Ethnic group	WOMEN Core	Boost	Population	MEN Core	Boost	Population
	%	%	%	%	%	%
White	62.8	69.6	71.8	64.5	68.7	71.8
Mixed	2.7	2.9	2.4	1.6	2.2	2.4
Asian	17.3	13.6	11.9	18.7	15.7	12.9
Black	13.6	10.2	10.4	12.5	9.4	9.4
Chinese/other	3.6	3.8	3.6	2.7	4.0	3.4
Base (unweighted)	831	2672	3066300	732	2145	2949900
Base (weighted)	868	2808		764	2244	

Population estimates are from experimental Population Estimates by ethnic group for local authority districts and higher administrative areas in England for 2005 published on 22 August 2007, hence the bases differ from those in the earlier tables.

Amongst those reporting the highest educational qualification gained, there are significant differences between the sample types in the type and level of qualifications reported by respondents (see Appendix D, Table D1). Generally, we would expect Boost respondents to have higher levels of education than Core respondents, because those with poor reading skills may be put off by the self-completion format.

Moreover, the size of these differences may be exacerbated by differences in the administration of the question. In the Core survey, the interviewer gave the respondent a show card with a detailed list of qualifications and asked the respondent to indicate which ones they had. The interviewer was able to assist any respondents who needed help, and also to probe to make sure all qualifications were mentioned. It was impractical to include an equally long list of qualifications in the Boost self-completion questionnaire, due to space limitations, so the number of answer categories was

reduced for the Boost survey. The more detailed information collected in the Core survey was re-coded to match the less detailed categories used in the Boost survey education variable. It is this variable that is used for comparing the two sample types, and the comparison shows that, as expected, Boost respondents are better educated than Core respondents. Both men and women in the Boost were more likely to have higher degrees, degrees and A-levels, whilst Core respondents were much more likely to have no qualifications. However, among Boost respondents, there was a very high degree of item non-response to the qualification question, with 9% skipping this question. The problem of missing qualification data was discussed in section 3.4, and it seems likely that the missing data is playing a role in the observed difference between the samples.

There were small but non-significant differences between the sample types in the respondents' economic activity profile and marital status (see Appendix D, Table D1). There were no significant differences in the profile of responding households by household type, HRP NS-SEC, HRP current economic activity, area-level deprivation quintiles, or PCT spearhead status.

When the data are weighted for non-response (see APPENDIX B), the differences in educational qualifications remain, but the significant difference in ethnic profile is removed. The weighting also removes the slight age bias of the two samples, making the age profiles of both samples closer to that of the population. The non-response weighting should therefore be expected to improve estimates for key health variables, as the profile of the achieved samples become more similar to that of the London population on a range of variables (e.g. age, ethnicity).

4.1 Summary

The demographic profiles of the Core and Boost samples were generally very close. They were similar in terms of age and sex, with both samples slightly under-representing younger men compared with the London population. This was corrected by the non-response weighting.

There were no significant differences in sample profile by the current economic activity of respondents, marital status, household composition, current economic activity of HRP, NS-SEC of HRP, area-level deprivation indicators, or PCT spearhead status.

There were significant differences in the ethnic profile of respondents, with a higher proportion of Core respondents from non-White minority ethnic groups; however, the Boost sample was a closer match with London-wide population estimates of ethnicity. The differences in the ethnic profile between the Core and Boost samples were no longer significant once non-response weights were applied, as the proportion of White respondents in the Core sample increased, making it closer to the population estimates.

There were also significant differences in respondents' educational qualifications. Boost respondents had higher levels of qualifications than Core respondents, and these differences remained even after non-response weights were applied. A far lower proportion of Boost respondents had no qualifications. However, it is likely that the differences in education levels were in part caused by differences in how the information about qualifications was collected as well as the high levels of missing data for Boost respondents. As a result, the qualification questions used in the Core and Boost surveys may not provide a reliable measure of differences between data collection modes.

In summary, comparisons of the two samples show few significant differences between the socio-demographic characteristics of the achieved Core and Boost households and individual adults. Moreover, after non-response weighting has been applied, both samples seem to provide reasonably close correspondence with the London population for the characteristics examined.

5 COMPARISON OF HEALTH VARIABLES

5.1 Introduction

Results for a number of key health variables were compared for the two samples, including: general health, limiting and long-term illness, GHQ12 score, smoking, drinking, fruit and vegetable consumption, physical activity and social support. Tests were carried out to assess whether any differences in the survey estimates were statistically significant.

The analyses were run twice:

1. The first comparison looked for differences in weighted results. For this comparison the samples were weighted separately using the same weighting scheme (see APPENDIX B). Throughout this section, this is referred to as the *weighted* data.
2. The second comparison looked for the effects of differential measurement error by adjusting the socio-demographic profile of the Core sample to match that of the Boost sample. Any remaining differences found after the matching should be attributable to measurement error rather than sample composition. Throughout this section, this is referred to as the *matched* data[11].

The analyses were carried out separately for men and women.

Generally, estimates that were significantly different in the weighted data were also different in the matched data. This suggests that the larger differences in the survey estimates were not due to differences in sample composition or differential non-response but due to differences in measurement error. Most results given are for weighted data, but references to the matched data are given where appropriate. All tables can be found in APPENDIX D.

5.2 Health measures

5.2.1 General health

There were surprisingly large differences in the responses to the question about general health. This question, identically worded in both the self-completion and face-to-face questionnaires, asks respondents to rate their own general health. There are five response categories: very good, good, fair, bad and very bad. The difference in the distribution of responses to this question was highly significant (p= 0.000) for both men and women. The pattern was the same for both sexes, with Core respondents much more likely to say that their health was very good. Boost respondents were more likely to give an answer in the middle of the distribution and state their general health was fair. This pattern held for both weighted and matched data; the weighted data are given in Table 15.

[11] The Core sample was adjusted to match the unweighted Boost sample, so neither sample is representative of London and cannot be used to make generalisations. The samples should only be compared with each other, not the population. See APPENDIX C for more details on the Propensity Score Matching.

Table 15 Self-assessed general health for weighted data by sample type and sex

Self-assessed general health	WOMEN Core	WOMEN Boost	MEN Core	MEN Boost
	%	%	%	%
Very good	32.8	21.3	35.8	23.8
Good	45.4	51.1	43.6	48.3
Fair	14.0	23.3	14.0	23.8
Bad	6.1	3.6	5.5	3.1
Very bad	1.6	0.6	1.1	0.9
P-value		*0.000*		*0.000*
Very good / good	78.3	72.4	79.4	72.2
Fair/bad/very bad	21.7	27.6	20.6	27.8
P-value		*0.000*		*0.000*
Base (unweighted)	*833*	*2701*	*736*	*2181*
Base (weighted)	*811*	*2489*	*783*	*2428*

Although item non-response was higher for Boost respondents, this is unlikely to be the cause of the difference as only 1% of Boost respondents missed the general health question.

To check whether the concurrent interviewing could be as having an effect on responses of Core respondents living in households with more than one household member, the results for individuals who were interviewed alone were compared to those who were interviewed concurrently. There was no difference in the distribution of responses, which suggests the presence of other households members had no effect.

One possibility is that being able to visualise the scale on a paper questionnaire may have encouraged Boost respondents to go for the middle answer category. Also at play may be a form of satisficing, whereby respondents give the answer they feel the interviewer expects, i.e. the respondent may not want to seem to be complaining to the interviewer and therefore say they feel very good.

5.2.2 GHQ12

The General Health Questionnaire (GHQ) is used as an indicator of psychological health in many general population surveys. The questions cover general levels of happiness, depression, anxiety, sleep disturbance and ability to cope during the four weeks prior to interview. The GHQ used in the HSE and London Boost surveys contains 12 questions that ask the respondent to assess their present state relative to their usual, or normal, state (Goldberg and Williams, 1988). The respondent is then given a summary score based on their answers, with a higher score indicating poorer health. The GHQ12 score was grouped into three bands; 0, 1-3 and 4+. A score of 4 or more indicates poor psychological health. The proportion of respondents with a zero GHQ12 score was higher for Core respondents than Boost respondents; Boost respondents were more likely to have a score of 4 or more. The difference in scores was found in both the weighted and matched data, and the pattern was also the same for each.

The GHQ12 questions were administered in a self-completion format for both the Core and Boost samples, and the question wording and layout were identical for each. Given

that the mode of administration is the same, the expectation is for there to be no differences between the two sample types, especially after they had been matched. The figures for the matched data are given in Table 16. The difference in scores was significant at the 5% level for both men and women.

Table 16 GHQ12 score for matched data by sample type and sex

GHQ12 score - grouped	WOMEN Core	WOMEN Boost	MEN Core	MEN Boost
	%	%	%	%
Score 0	57.7	51.3	63.5	56.9
Score 1-3	26.8	28.2	24.2	26.8
Score 4+	15.5	20.5	12.3	16.3
P-value		*0.005*		*0.011*
Base (unweighted)	*691*	*2628*	*620*	*2143*
Base (weighted)	*2286*	*2628*	*1866*	*2143*

The matching should rule out differences in sample composition, at least for those variables that were available for matching. There could, of course, still be differences between the samples that were not recorded.

The responses could also have been affected by the context in which the questions were given and the environment in which the self-completion took place. The Core respondents would have had an interviewer, and possibly other household members, present whilst they filled in their self-completion booklet, and this may have affected their responses (e.g. to be more positive). By contrast, the Boost respondents were left the self-completion booklet to complete on their own, so may have had more privacy when providing their answers. Again we looked at the results for Core responses by number of people present during the interview. There were no significant differences in the responses of individuals interviewed alone and the responses of individuals interviewed concurrently. However, as mentioned, an interviewer would always have been present when Core respondents were completing their booklets, so they were never entirely alone, as would have been possible for Boost respondents when they completed their booklets.

The position of the question in the interview could be affecting the responses. The GQH questions were near the start of the Boost individual interview, whereas the self-completion booklet was administered near the end of the Core HSE interview, when the Core respondent was used to answering questions.

5.2.3 Long-standing illness

There were no significant differences in the proportion of respondents reporting long-standing or limiting long-standing illness. No patterns were seen for either men or women in the weighted or matched comparisons.

5.3 Health-related behaviours

5.3.1 Cigarette smoking

The two key estimates for smoking behaviour were current smoking status and number of cigarettes smoked per day. In addition there were two questions included in the household questionnaire about the smoking habits of all household members; whether any person in the household smokes cigarettes and, if so, how many.

There were no significant differences in the current smoking status of Core and Boost respondents; both were as likely to classify themselves as ex-, current or non-smokers. This held for both men and women and for estimates from weighted and matched data.

All current smokers were asked to record the number of cigarettes they usually smoked per day[12]. There were no significant differences in the number of cigarettes smoked by men. Men in the Boost sample reported a slightly higher number of cigarettes smoked per day, with a weighted mean of 13.5, compared with 12.4 for Core men. However, this difference was not significant in either the weighted or matched data.

The same pattern was seen for women: the number of cigarettes smoked was higher for Boost than Core respondents. Women smokers in the Boost sample recorded a weighted mean of 12.8 cigarettes per day, whereas the mean for Core respondents was 10.1. This difference was significant at the 5% level. The figures for the matched data were similar.

This suggests that questionnaire mode may have an effect on the responses of women smokers but not men. Women appear to report less smoking in a face-to-face interview than in a self-completion questionnaire. Since the difference was found in the matched estimates, it appears that it is not caused by differences in sample composition or response but is best explained by data collection mode. Similar results have been found in other research looking at the reporting of 'anti-social' activities; these studies have shown that self-completion questionnaires are more likely than personal interviews to elicit honest responses about potentially sensitive behaviours (Tourangeau and Smith 1996, Schwarz et al 1991). Indeed, the findings for alcohol consumption show a similar pattern (see Section 5.2.2).

There were no significant differences between sample types for the household smoking questions. These questions were asked during the household interview and were administered by the interviewer in both surveys. We would not expect mode differences for these questions, but the analyses indicate there are also no significant differences due to sample composition or response bias.

5.3.2 Alcohol consumption

The key estimates on alcohol consumption include: the usual frequency of drinking in the last year, whether or not the respondent drank in the last week, the number of units drunk on the heaviest drinking day in the last week, and the proportion of respondents who were binge drinking in the last week. Binge drinking is defined as drinking more than twice the daily-recommended amount on a single day, which is more than 8 units for men and more than 6 units for women on the heaviest drinking day in the last week. The analysis was only carried out on respondents aged 25 or over. Some Core

[12] This information was missing for 2% of women and 3% of men in the Boost sample, see Section 3.4.

respondents aged 18-24 were given the drinking questions in a self-completion format, hence the comparison by modes would be confounded if they were included.

Responses to the questions on whether the respondent had drunk in the last 7 days and how frequently they usually drank in the last 12 months were very similar for both sample types. Boost respondents were slightly less likely to say they ever drank alcohol, but the difference was not significant for either sex for weighted or matched data.

Respondents were asked to record the number and type of alcoholic drinks consumed on the day that they drank the most in the previous week. The information was recorded as glasses or bottles drunk and this was converted to units of alcohol by the research team[13]. The mean number of units drunk was significantly higher for both men and women Boost respondents. Using weighted data, men in the Boost sample drank 8.8 units on the heaviest day in the last week, compared with 7.1 for Core respondents. The corresponding figures for women were 6.0 and 4.7, respectively.

The difference between the mean units drunk was also significant for the matched data; again the Boost sample recorded more units drunk than the Core sample, which suggests an effect due to mode of data collection. The mean number of units consumed for both weighted and matched data are shown in Table 17.

Table 17 Mean units consumed on the heaviest drinking day in the last week for respondents aged 25 and over, by sample type and sex

	WOMEN			MEN		
	Core	Boost	P-value	Core	Boost	P-value
	Units	Units		Units	Units	
Weighted	4.7	6.0	0.000	7.1	8.8	0.000
Matched	4.6	5.9	0.000	6.5	8.6	0.000
Base (unweighted)[a]	358	1079		425	1167	

[a] All respondents aged 25 and over who drank in the last week

The differences in the number of units recorded by each sample type are also reflected in the data on binge drinking. Using weighted data, 44% of women in the Boost sample who had drunk in the last 7 days had been binge drinking, compared with 33% of females in the Core. The corresponding weighted figures for men were 50% for Boost and 34% for Core. The pattern was the same in the matched data.

In summary, the two surveys gave similar estimates of the proportion of men and women who drank in the last week and of the usual frequency of drinking. However, there were large differences in the number of units recorded, with men and women in the Boost reporting a greater number of units than Core respondents.

As mentioned in Section 5.2.1, these results are consistent with those found for the smoking questions, as well as with other research about reporting sensitive behaviours, which shows that respondents tend to be more honest in self-completion questionnaires and are likely to under-report in face-to-face interviews.

However, the self-completion format may lead to other problems: e.g. data quality may suffer if complex filtering is required, or if the questions are cognitively demanding and

[13] Details on the conversion of questionnaire data into alcoholic units is given in Volume 1 of the 2006 HSE main report. http://www.ic.nhs.uk/pubs/hse06cvdandriskfactors

respondents struggle accurately to record the information required without an interviewer there to assist them. This was in fact found when comparing the number of missing values on the variable which computed the units drunk on the heaviest drinking day: while only 0.5% were missing for Core respondents, the level was much higher for Boost respondents at 5.9%.

5.3.3 Fruit and vegetable consumption

Respondents were asked to give details about the different types of fruit and vegetables they had consumed the previous day. This information was then used to calculate the number of portions of fruit and vegetables eaten on that day, which was used to check whether respondents were meeting government recommendations for consuming at least five potions of fruit and vegetables per day[14].

Boost respondents reported consuming more portions of fruit and vegetables than Core respondents. The differences between sample types are quite large. Using weighted data, 54% of men in the Boost sample met the government's 5-a-day recommendation compared with 38% of men in the Core. Similarly, 55% of women in the Boost consumed five or more portions compared with 42% of women in the Core (Table 18). These differences were also found in the matched data.

Table 18 Grouped portions of fruit and vegetables for weighted data by sample type and sex

Portions of fruit and vegetables - grouped	WOMEN Core	WOMEN Boost	MEN Core	MEN Boost
	%	%	%	%
Less than 1	4.7	3.9	7.6	5.6
1 to 2	26.0	14.7	26.4	16.7
3 to 4	27.3	26.9	28.1	23.7
5 or more	42.0	54.5	37.9	54.1
P-value		*0.000*		*0.000*
Base (unweighted)	*833*	*2647*	*735*	*2133*
Base (weighted)	*811*	*2436*	*782*	*2375*

An examination of the original data from which the summary variable was derived showed that the number of portions recorded by Boost respondents was consistently higher than the number recorded by Core respondents. Boost respondents were more likely to say they had eaten each category of fruit and vegetables listed and, where a particular category of fruit or vegetable had been recorded, they generally entered a greater number of portions. For the majority of fruit and vegetable categories, the mean number of portions consumed by both men and women in the Boost sample was significantly higher than for Core respondents. The differences in the recorded amounts were consistently higher and were not caused by a small number of unusually high responses or outliers. The proportion of respondents eating each category of fruit and vegetables, and the mean number of portions recorded is given in Appendix D, Tables D5, D6 and D7.

[14] When generating the summary measure, portions of fruit juice, dried fruit and pulses were capped, so they could only ever equal a maximum of one portion each, regardless of how many were actually consumed. This was done in accordance with Food Standards Agency guidelines. More details on the summary measure can be found in Volume 1 of the 2006 HSE main report http://www.ic.nhs.uk/pubs/hse06cvdandriskfactors

The format of the fruit and vegetable module appears to be causing Core and Boost respondents to report differently. It could be that Boost respondents were not strictly recording one day's worth of vegetable consumption (although it is clear in the instructions that the questions refer to food consumed the previous day) or were having problems with the measurements.

The fruit and vegetable modules vary considerably between the Core and Boost surveys, with important differences in question format. During the Core CAPI interview, respondents were asked whether they had eaten a certain category of fruit or vegetables; for those replying in the affirmative, the interviewer asked a series of follow-up questions to establish how many portions had been eaten. However, the extensive filtering involved in the CAPI questionnaire is not appropriate for a self-completion format, so the Boost module included a list of all the categories of fruit and vegetables, with boxes for the respondent to record the amount eaten. Such a format may have encouraged respondents to give a positive answer to each type of fruit and vegetable listed. By contrast, it is also possible that the CAPI format may encourage under-reporting, since it becomes clear to respondents that a positive response to the initial question leads to an extra set of questions.

5.3.4 Physical activity

The physical activity measure examined was the proportion of respondents meeting the government's recommendation of 30 minutes or more of at least moderate intensity exercise on five or more days a week. Both the Core and Boost surveys collected participation in various sports and other activities. Respondents were asked to record the number of days on which they carried out the activity, the average amount of time spent doing the activity, and whether the activity caused them to get out of breath or sweaty.

In the face-to-face interview, respondents were presented with a show card listing 10 different activities, which were also read out one by one by the interviewer. If respondents had participated in an activity, they were asked a series of follow-up questions on the amount of time they spent doing that activity and whether they got out of breath. The amount of filtering used in the CAPI was much too complex for a self-completion format, so Boost respondents were provided with a list of the same 10 activities (and space to add up to three extra). The list included boxes for respondents to record the number of days they did each activity in the last four weeks, and the list was repeated for respondents to record the amount of time they spent doing each activity, and to tick whether or not they had become out of breath or sweaty.

Some of the questions were asked in a similar way on the Core and Boost surveys, including those on housework, gardening and other manual work. However, Boost respondents were required to follow the routing for these questions, in contrast to the Core respondents where the CAPI routing is done automatically. It appears that a small number of Boost respondents missed the filters or skipped questions they should have completed and so were missing on the physical activity derived variable.

The summary activity variable was derived using the information collected on separate sports and activities. The summary variable records the average number of days per

week that respondents did at least 30 minutes of a moderate to vigorous activity in the four-week period before the interview and defines three physical activity groups[15]:

- Group 1 = 'high', at least 30 minutes of moderate activity on five or more days per week on average in the last four weeks
- Group 2 = 'medium', at least 30 minutes of moderate activity on one to four days per week on average in the last four weeks, and
- Group 3 = 'low', at least 30 minutes of moderate activity on less than one day per week on average in the last four weeks.

The analysis includes respondents aged 16-64 only, since the physical activity module was not asked of all older respondents in the Core survey. The summary variable for physical activity shows that Boost respondents were generally less active and were much less likely to meet the government's recommendations. The figures were very similar for weighted and matched data; the differences were in the same direction and were significant for both analyses. The weighted data are given in Table 19, and show that: 22% of men in the Boost and 44% of men in the Core were in Group 1 (high), whereas 44% of Boost men and 28% of Core men were in Group 3 (low). For women the weighted figures were 18% of the Boost and 34% of the Core in Group 1 (high) and 43% of Boost and 32% of the Core in Group 3 (low).

Table 19 Summary physical activity level for respondents aged 16-64 for weighted data by sample type and sex

Summary physical activity level	WOMEN		MEN	
	Core	Boost	Core	Boost
	%	%	%	%
Group 1 - high	33.5	17.8	44.3	22.0
Group 2 - medium	34.1	39.3	27.5	34.6
Group 3 - low	32.4	42.9	28.2	43.5
P-value		*0.000*		*0.000*
Base (unweighted) [a]	695	1957	620	1593
Base (weighted)	682	1837	681	1841

[a] All respondents aged 16-64

These are large and highly significant differences for both men and women. The physical activity module is cognitively demanding for respondents, who are asked to recall how often they have participated in various activities over the last four weeks. For Boost respondents, there was no interviewer on hand to help with this task, whereas Core respondents would have been able to receive help. Both the complexity of the filtering and the availability of the interviewer are likely to have influenced the results.

This can be seen at a basic level of missing items, as described earlier in Section 3.4. Boost respondents had much higher levels of missing data for the activity questions. Since the individual questions are combined to derive the summary activity measure, missing just one of the component variables means the summary variable cannot be computed. The summary physical activity measure is missing for 16% of Boost respondents, compared with 0.2% of Core respondents. Such a high level of missing data will almost certainly affect the quality of the Boost activity estimates and it is also likely to bias the results, since those who participated in few or no activities are more

[15] Details on the physical activity summary variable can be found in Volume 1 of the 2006 HSE main report. http://www.ic.nhs.uk/pubs/hse06cvdandriskfactors

likely to have complete data enabling the summary activity category to be derived. By contrast, the more activities reported, the more additional questions that need to be answered, and thus the greater the scope for missing a question which would exclude more active individuals from the summary variable.

The data shown in Table D8 (Appendix D) suggest that this is likely to be at least part of the explanation. The actual rates of participation in each activity are generally higher for Boost respondents than for Core respondents, and for activities where the differences in participation are statistically significant, this was always the case. Despite having higher participation rates, the summary measure showed Boost respondents to be less active, because the more activities the respondent participated in, the more likely they were to have missing data.

There were 612 Boost respondents with a missing physical activity summary variable. Of these 276 (45%) were doing at least one sport but were missing information on other components, such as time spent doing manual labour, heavy housework, etc. In addition, there were 519 cases where the Boost respondent had given information on all components of the physical activity section and had been assigned a summary variable but were still missing some information on sport activities, i.e. their second or third sporting activity contained missing information and could not count towards the summary. This would underestimate the proportion of respondents doing moderate levels of physical activity.

Another reason for the lower overall activity level of the Boost respondents, is that, for a particular activity, they were less likely than Core respondents to say they were made out of breath or sweaty.

There may also be other factors at work. During the Core interview the interviewers are careful to explain to respondents that they only want activities carried out in the previous four weeks. Some respondents, particularly those that carry out seasonal sports, may feel this to be unfair as the answers they are being asked to provide are not typical. The Boost respondents do not have an interviewer to give them such strict guidance and may include activities that were outside the relevant time period. This may be contributing towards the higher participation rates for the Boost respondents.

5.3.5 Social support

The section on social support in the Boost questionnaire was taken directly from the HSE self-completion booklet, hence the format and layout of the questions were identical. The social support questions collect information on the amount of support and encouragement an individual receives from friends and family. There are seven social support questions that ask about different emotional aspects of social support. A single scale is produced by assigning a score to the responses from each question, ranging from a score of 1, if the respondent reports a lack of social support, to a score of 3 for no lack. The scores for all seven questions are added together to give an overall score with a maximum value of 21. This is grouped into three categories of social support: a score of 21 is no lack, less than 21 but more than 18 is some lack and less than 18 is severe lack.

Women in the Boost sample were more likely to be at either end of the scale than were women in the Core, i.e. they were more likely to say they have a severe lack, or no lack, of social support. These findings are opposite to the pattern seen for self-reported general health, where the Boost respondents tended to give answers that placed them

in the middle category. The differences are significant for both weighted and matched comparisons, although the actual percentage differences are not especially large.

There is a different pattern for men. In the Boost, men were more likely to report no lack of social support and less likely to say they have some or severe lack. The differences are significant for weighted estimates, but are no longer significant for the matched comparisons.

Table 20 Perceived social support score for matched data by sample type and sex

Perceived lack of social support - grouped	WOMEN Core	WOMEN Boost	MEN Core	MEN Boost
	%	%	%	%
No lack	61.7	63.8	54.6	59.0
Some lack	25.0	20.2	23.6	21.3
Severe lack	13.4	16.0	21.8	19.7
P-value		*0.036*		*0.223*
Base (unweighted)	*701*	*2665*	*626*	*2133*
Base (weighted)	*2323*	*2665*	*1881*	*2133*

Whilst both questionnaires were administered in a self-completion format, the interviewer was still present for the Core respondents, which may have affected respondents' answers. The responses given for Core sample members were checked to see if the presence of other household members affected the response, but this was found not to be the case.

5.4 Summary

The aim of these analyses was to look for differences in measures of health and health-related behaviours by mode of data collection. There were no significant differences between the Core and Boost samples for a number of variables including: long-term illness, limiting long-term illness, current rates of smoking, whether respondents drank alcohol and how often they usually drank. However, there were a number of differences, some quite large, between some key measures including: general health, GHQ12 score, number of cigarettes smoked, number of alcohol units consumed on the heaviest drinking day in the last week, portions of fruit and vegetables consumed, levels of physical activity, and, to a lesser extent, perceived social support (among women only).

The estimates were similar between sample types for simple measures of whether or not respondents currently smoked cigarettes or consumed alcohol. However, significant differences were found between the Boost and the Core for amounts of consumption. The Boost data showed significantly higher estimates of the number of cigarettes smoked by women and the number of alcohol units drunk by both men and women. In line with previous research, this suggests that respondents are more likely to report socially undesirable behaviour in a self-completion format than in a face-to-face interview.

However, Boost respondents also reported much higher levels of fruit and vegetable consumption: this was found not only for the overall summary measure, but also for each of the separate categories of fruit and vegetables included in the questionnaire.

This is contrary to the hypothesis that respondents are more likely to be honest in a self-completion format, as it appears that Boost respondents may have over-reported socially desirable behaviour. The explanation may be due to a combination of factors including: Core respondents in the face-to-face interview avoiding having follow-up questions to answer; and, lack of interviewer guidance for Boost respondents, who may not have read or fully understood the instructions in the self-completion booklet.

There were also large differences in the level of physical activity reported by Core and Boost respondents. While Boost respondents reported far lower levels of physical activity, their reported rates of participation in sports and other activities were generally higher. However, the high level of missing items for the more active Boost respondents explains a large amount of the difference between sample types and highlights the difficulties of including complex routing in a self-completion format.

As well as differences in mode of data collection between the Core and Boost surveys, there were also differences in questionnaire context and the presence or absence of the interviewer. For example, Boost respondents were more likely to state their general health to be good or fair, while Core respondents were more likely to use the extreme categories and say their general health was very good or bad/ very bad. These differences may well be due to the interviewer/respondent interaction.

Finally, a few questions were asked in a self-completion format in both the Core and Boost surveys, yet had different results. The GHQ12 and social support modules were asked identically in both sample types, yet there were still large differences in the results, e.g. Boost respondents scored higher on the GHQ12 questions than Core respondents. It may be that the differences were caused by the context in which the questionnaire was placed: having an interviewer present whilst completing the booklet could have affected Core respondents' answers, even though the interviewer was not actually asking the questions.

For those measures where the estimates on the matched samples are significantly different between the Core and the Boost surveys, it is difficult to determine how much of the difference may be caused by data collection mode or other effects, and how much of the difference may be real.

6 COMPARISON OF HEIGHT AND WEIGHT MEASUREMENTS

6.1 Height and weight

Height and weight were collected from respondents in both the Core and Boost surveys. Interviewers were asked to measure respondents but if this was not possible then self-reported heights and weights were obtained. This section compares: response to the height and weight measurements for each sample type; mean heights and weights measured by the interviewer for each sample type; interviewer-measured with self-reported height and weight measurements (in the Boost sample only); and mean heights and weights for all respondents with a measurement (whether interviewer-measured or self-reported). Height and weight were used to calculate Body Mass Index (BMI), and mean BMI and the proportion of respondents in each BMI category are also compared across sample types. The analyses were done separately for men and women.

Interviewers measured the height and weight of Core respondents following strict protocols to ensure accurate and robust readings. The majority of adult respondents allowed themselves to be measured: height was obtained for 83% of Core respondents and weight for 81%. A small number of exclusions were allowed, e.g. if the respondent was unable to stand, unsteady or found it difficult or painful to stand straight. Pregnant women were not eligible for the weight measurement. If the height or weight measurement was not attempted or the respondent refused, the interviewer asked the respondent to estimate their height or weight. This was done for 13% of the sample.

The procedure was slightly different for the Boost sample, as respondents could have their height and weight collected twice, once by an interviewer, if the respondent was available at either interviewer visit, and once as a self-report in the self-completion questionnaire. Interviewers measured height and weight during either the household interview or the return visit to collect the questionnaires, and followed the same protocols as Core interviewers. However, Boost interviewers did not make return trips to the address specifically to measure household members who may have been missed at earlier visits; as a result, height and weight measurements were not attempted for 23% of the Boost sample, and there are height and weight measurements for just over half (55%) of Boost respondents (which is 71% of the individuals asked).

In addition to the measurements collected by the interviewer, Boost respondents were asked to record their height and weight in the self-completion questionnaire[16]. There are self-reported measures of height for 88% of Boost respondents and weight for 73%. Just under half of Boost respondents had both interviewer-measured and self-reported height and weight: 2414 (49%) of Boost respondents have both measures of height and 2107 (43%) have both measures of weight. The data for this analysis are unweighted because the comparison is made of repeated readings for the same individuals. The estimates are not being used to make inferences about the population.

[16] Some respondents would have filled in the self-completion questionnaire after the interviewer had taken their measurements, others before. We would expect more accurate measurements from respondents who filled in the self-completion form after measurements were taken, but there is no means of identifying these respondents.

6.1.1 Comparison of interviewer-measured height and weight

The analysis includes only those cases where an interviewer took the height or weight measurement. As expected, the mean height and weight for the two samples are very close. There are no significant differences between the two samples for either men or women. The mean interviewer-measured height and weight by sample type are given in Table 21.

Table 21 Interviewer-measured height (cms) and weight (kgs) by sample type

		Unweighted sample size	Minimum	Maximum	Mean	Std. Deviation	P value
Height							
Women	Core	691	103.5	183.0	161.4	0.32	0.355
	Boost	1563	116.2	188.5	161.1	0.20	
Men	Core	622	151.1	197.0	174.3	0.35	0.687
	Boost	1236	124.0	198.0	174.1	0.24	
Weight							
Women	Core	663	33.2	122.0	67.4	0.47	0.797
	Boost	1533	35.2	155.6	67.5	0.34	
Men	Core	614	44.6	130.0	80.8	0.68	0.443
	Boost	1229	10.5	222.4	80.1	0.40	

All heights recorded by an interviewer

6.1.2 Interviewer-measured and self-reported heights and weights (Boost respondents only)

Paired t-tests showed the differences in interviewer-measured and self-reported mean height and weight were significantly different. Self-reported height was significantly higher than height measured by the interviewer, across both sexes and all age groups. Self-reported weight was significantly lower than weight measured by the interviewer, among both men and women. The mean height and weight for Boost sample members with both measured and self-reported heights and weights is given in

Table 22.

Table 22 Boost respondents with interviewer-measured and self-reported height (cms) and weight (kgs)

		Unweighted sample size	Minimum	Maximum	Mean	Std. Deviation	P value
Height							
Women	Self-report	1318	129.5	182.9	162.1	7.4	0.000
	Interviewer-measured	1318	136.0	188.5	161.6	7.0	
Men	Self-report	1096	129.5	233.7	175.4	8.1	0.000
	Interviewer-measured	1096	146.0	198.0	174.4	7.8	
Weight							
Women	Self-report	1096	33.6	142.0	66.2	13.8	0.000
	Interviewer-measured	1096	37.1	155.6	67.2	14.1	
Men	Self-report	959	31.0	154.0	79.7	14.6	0.000
	Interviewer-measured	959	10.5	222.4	80.4	15.6	

6.1.3 Comparison of all respondents with height and weight data

The next analysis compares height and weight for all respondents with data available: i.e. self-reported height and weight is used if interviewer-measured data were not obtained. The results show that the mean height of Boost respondents is significantly higher than that of Core respondents, for both men and women. This is explained by self-reported height being higher than interviewer-measured height, and the fact that self-reported height was much more common in the Boost sample than in the Core sample. Looking at weight, Boost respondents have a slightly lower mean weight than Core respondents, but the difference is not statistically significant for ether sex. The mean height and weight of all respondents with a measure available is given in Table 23.

Table 23 Height (cms) and weight (kgs) using all available measures, by sample type

		Unweighted sample size	Minimum	Maximum	Mean	Std. Deviation	P value
Height							
Women	Core	811	103.5	183.0	161.4	0.30	0.008
	Boost	2608	116.2	188.5	162.3	0.16	
Men	Core	711	151.1	198.1	174.5	0.32	0.031
	Boost	2116	124.0	213.4	175.3	0.19	
Weight							
Women	Core	784	21.3	132.0	67.5	0.50	0.149
	Boost	2335	35.2	155.6	66.7	0.28	
Men	Core	707	44.6	142.5	80.9	0.74	0.594
	Boost	1936	10.5	222.4	80.5	0.33	

Base: All respondents with a measure of height or weight available

The difference between the two measures was also tested using the methods outlined by Bland and Altman in their 1986 paper. The aim of the test is to measure the amount of agreement between the height and weight measures made by the interviewer and the respondents' self-reports. The difference of the two measures was calculated (interviewer-measured minus self-report) and the mean difference found. A set of confidence intervals was generated for the mean difference, which are the intervals in which the mean difference is expected to fall 95% of the time. Bland and Altman refer to these as the 'limits of agreement'; wide confidence intervals demonstrate a lack of precision between the two measures.

The mean difference for height is negative because interviewer-measured heights are generally lower than the self-reported height. The reverse was true for weight, hence the mean difference for weight is positive. The mean differences and limits of agreement are given in Table 24.

Table 24 Mean difference in height (cm) and weight (kg)

	Height		Weight	
	Women	Men	Women	Men
Mean difference between two estimates	-0.50	-1.05	0.96	0.71
Std dev of difference	3.77	3.88	4.77	5.78
Estimated limits of agreement				
Lower	-7.9	-8.7	-8.4	-10.6
Upper	6.9	6.6	10.3	12.0

For both men and women, the limits of agreement for height and weight are fairly wide, indicating poor agreement between the interviewer measurements and self-reports. This reinforces the findings of the paired t-test, which showed a significant difference in the means between the measurements and self-reports.

Interviewer-measured height for women may be 8 cm below or 7 cm above self-reported height. This means there is a large interval in which self-reported height will fall, although it is more likely to fall below the interviewer measured height. Interviewer-measured weight for women may be 8 kg below or 10 kg above self-reported weight, which again is a wide interval. Self-reported weight is more likely to be below the interviewer measurement. The largest interval is for men's weights: the interviewer measurement may be 11 kg below or 12 kg above the self-reported weight. Interviewer-measured height for men could be 8 cm below or 10 cm above the self-reported height.

6.2 Body Mass Index

6.2.1 Body Mass Index (BMI) measurement

Height and weight are used to generate Body Mass Index (BMI). BMI is a measurement that allows for differences in weight due to height and can be used to estimate the proportion of respondents who are overweight or obese. It is defined as weight (kg) / height (m^2). BMI was calculated for all respondents for whom both valid height and

weight measurements were recorded. Respondents were classified into the following BMI groups:

BMI	Description
Less than 18.5	Underweight
18.5 to less than 25	Normal
25 to less than 30	Overweight
30 or more	Obese

Similar to the height and weight comparisons, the following analysis was carried out twice: firstly, for those with interviewer-measured heights and weights; and secondly, including those with self-reported height and weight when interviewer measurements were not available. The results are shown in Table D9, Appendix D.

6.2.2 Comparison of BMI for interviewer-measured height and weight

Mean BMI and the proportion of respondents in each of the BMI categories was compared by sample type for those respondents who had interviewer-measured height and weight. There are no significant differences for either men or women in mean BMI or in the distribution of respondents between the BMI categories. As expected, the figures for the Core and Boost are all very close.

6.2.3 Comparison of BMI for all respondents with height and weight data

When including self-reports for respondents where interviewer measurements were not available, it was found that Boost respondents had a lower mean BMI and were less likely to be classified as obese than Core respondents. These differences were significant for women but not for men.

6.3 Summary

There are no differences in the height, weight and BMI of Core and Boost respondents when looking at interviewer-measured data only. The readings taken by interviewers are accurate and consistent. However, when self-reported heights and weights are included, significant differences are found between the sample types. As shown in Section 6.1.2, respondents tend to report being taller and lighter than they actually are, so estimates of height and weight are biased when self-reported. Since the Boost sample contains a much larger proportion of self-reported heights and weights, significant differences are introduced between the sample types when self-reports are included.

The differences between Core and Boost respondents in height and weight measurements are reflected in differences in BMI. There are no differences in BMI between sample types when the analysis includes interviewer measurements only. However, when self-reported heights and weights eare included, mean levels of obesity are under-estimated for Boost respondents.

7 REFERENCES

Bland JM, Altman DG (1986):Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986,**1**,307-10.

Craig R and Mindell J (2006) Health survey for England 2006: Volume 1 Cardiovascular disease and risk factors in adults.

Tourangeau R, Smith TW (1996): Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context. The Public Opinion Quarterly 1996, Vol 60, No 2, 275-304

Schwarz N, Strack F, Hippler HJ and Bishop G (1991) The impact of administration mode on response effects in survey measurement. Applied Cognitive Psychology, 5(3), 193-212

APPENDIX A VARIABLES INCLUDED IN THE ANALYSIS

Individual
Age (6 groups)
Sex
Ethnicity of respondent
Highest education qualification
Respondent economic activity status
Marital status

Household
(D[a]) Household Type
(D) NS-SEC 5 variable classification (HRP[b])
HRP[b] economic activity status

Area
IMD population quintiles for London only
PCT spearhead status (new classification of PCTs)

Health
(D) Self-assessed general health - grouped
Long-standing illness
(D) Limiting long-standing illness
(D) GHQ12 Score - grouped (0,1-3,4+)

Health related behaviours
Whether any household member smokes
Number of household members who smoke
(D) Cigarette Smoking Status - Never/Ex-regular/Ex-occasional/Current
(D) Cigarette Smoking Status - Banded current smokers
(D) Cigarette Smoking Status - Current/Ex-Regular/Never-Regular
(D) Number of cigarettes smoke a day - including non-smokers
(D) Frequency drink alcohol in last 12 months - including non-drinkers
Whether drank in last 7 days
Binge drinking in last week
(D) Portions of fruit (including fruit juice) & veg yesterday
(D) Grouped portions of fruit (including fruit juice) & veg yesterday
(D) New summary physical activity level
(D) Perceived social support score - grouped

Measurements
Interviewer-measured height (Boost respondents)
Interviewer-measured weight (Boost respondents)
Self-reported height (Boost respondents)
Self-reported weight (Boost respondents)
Height (Core respondents)
Weight (Core respondents)
BMI - based on interviewer measurements only (Core and Boost)
BMI - (interviewer-measured) grouped:<18.5,18.5-<25,25-<30,30+
BMI - including estimated self-reports (Core and Boost)
BMI - (all measures) grouped:<18.5,18.5-<25,25-<30,30+

[a] Derived variable

[b] HRP: Household Reference Person, i.e. the person in the household responsible for the accommodation, and if more than one, the person with the highest income.

APPENDIX B OVERVIEW OF WEIGHTING SCHEME

The London Boost sample was drawn from the 31 Primary Care Trusts[17] (PCTs) in London. It was run alongside the Heath Survey for England in 2006 to enable the data from the two samples to be analysed in conjunction. The analysis was carried out on data from the London Boost sample (for clarity this will be referred to as the Boost) and London addresses in the Core HSE sample (hereafter referred to as the Core).

Three sets of weights were generated for this project to enable different types of comparisons to be carried out. These were:

1. Selection weights
2. Non-response weights
3. Propensity score matching weights.

The selection weights correct for the effects of over-sampling addresses in smaller PCTs in the London Boost and the selection of households at multi-household addresses. The selection weights ensure the distribution of Boost addresses is correct across PCTs. By applying the selection weights it is possible to make assertions about the differences in the size of unweighted estimates as though the two samples had both been drawn with equal probability.

The non-response weights correct for the effects of differential non-response. The same methodology was used to create non-response weights for each sample, which is described in more detail below. These weights allow a comparison of key estimates that have been corrected for both unequal selection and non-response.

The propensity score matching weights adjust the Core sample to make it match the Boost on a number of socio-demographic variables. These are discussed in APPENDIX C.

Selection weights

Address selection weight (w_0)

The Boost sample was designed to give an equal number of interviews per PCT, with an additional boost in Camden and Islington. Six Primary Sampling Units (PSUs [18]) were selected from each London PCT, except Camden and Islington, where 14 PSUs were selected in each PCT. In addition, a larger number of addresses were selected from inner London PCTs, as the response rates in inner London were expected to be lower. There were 40 addresses selected in each inner London PCT, compared with 34 in outer London.

Address selection weights are needed for the Boost sample as the sampling probabilities for Boost addresses vary by PCT. The smaller PCTs have larger selection probabilities as the addresses within them were being sampled at a higher rate. Without selection weights the smaller PCTs would be over-represented in the sample. The Core sample does not require selection weights since each Core address had an equal chance of being selected.

[17] The sampling was carried out in September 2005, and the 2005 PCT codes are used in this report.
[18] PSUs were postcode sectors or groups of postcode sectors. Postcode sectors with fewer than 500 addresses were grouped with neighbouring sectors.

The address selection weight (w_0) was generated as the inverse of the address and PSU selection probabilities. This was equal to 1 for the Core sample, since the addresses in the Core were selected with equal probability.

Household selection weight (w_1)
A small number of addresses (<1%) selected from the PAF contain more than one household. At these addresses the interviewers carry out a selection procedure to identify which households to include in the sample. The same procedure was used for both samples; interviewers selected up to three households at each address; if more than three households were found the interviewer selected three at random.

The household selection weight (w_1) corrects for this selection of households and prevents households in multi-occupied addresses from being under-represented in the issued sample. This weight is equal to the number of households identified divided by the number selected. This weight was trimmed at 2 to avoid any large values.

The final selection weights were the product of the household and address selection weights (w_{01}). The weights were scaled to match the achieved sample size. This weight (selectin) is used for the comparison of key estimates before corrections were made for differential non-response.

Calibration weight for participating households (w_2)
Calibration weighting was used to generate weights for the participating households using CALMAR. The same variables were used for both Boost and Core samples. The achieved household sample was calibrated so that the distributions for age/sex and region for the household members matched the adjusted ONS 2005 mid-year household population estimates[19] (Tables B1 and B2). The region variable used was an inner/outer London split based on PCT. The calibration weight generated for a particular household depended upon the age/sex profiles of all household members and the region within which it was situated. Since this is a household-level weight, information on all household members was used, including children. The household and address selection weights were used as initial values when generating the calibration weights (w_2).

[19] These were the most recent estimates available when the HSE was being weighted. For the sake of comparison the same estimates were subsequently used for the Boost.

Apx. Table B1 ONS mid-2005 household population estimates for London by age and sex

Age (grouped)	Men	Women
0-4	261,377	251,084
5-10	260,458	252,762
11-15	214,191	203,149
16-24	430,969	430,607
25-34	734,464	722,543
35-44	650,592	616,758
45-54	430,424	444,336
55-64	314,280	342,595
65-74	210,890	239,429
75+	160,099	242,261
All London	3,667,744	3,745,524

Apx. Table B2 ONS mid-2005 household population estimates for inner/outer London PCTs

Region	Estimate
Inner London[1]	2,920,805
Outer London	4,492,463
All London	7,413,268

[1]Inner London PCT codes are: 5C3, 5C4, 5C5, 5C9, 5H1, 5K7, 5K8, 5LA, 5LC, 5LD, 5LE, 5LF and 5LG. Outer London PCT codes are: 5A4, 5A5, 5A7, 5A8, 5A9, 5AT, 5C1, 5C2, 5HX, 5HY, 5K5, 5K6, 5K9, 5M6, 5M7, 5NA, 5NC and TAK.

The aim of the calibration weighting was to reduce non-response bias resulting from differential non-response at the household level. The calibration weights generated (w_2) were re-scaled so that the sum of the weights equalled the number of participating households. This weight is the household weight for the Core sample (hhwt).

Adult interview weight (w_3)

There were no selection weights for adult respondents since all adults in responding households were selected for the Core sample. Non-response weights were generated to reduce possible non-response bias caused by individuals in responding households not completing individual interviews. Response was lower for the Boost sample; for the Boost 66% of adults in households with more than one adult completed an individual interview, compared with 86% for the Core.

A weighted (by hhwt) logistic regression model was fitted. The outcome was whether the interview was completed or not. The following variables were entered as covariates:

- Age group by sex,
- Household type, and
- Inner/outer London

The logistic regression model shows the relationship of these measures with response (see Table B3). The adult non-response weight (w_3) was calculated as the inverse of the predicted probabilities of response estimated from the regression model.

Respondents in single adult households were not included in the modelling and were given a non-response weight of 1.

The weights were trimmed at the 0.5% tails to remove extreme values. The interview weights for the core sample were calculated as:

intwt = w_2 x w_3

The weights were re-scaled so that the sum of the weights equalled the size of the achieved sample.

Apx. Table B3 Adult individual non-response model for Core HSE data

	B	S.E.	Wald	df	Sig.	Exp(B)
Age group			25.82	13	0.02	
Men 16-24					(baseline)	
Men 25-34	0.28	0.27	1.04	1	0.31	1.32
Men 35-44	0.28	0.29	0.88	1	0.35	1.32
Men 45-54	-0.06	0.30	0.04	1	0.85	0.94
Men 55-64	1.14	0.45	6.40	1	0.01	3.12
Men 65-74	0.94	0.55	2.99	1	0.08	2.57
Men 75+	1.72	0.87	3.96	1	0.05	5.60
Women 16-24	0.27	0.29	0.88	1	0.35	1.31
Women 25-34	0.36	0.28	1.68	1	0.20	1.44
Women 35-44	0.95	0.35	7.46	1	0.01	2.57
Women 45-54	1.05	0.34	9.51	1	0.00	2.86
Women 55-64	0.46	0.38	1.44	1	0.23	1.58
Women 65-74	0.31	0.44	0.49	1	0.48	1.36
Women 75+	0.99	0.67	2.17	1	0.14	2.70
Household type			44.31	4	0.00	
2 adults, both 16-59, no children					(baseline)	
Small family	0.39	0.25	2.56	1	0.11	1.48
Large family	0.38	0.32	1.40	1	0.24	1.47
Large adult household	-0.76	0.18	17.01	1	0.00	0.47
2 adults, 1 or both aged 60+, no children	0.17	0.35	0.24	1	0.62	1.19
Region			3.24	1	0.07	
Inner London					(baseline)	
Outer London	0.25	0.14	3.24	1	0.07	1.29
Constant	1.17	0.26	19.44	1	0.00	3.21

Notes:
1. The response is 1 = response, 0 = non-response.
2. The model R^2 = 0.062 (Cox and Snells).
3. B is the estimate coefficient with standard error S.E.
4. The Wald-test measures the impact of the categorical variable on the model with the appropriate number of degrees of freedom df. If the test is significant (sig < 0.05) then the categorical variable is considered to be 'significantly associated' with the response variable and therefore included in the model.
5. The Wald test for each level of the categorical variable is also shown. This tests the difference between that level and the baseline category.

Apx. Table B4 Adult individual non-response model for London Boost data

	B	S.E.	Wald	df	Sig.	Exp(B)
Age group			85.4	13	0.00	
Men 16-24					(baseline)	
Men 25-34	0.15	0.11	1.8	1	0.18	1.17
Men 35-44	0.31	0.12	6.8	1	0.01	1.37
Men 45-54	0.74	0.13	30.4	1	0.00	2.10
Men 55-64	0.36	0.16	5.2	1	0.02	1.43
Men 65-74	0.07	0.19	0.1	1	0.70	1.08
Men 75+	-0.36	0.25	2.1	1	0.15	0.70
Women 16-24	0.34	0.12	7.4	1	0.01	1.40
Women 25-34	0.27	0.12	5.3	1	0.02	1.30
Women 35-44	0.67	0.12	29.0	1	0.00	1.96
Women 45-54	0.67	0.13	26.5	1	0.00	1.95
Women 55-64	0.33	0.15	4.5	1	0.03	1.39
Women 65-74	-0.13	0.21	0.4	1	0.54	0.88
Women 75+	-0.35	0.26	1.9	1	0.17	0.70
Household type			30.9	4	0.00	
2 adults, both 16-59, no children					(baseline)	
Small family	0.64	0.08	57.6	1	0.00	1.90
Large family	0.30	0.11	7.7	1	0.01	1.35
Large adult household	-0.06	0.07	0.8	1	0.38	0.94
2 adults, 1 or both aged 60+, no children	0.19	0.17	21.2	1	0.00	1.21
Region			5.8	1	0.02	
Inner London					(baseline)	
Outer London	0.13	0.06	5.8	1	0.02	1.14
Constant	-0.13	0.11	1.6	1	0.21	0.87

Notes:
1. The response is 1 = response, 0 = non-response.
2. The model $R^2 = 0.084$ (Cox and Snells).
3. B is the estimate coefficient with standard error S.E.
4. The Wald-test measures the impact of the categorical variable on the model with the appropriate number of degrees of freedom df. If the test is significant (sig < 0.05) then the categorical variable is considered to be 'significantly associated' with the response variable and therefore included in the model.
5. The Wald test for each level of the categorical variable is also shown. This tests the difference between that level and the baseline category.

APPENDIX C PROPENSITY SCORE MATCHING

Propensity score matching (PSM) is a method that allows cases from a treatment sample (in this case the HSE Core sample in London) to be matched to cases from a control sample (the London Boost Sample). The matching controls for differences in sample profile; in this case the socio-demographic profile of the Core sample is adjusted to make it match that of the Boost. Matching the samples means any differences in survey estimates should be attributable to measurement error and not sample composition.

PSM is based on the following steps:

1. A propensity model is fitted using logistic regression, the binary outcome variable is whether the case belongs to the treatment or control sample;
2. A predicted score is generated by the model for each case, this is the 'propensity score';
3. The propensity scores are then used to match the treatment and control samples[20]. The matched samples are then analysed together.

The first step was to model the differences between the two sample profiles. The probability, or propensity, of the respondent being in either the Core or Boost sample was estimated using a logistic regression model. Sample type was used as the dependent variable and a number of socio-demographic characteristics were used as predictors. The predicted probabilities were saved as propensity scores. These scores measure the propensity of a respondent to be in either the Core or Boost sample, depending on their socio-demographic characteristics. The propensity scores were used to match the samples, since respondents with similar scores should be similar in terms of the characteristics in the model. A full list of the demographic and household characteristics used in the regression model is given in Table C1.

Apx. Table C1 Variables included in the logistic regression

Age group
Sex
Ethnicity
Marital status
NS-Sec of HRP
Current economic activity of HRP
Number of adults in the household (16+)
Number of children in the household (0-15)
Household type
Lone parent household
Index of Multiple Deprivation score 2006 (Super Output Area level)
PCT spearhead status
% of the population from a non-white background (Postcode sector level measure using data from 2001 Census)
% of the population from a non-manual occupation (Postcode sector level measure using data from 2001 Census)

[20] The propensity score matching was carried out in Stata V9 using the psmatch2 command

There were some variables that were unsuitable for using in the model. These were household income and the respondent's current economic activity and education.

These variables were excluded because there were fundamental differences in the wording or format of the Core and Boost questions, which meant that it was unclear whether the response categories were measuring the same thing. These variables also contained large numbers of missing cases for Boost respondents

Men and women were matched separately. For women the variables that varied significantly by sample type were NS-SEC of HRP, respondent ethnicity and IMD score of the local area. For men the significant variables were household type, NS-SEC of HRP, respondent ethnicity and IMD score of the local area.

Women in the Core sample were more likely to have a HRP in an intermediate NS-SEC category than Boost respondents, who were more likely to have a HRP at either end of the NS-SEC scale (either higher managers or had never worked). Women in the Boost were more likely to be from a White ethnic background and were more likely to live in more deprived areas, since being in the Boost sample was associated with a higher IMD score.

Similar patterns were seen for men. Core respondents were more likely to have a HRP in a middle NS-SEC category. Boost respondents were more likely to have a HRP who was either in management or had never worked. Men in the Boost respondents were also more likely to be from a White ethnic background and live in more deprived areas. The full models are given in Tables C2 and C3, below.

Apx. Table C2 Propensity model for women

	B	S.E.	Wald	df	Sig.	Exp(B)
NS-SEC of HRP (8 variable classification)			16.8	7	0.019	
Higher managerial and professional occupations			(baseline)			1.00
Lower managerial and professional occupations	-0.22	0.13	3.0	1	0.082	0.80
Intermediate occupations	-0.09	0.16	0.3	1	0.596	0.92
Small employers and own account workers	-0.23	0.17	1.8	1	0.180	0.79
Lower supervisory and technical occupations	-0.33	0.19	3.1	1	0.080	0.72
Semi-routine occupations	-0.46	0.15	9.3	1	0.002	0.63
Routine occupations	-0.38	0.17	5.0	1	0.025	0.68
Never worked and long term unemployed	0.11	0.19	0.3	1	0.577	1.11
Ethnicity of respondents			24.7	4	0.000	
White			(baseline)			1.00
Mixed	-0.05	0.25	0.0	1	0.830	0.95
Asian	-0.39	0.12	11.3	1	0.001	0.68
Black	-0.55	0.13	17.8	1	0.000	0.58
Chinese/other	-0.19	0.21	0.8	1	0.365	0.82
Age group			2.0	5	0.851	
16-24			(baseline)			1.00
25-34	-0.10	0.14	0.5	1	0.489	0.91
35-44	-0.16	0.14	1.3	1	0.262	0.85
45-54	-0.05	0.15	0.1	1	0.725	0.95
55-64	-0.01	0.17	0.0	1	0.958	0.99
65+	-0.11	0.15	0.5	1	0.463	0.89
IMD score	0.02	0.00	24.6	1	0.000	1.02
Constant	1.22	0.17	53.6	1	0.000	3.39

Notes: 1. The response is 1 = Boost, 0 = Core.

2. The model $R^2 = 0.019$ (Cox and Snells).

3. B is the estimate coefficient with standard error S.E.

4. The Wald-test measures the impact of the categorical variable on the model with the appropriate number of degrees of freedom df. If the test is significant (sig < 0.05) then the categorical variable is considered to be 'significantly associated' with the response variable and therefore included in the model.

5. The Wald test for each level of the categorical variable is also shown. This tests the difference between that level and the baseline category.

Apx. Table C3 Propensity model for men

	B	S.E.	Wald	df	Sig.	Exp(B)
Household type			18.0	6	0.006	
1 adult aged 16-59, no children			(baseline)			1.00
2 adults, both 16-59, no children	0.19	0.16	1.4	1	0.235	1.21
Small family	0.40	0.17	5.6	1	0.018	1.49
Large family	0.21	0.20	1.1	1	0.301	1.23
Large adult household	0.58	0.16	13.2	1	0.000	1.78
2 adults, 1 or both aged 60+, no children	0.34	0.22	2.4	1	0.118	1.41
1 adult, aged 60+, no children	0.12	0.26	0.2	1	0.657	1.12
NS-SEC of HRP (8 variable classification)			17.4	7	0.015	
Higher managerial and professional occupations			(baseline)			
Lower managerial and professional occupations	0.00	0.13	0.0	1	0.993	1.00
Intermediate occupations	-0.10	0.18	0.3	1	0.578	0.90
Small employers and own account workers	-0.10	0.17	0.4	1	0.552	0.90
Lower supervisory and technical occupations	-0.38	0.17	4.6	1	0.031	0.69
Semi-routine occupations	-0.34	0.17	4.2	1	0.041	0.71
Routine occupations	-0.37	0.17	4.8	1	0.028	0.69
Never worked and long term unemployed	0.35	0.24	2.1	1	0.147	1.41
Ethnicity of respondents			20.6	4	0.000	
White			(baseline)			1.00
Mixed	0.37	0.33	1.3	1	0.263	1.44
Asian	-0.30	0.13	5.6	1	0.018	0.74
Black	-0.56	0.15	14.8	1	0.000	0.57
Chinese/other	0.12	0.25	0.2	1	0.647	1.12
Age group			3.9	5	0.569	
16-24			(baseline)			1.00
25-34	0.04	0.16	0.1	1	0.803	1.04
35-44	0.19	0.17	1.3	1	0.255	1.21
45-54	0.26	0.17	2.4	1	0.123	1.30
55-64	0.15	0.19	0.7	1	0.417	1.17
65+	0.18	0.22	0.6	1	0.423	1.20
IMD score	0.01	0.00	10.4	1	0.001	1.01
Constant	0.57	0.22	6.6	1	0.010	1.78

Notes:
1. The response is 1 = Boost, 0 = Core.
2. The model $R^2 = 0.017$ (Cox and Snells).
3. B is the estimate coefficient with standard error S.E.
4. The Wald-test measures the impact of the categorical variable on the model with the appropriate number of degrees of freedom df. If the test is significant (sig < 0.05) then the categorical variable is considered to be 'significantly associated' with the response variable and therefore included in the model.
5. The Wald test for each level of the categorical variable is also shown. This tests the difference between that level and the baseline category.

The two samples are fairly similar in size; this means one-to-one matching would have been unsuitable. Propensity score matching usually requires a large group of controls to find suitable one-on-one matches for a smaller treatment group. It would have been difficult to carry out one-to-one matching (without replacement) between the Boost and Core samples, and the kernel method of matching was used instead

The kernel function matches each respondent in the Boost to *all* the respondents in the Core. The respondents in the Boost that are matched are weighted according to their proximity to each Core respondent using their propensity score. The kernel function weights the contribution of each respondent in the Core sample, with higher weights for respondents who are a better match. Exact matches get a large weight; poorer matches contribute less and get a smaller weight. All members of the Core are used but the poorer matches have such small weights they have little effect.

The output of the matching process is a weight which, when applied to the Core data, makes the Core match the Boost sample in terms of the socio-demographic variables in the model. The respondents in the Boost sample are each given a weight of 1. The weights for the Core sample vary and ensure that the profiles of the two groups match. The Core sample was adjusted to match the unweighted Boost sample, so neither sample is representative of the population of London and cannot be used to make generalisations. The samples should only be compared to each other, and not to the London population. When running analyses on matched combined data we can attribute any differences in the outcome variables to the survey mode and not to differences in sample composition.

APPENDIX D TABLES

Apx. Table D1 Individual level socio-demographic characteristics by sample type and sex

	WEIGHTED BY NON-RESPONSE WEIGHT				WEIGHTED BY SELECTION WEIGHT ONLY				MATCHED DATA (UNWEIGHTED)			
	WOMEN		MEN		WOMEN		MEN		WOMEN		MEN	
	Core	Boost	Core	Boost	Core	Boost	Core	Boost	Core	Boost	Core	Boost
	%	%	%	%	%	%	%	%	%	%	%	%
Age group												
16-24	14.3	14.6	14.8	15.3	12.2	13.0	12.2	12.2	11.6	13.4	11.5	12.1
25-34	23.2	23.0	24.4	24.1	22.0	21.4	22.6	20.1	22.0	21.8	21.9	21.3
35-44	20.5	21.1	22.3	22.6	23.6	21.6	21.9	21.4	21.7	20.4	21.1	20.6
45-54	14.8	15.2	14.9	15.0	15.3	15.9	15.4	17.9	15.2	15.8	15.7	17.3
55-64	11.4	11.5	10.8	10.7	11.2	12.3	13.2	13.3	12.5	12.5	13.9	13.3
65+	15.8	14.6	12.8	12.3	15.8	15.8	14.8	15.0	16.9	16.0	15.9	15.4
Base (unweighted)	*833*	*2726*	*736*	*2202*	*833*	*2726*	*736*	*2202*	*833*	*2726*	*736*	*2202*
Base (weighted)	*811*	*2508*	*783*	*2452*	*870*	*2871*	*768*	*2302*	*2736*	*2726*	*2206*	*2202*
P-value		*0.978*		*0.998*		*0.809*		*0.637*		*0.774*		*0.937*
Ethnicity												
White	64.8	69.1	65.8	68.3	62.8	69.6	64.5	68.7	70.6	70.6	70.9	70.0
Mixed	2.5	2.8	1.5	2.1	2.7	2.9	1.6	2.2	2.9	3.0	1.7	2.5
Asian	16.2	13.7	16.1	15.9	17.3	13.6	18.7	15.7	13.1	12.5	15.5	14.7
Black	12.5	10.3	13.8	9.4	13.6	10.2	12.5	9.4	9.6	10.0	9.0	8.8
Chinese/other	4.0	4.0	2.9	4.3	3.6	3.8	2.7	4.0	3.9	3.9	2.9	3.9
Base (unweighted)	*831*	*2672*	*732*	*2145*	*831*	*2672*	*732*	*2145*	*831*	*2672*	*732*	*2145*
Base (weighted)	*809*	*2454*	*779*	*2384*	*868*	*2808*	*764*	*2244*	*2729*	*2672*	*2194*	*2145*
P-value		*0.373*		*0.150*		*0.046*		*0.098*		*0.991*		*0.673*
Highest education qualification												
Higher degree/ Degree/NVQ4/5	35.9	36.0	36.1	41.9	34.2	33.6	34.6	39.8	36.2	35.6	36.4	40.8
NVQ3/GCE A Level equiv - any grade	10.8	15.4	12.5	15.9	10.1	15.4	11.1	14.6	10.2	14.8	11.4	14.9
NVQ1/2 GCE O Level equiv - any grade	25.1	28.8	25.7	24.9	26.1	29.5	26.1	26.0	24.8	28.4	24.8	24.4
Other (eg City and Guilds, RSA/OCR, BTEC	2.4	3.0	3.7	3.9	2.8	3.0	4.3	4.6	2.8	2.9	4.3	4.3
No qualification	25.7	16.8	22.1	13.5	26.9	18.5	23.9	15.1	26.0	18.2	23.0	15.6
Base (unweighted)	*832*	*2479*	*732*	*2008*	*832*	*2479*	*732*	*2008*	*832*	*2479*	*732*	*2008*
Base (weighted)	*810*	*2294*	*779*	*2246*	*869*	*2603*	*764*	*2095*	*2733*	*2479*	*2194*	*2008*
P-value		***0.000***		***0.000***		***0.000***		***0.000***		***0.000***		***0.002***
Respondent economic activity												
Full time study	7.8	8.8	8.6	11.2	6.7	7.9	7.9	9.3	6.4	8.5	7.0	8.9
In paid work	50.8	54.4	65.6	67.1	48.5	51.6	64.2	65.4	49.1	51.8	64.7	65.1
Looking for work	2.8	2.5	5.4	3.6	2.7	2.9	4.8	3.6	2.6	2.9	4.3	3.5
Sick (temp or long term)	2.5	3.0	4.9	3.7	2.8	3.3	4.9	4.4	2.9	3.5	4.6	4.7
Retired	17.6	14.2	14.7	12.7	17.6	15.8	17.1	15.4	19.0	15.8	18.4	15.8
Looking after home	18.6	17.0	0.8	1.8	21.8	18.5	1.1	2.0	19.9	17.6	1.0	1.9
Base (unweighted)	*821*	*2479*	*727*	*2005*	*821*	*2479*	*727*	*2005*	*821*	*2479*	*727*	*2005*
Base (weighted)	*805*	*2292*	*772*	*2223*	*863*	*2610*	*759*	*2098*	*2711*	*2479*	*2181*	*2005*
P-value		*0.209*		*0.057*		*0.233*		*0.359*		*0.094*		*0.199*

Table D1 (continued)

	WEIGHTED BY NON-RESPONSE WEIGHT				WEIGHTED BY SELECTION WEIGHT ONLY				MATCHED DATA (UNWEIGHTED)			
	WOMEN		MEN		WOMEN		MEN		WOMEN		MEN	
	Core	Boost	Core	Boost	Core	Boost	Core	Boost	Core	Boost	Core	Boost
	%	%	%	%	%	%	%	%	%	%	%	%
Marital status												
Single, never married	35.7	34.0	45.6	41.9	32.5	32.7	39.5	34.8	33.3	34.9	39.1	36.7
Married	44.0	47.3	44.8	48.8	46.0	44.9	49.9	54.6	44.9	43.0	50.6	52.3
Separated	3.1	3.7	2.5	2.4	3.4	4.7	2.3	2.4	3.1	4.5	2.2	2.5
Divorced	8.5	7.1	5.0	5.0	9.3	8.4	5.7	5.7	9.5	8.3	5.4	5.9
Widowed	8.7	7.9	2.0	1.8	8.8	9.4	2.6	2.5	9.3	9.3	2.7	2.5
Base (unweighted)	*831*	*2679*	*736*	*2157*	*831*	*2679*	*736*	*2157*	*831*	*2679*	*736*	*2157*
Base (weighted)	*810*	*2465*	*783*	*2395*	*868*	*2821*	*768*	*2253*	*2729*	*2679*	*2206*	*2157*
P-value		*0.302*		*0.571*		*0.506*		*0.269*		*0.273*		*0.786*

Apx. Table D2 Household level socio-demographic characteristics by sample type

	WEIGHTED BY NON-RESPONSE WEIGHT		WEIGHTED BY SELECTION WEIGHT ONLY	
	Core %	Boost %	Core %	Boost %
Household type				
1 adult aged 16-59, no children	18.0	15.0	17.0	13.9
2 adults, both 16-59, no children	18.7	17.2	17.0	16.1
Small family	18.2	19.4	20.0	20.6
Large family	5.3	5.7	6.4	6.8
Large adult household	16.1	18.6	15.1	18.3
2 adults, 1 or both aged 60+, no child	10.5	11.3	11.3	11.5
1 adult, aged 60+, no children	13.2	12.7	13.3	12.8
Base (unweighted)	*963*	*3873*	*963*	*3873*
Base (weighted)	*967*	*3873*	*966*	*3882*
P-value		*0.277*		*0.187*
NS_SEC of HRP (8 groups)				
Managerial and professional occupations	45.1	47.2	44.2	46.3
Intermediate occupations	10.6	11.1	10.6	11.4
Small employers & own account workers	8.9	9.6	9.1	9.5
Lower supervisory and technical occupations	7.8	7.4	8.0	7.3
Semi-routine and routine occupations	23.0	19.8	23.4	20.4
Never worked and long term unemployed	4.6	4.8	4.7	5.1
Base (unweighted)	*959*	*2868*	*959*	*2868*
Base (weighted)	*957*	*2894*	*961*	*2885*
P-value		*0.458*		*0.385*
Economic activity of HRP				
In paid employment or self-employment (or away temporarily)	63.2	63.4	62.1	62.6
Looking or waiting for paid work or training	2.7	2.7	2.9	2.8
In full time study or govt training	2.6	2.0	2.6	1.9
Sick - temp or long-term	3.6	4.8	3.7	4.6
Retired from paid work	20.7	21.9	21.3	22.1
Looking after the home or family	7.1	5.2	7.5	6.0
Base (unweighted)	*953*	*2855*	*953*	*2855*
Base (weighted)	*952*	*2883*	*955*	*2881*
P-value		*0.186*		*0.325*

Table D2 continued

	WEIGHTED BY NON-RESPONSE WEIGHT		WEIGHTED BY SELECTION WEIGHT ONLY	
	Core %	Boost %		Core %
IMD (London quintiles)				
<11.87-least deprived	18.5	20.1	19.9	21.0
11.88-19.24	20.3	20.6	20.6	20.9
19.25-27.65	21.3	19.5	20.9	19.4
27.67-37.92	21.8	18.4	21.1	18.3
>37.93-most deprived	18.1	21.3	17.5	20.4
Base (unweighted)	*963*	*3873*	*963*	*3873*
Base (weighted)	*967*	*3873*	*966*	*3882*
P-value		*0.591*		*0.711*
PCT status				
Non-spearhead	64.8	65.4	66.6	67.4
Spearhead	35.2	34.6	33.4	32.6
Base (unweighted)	*963*	*3873*	*963*	*3873*
Base (weighted)	*967*	*3873*	*966*	*3882*
P-value		*0.277*		*0.760*

Note: There is no matched comparison because Propensity Score Matching was only carried out at individual level. HRP NS-SEC and HRP economic activity were only coded for households where there was a complete individual interview.

Apx. Table D3 Health measures by sample type and sex

	WEIGHTED BY NON-RESPONSE WEIGHT				WEIGHTED BY SELECTION WEIGHT ONLY				MATCHED DATA (UNWEIGHTED)			
	WOMEN		MEN		WOMEN		MEN		WOMEN		MEN	
	Core	Boost	Core	Boost	Core	Boost	Core	Boost	Core	Boost	Core	Boost
	%	%	%	%	%	%	%	%	%	%	%	%
Self-assessed general health												
Very good	32.8	21.3	35.8	23.8	30.7	20.9	34.3	23.3	31.5	20.8	35.0	23.0
Good	45.4	51.1	43.6	48.3	46.6	50.6	43.8	47.4	45.9	50.6	43.0	47.4
Fair	14.0	23.3	14.0	23.8	14.8	23.8	14.6	24.4	14.3	23.4	14.7	24.6
Bad	6.1	3.6	5.5	3.1	6.2	3.9	5.9	3.8	6.4	4.3	5.9	3.9
Very bad	1.6	0.6	1.1	0.9	1.8	0.7	1.4	1.0	1.8	0.9	1.4	1.1
Base (unweighted)	*833*	*2701*	*736*	*2181*	*833*	*2701*	*736*	*2181*	*833*	*2701*	*736*	*2181*
Base (weighted)	*811*	*2489*	*783*	*2428*	*870*	*2846*	*768*	*2279*	*2736*	*2701*	*2206*	*2181*
P-value		***0.000***		***0.000***		***0.000***		***0.000***		***0.000***		***0.000***
(D) GHQ12 score -grouped												
score 0	58.3	51.6	62.2	56.9	57.9	51.8	64.3	58.0	57.7	51.3	63.5	56.9
score 1-3	26.5	28.7	25.1	27.5	27.0	28.3	23.4	26.3	26.8	28.2	24.2	26.8
score 4+	15.2	19.7	12.6	15.6	15.0	19.9	12.3	15.6	15.5	20.5	12.3	16.3
Base (unweighted)	*691*	*2628*	*620*	*2143*	*691*	*2628*	*620*	*2143*	*691*	*2628*	*620*	*2143*
Base (weighted)	*666*	*2426*	*665*	*2381*	*719*	*2771*	*646*	*2238*	*2286*	*2628*	*1866*	*2143*
P-value		***0.006***		*0.093*		***0.009***		*0.026*		***0.005***		*0.011*
Long-standing illness or disability												
Yes	32.7	31.3	30.3	28.4	33.9	31.8	33.4	29.9	34.5	32.3	33.9	31.0
No	67.3	68.7	69.7	71.6	66.1	68.2	66.6	70.1	65.5	67.7	66.1	69.0
Base (unweighted)	*2676*	*833*	*2174*	*736*	*2676*	*833*	*2174*	*736*	*2676*	*833*	*2174*	*736*
Base (weighted)	*2464*	*811*	*2421*	*783*	*2819*	*870*	*2271*	*768*	*2676*	*2736*	*2174*	*2206*
P-value		*0.533*		*0.383*		*0.306*		*0.136*		*0.294*		*0.224*
(D) Limiting long-standing illness or disability												
Limiting LI	21.4	18.6	18.8	16.9	22.5	19.5	21.5	17.9	23.0	20.0	21.9	18.6
Non limiting LI	11.3	12.7	11.5	11.5	11.4	12.3	12.0	12.0	11.4	12.3	12.0	12.4
No LI	67.3	68.7	69.7	71.6	66.1	68.2	66.6	70.1	65.5	67.7	66.1	69.0
Base (unweighted)	*2676*	*833*	*2174*	*736*	*2676*	*833*	*2174*	*736*	*2676*	*833*	*2174*	*736*
Base (weighted)	*2464*	*811*	*2421*	*783*	*2819*	*870*	*2271*	*768*	*2676*	*2736*	*2174*	*2206*
P-value		*0.231*		*0.532*		*0.208*		*0.144*		*0.215*		*0.193*

Apx. Table D4 Health-related measures by sample type and sex

	WEIGHTED BY NON-RESPONSE WEIGHT				WEIGHTED BY SELECTION WEIGHT ONLY				MATCHED DATA (UNWEIGHTED)			
	WOMEN		MEN		WOMEN		MEN		WOMEN		MEN	
	Core	Boost	Core	Boost	Core	Boost	Core	Boost	Core	Boost	Core	Boost
	%	%	%	%	%	%	%	%	%	%	%	%
(D) Cigarette smoking status												
Never smoked cigarettes at all	58.2	59.5	47.4	49.9	59.0	58.8	48.0	48.9	55.6	57.6	46.6	47.7
Used to smoke cigarettes occasionally	7.5	8.5	6.7	7.6	6.4	8.1	6.1	7.5	7.0	8.3	6.6	7.4
Used to smoke cigarettes regularly	15.3	13.2	21.1	17.2	15.5	13.8	21.2	19.2	17.4	14.6	23.1	19.6
Current cigarette smoker	19.0	18.8	24.9	25.4	19.0	19.3	24.8	24.3	20.0	19.5	23.7	25.4
Base (unweighted)	*827*	*2680*	*728*	*2142*	*827*	*2680*	*728*	*2142*	*827*	*2680*	*728*	*2142*
Base (weighted)	*805*	*2468*	*774*	*2395*	*862*	*2821*	*760*	*2243*	*2716*	*2680*	*2181*	*2142*
P-value		*0.568*		*0.284*		*0.410*		*0.504*		*0.282*		*0.331*
(D) Number of cigarettes smoked (including unknown)												
Current smoker, under 10 a day	9.7	6.8	10.3	9.8	9.5	7.1	9.6	8.9	9.9	7.1	9.4	9.1
Current smoker, 10 to under 20 a day	6.1	6.3	7.6	8.0	6.0	6.4	8.2	7.8	6.3	6.2	7.6	8.2
Current smoker, 20 or more a day	3.0	3.6	7.1	5.5	3.1	3.8	7.2	5.3	3.5	3.8	7.0	5.5
Not a current smoker	81.0	81.1	75.0	73.8	81.0	80.6	75.1	74.9	80.0	80.4	76.1	73.8
Current smoker, amount unknown	0.3	2.2	0.0	3.2	0.4	2.2	0.0	3.1	0.4	2.5	0.0	3.4
Base (unweighted)	*827*	*2680*	*728*	*2142*	*827*	*2680*	*728*	*2142*	*827*	*2680*	*728*	*2142*
Base (weighted)	*805*	*2468*	*774*	*2395*	*862*	*2821*	*760*	*2243*	*2716*	*2680*	*2181*	*2142*
P-value		***0.001***		*0.682*		*0.170*		*0.449*		*0.173*		*0.619*
(D) Number of cigarettes smoked (excluding unknown)												
Current smoker, under 10 a day	9.7	6.9	10.3	9.9	9.5	7.2	9.6	9.1	9.9	7.3	9.4	9.4
Current smoker, 10 to under 20 a day	6.1	6.5	7.6	8.2	6.1	6.5	8.2	8.1	6.3	6.3	7.6	8.5
Current smoker, 20 or more a day	3.0	3.7	7.1	5.7	3.1	3.9	7.2	5.5	3.5	3.9	7.0	5.7
Not a current smoker	81.2	83.0	75.0	76.2	81.3	82.4	75.1	77.3	80.3	82.5	76.1	76.4
Base (unweighted)	*824*	*2639*	*728*	*2110*	*824*	*2639*	*728*	*2110*	*824*	*2639*	*728*	*2110*
Base (weighted)	*802*	*2434*	*773*	*2355*	*859*	*2785*	*760*	*2211*	*2705*	*2639*	*2181*	*2110*
P-value		*0.096*		*0.682*		*0.170*		*0.449*		*0.173*		*0.619*
Smoker in household?												
Yes	23.7	23.8	23.7	24.9	23.0	23.4	23.3	23.6	23.2	23.3	22.6	23.7
No	76.3	76.2	76.3	75.1	77.0	76.6	76.7	76.4	76.8	76.7	77.4	76.3
Base (unweighted)	*832*	*2687*	*736*	*2168*	*832*	*2687*	*736*	*2168*	*832*	*2687*	*736*	*2168*
Base (weighted)	*810*	*2475*	*783*	*2418*	*869*	*2830*	*768*	*2270*	*2732*	*2687*	*2206*	*2168*
P-value		*0.959*		*0.691*		*0.858*		*0.917*		*0.946*		*0.639*

	WEIGHTED BY NON-RESPONSE WEIGHT				WEIGHTED BY SELECTION WEIGHT ONLY				MATCHED DATA (UNWEIGHTED)			
	WOMEN		MEN		WOMEN		MEN		WOMEN		MEN	
	Core	Boost	Core	Boost	Core	Boost	Core	Boost	Core	Boost	Core	Boost
	%	%	%	%	%	%	%	%	%	%	%	%
Number of smokers in household												
None	76.3	77.0	76.3	75.8	77.0	77.4	76.7	77.0	76.8	77.5	77.4	76.9
1	16.5	15.2	15.7	15.6	17.0	15.7	16.9	15.5	16.8	15.7	15.5	15.6
2	7.2	7.8	8.0	8.6	6.0	6.9	6.4	7.5	6.4	6.8	7.1	7.5
Base (unweighted)	*833*	*2736*	*736*	*2206*	*833*	*2736*	*736*	*2206*	*833*	*2736*	*736*	*2206*
Base (weighted)	*811*	*2517*	*783*	*2456*	*870*	*2883*	*768*	*2306*	*2736*	*2736*	*2206*	*2206*
P-value		*0.746*		*0.931*		*0.631*		*0.615*		*0.799*		*0.943*
(D) Frequency drank alcohol in last 12 months: including non-drinkers												
Almost every day	8.8	6.1	14.7	13.9	8.6	6.4	14.7	14.1	9.5	6.8	15.5	14.9
Five or six days a week	3.9	4.1	7.5	7.4	3.7	4.2	7.0	7.6	4.3	4.1	7.7	7.6
Three or four days a week	11.3	13.5	15.7	15.9	9.9	12.6	14.3	15.4	11.1	13.1	14.9	15.7
Once or twice a week	20.5	22.0	26.9	25.5	20.2	21.0	26.3	24.5	21.9	21.6	26.6	24.0
Once or twice a month	12.1	11.7	10.0	10.8	11.8	11.8	9.9	10.5	12.0	11.8	10.2	10.6
Once every couple of months	8.8	9.1	3.4	6.2	8.8	9.6	3.6	6.8	8.6	9.5	3.6	6.3
Once or twice a year	8.4	10.0	5.4	4.6	8.7	10.3	6.0	5.2	8.5	10.0	5.9	5.0
Not at all in the last 12 months/Non-drinker	26.3	23.5	16.4	15.7	28.3	24.1	18.2	15.9	24.2	23.2	15.7	15.8
Base (unweighted)	*731*	*2297*	*644*	*1918*	*731*	*2297*	*644*	*1918*	*731*	*2297*	*644*	*1918*
Base (weighted)	*694*	*2068*	*663*	*2052*	*762*	*2430*	*670*	*1999*	*2410*	*2297*	*1942*	*1918*
P-value		*0.193*		*0.429*		*0.139*		*0.207*		*0.360*		*0.363*
(D) Whether drank in the last 7 days (excluding those who did not drink in the last 12 months)												
Yes	68.5	66.0	80.0	77.7	66.9	64.5	79.1	77.0	69.3	65.8	79.5	77.9
No	31.5	34.0	20.0	22.3	33.1	35.5	20.9	23.0	30.7	34.2	20.5	22.1
Base (unweighted)	*535*	*1814*	*539*	*1635*	*535*	*1814*	*539*	*1635*	*535*	*1814*	*539*	*1635*
Base (weighted)	*512*	*1642*	*555*	*1753*	*547*	*1903*	*549*	*1704*	*1831*	*1814*	*1640*	*1635*
P-value		*0.334*		*0.287*		*0.328*		*0.293*		*0.118*		*0.417*
(D) Binge drinking in the last 7 days (excl those who did not drink in last 7 days)												
No binge drinking in last week	66.8	56.5	65.8	49.9	67.6	58.4	68.5	53.2	67.3	58.0	68.7	52.3
Been binge drinking in last week	33.2	43.5	34.2	50.1	32.4	41.6	31.5	46.8	32.7	42.0	31.3	47.7
Base (unweighted)	*358*	*1114*	*426*	*1217*	*358*	*1114*	*426*	*1217*	*358*	*1114*	*426*	*1217*
Base (weighted)	*349*	*1012*	*443*	*1303*	*364*	*1144*	*433*	*1246*	*1262*	*1114*	*1302*	*1217*
P-value		***0.003***		***0.000***		***0.005***		***0.000***		***0.005***		***0.000***
(D) Binge drinking in the last 7 days (incl all non-drinkers)												
No binge drinking in last week	33.6	28.4	44.0	32.6	32.3	28.5	44.3	34.3	35.3	29.1	46.1	34.2
Been binge drinking in last week	16.7	21.8	22.9	32.7	15.5	20.3	20.4	30.1	17.1	21.1	21.0	31.2
All who did not drink in last 12mths/7 days	49.6	49.7	33.2	34.7	52.1	51.2	35.4	35.6	47.6	49.8	33.0	34.6
Base (unweighted)	*730*	*2217*	*644*	*1860*	*730*	*2217*	*644*	*1860*	*730*	*2217*	*644*	*1860*

	WEIGHTED BY NON-RESPONSE WEIGHT				WEIGHTED BY SELECTION WEIGHT ONLY				MATCHED DATA (UNWEIGHTED)			
	WOMEN		MEN		WOMEN		MEN		WOMEN		MEN	
	Core	Boost	Core	Boost	Core	Boost	Core	Boost	Core	Boost	Core	Boost
	%	%	%	%	%	%	%	%	%	%	%	%
Base (weighted)	*692*	*2014*	*663*	*1995*	*761*	*2345*	*670*	*1935*	*2604*	*2217*	*1942*	*1860*
P-value		*0.018*		***0.000***		*0.025*		***0.000***		*0.014*		***0.000***
(D) Grouped portions of fruit and veg (inc. juice) yesterday												
<1	4.7	3.9	7.6	5.6	4.5	4.1	7.8	5.2	4.5	4.1	7.2	5.3
1-2	26.0	14.7	26.4	16.7	27.1	14.9	27.5	16.8	26.6	14.9	26.9	16.9
3-4	27.3	26.9	28.1	23.7	27.1	27.0	27.5	23.8	27.4	26.5	27.8	24.4
5+	42.0	54.5	37.9	54.1	41.2	54.0	37.2	54.1	41.5	54.6	38.2	53.4
Base (unweighted)	*833*	*2647*	*735*	*2133*	*833*	*2647*	*735*	*2133*	*833*	*2647*	*735*	*2133*
Base (weighted)	*811*	*2436*	*782*	*2375*	*870*	*2789*	*767*	*2233*	*2736*	*2647*	*2202*	*2133*
P-value		***0.000***		***0.000***		***0.000***		***0.000***		***0.000***		***0.000***
(D) Summary physical activity level												
Group 1 - high	33.5	17.8	44.3	22.0	*32.6*	*18.0*	44.5	21.3	34.0	18.0	45.0	21.3
Group 2 - medium	34.1	39.3	27.5	34.6	*34.7*	*39.6*	26.6	34.6	34.6	39.3	27.0	34.3
Group 3 -low	32.4	42.9	28.2	43.5	*32.7*	*42.4*	28.9	44.1	31.4	42.7	29.0	44.3
Base (unweighted)	*695*	*1957*	*620*	*1593*	*695*	*1957*	*620*	*1593*	*695*	*1957*	*620*	*1593*
Base (weighted)	*682*	*1837*	*681*	*1841*	*732*	*2060*	*652*	*1671*	*2269*	*1957*	*1848*	*1593*
P-value		***0.000***		***0.000***		***0.000***		***0.000***		***0.000***		***0.000***
(D) perceived social support score - grouped												
No lack	60.0	64.2	52.8	58.8	61.7	65.2	52.1	59.8	61.7	63.8	54.6	59.0
Some lack	25.8	20.2	23.8	22.0	25.0	20.0	25.7	21.5	25.0	20.2	23.6	21.3
Severe lack	14.2	15.6	23.3	19.1	13.4	14.8	22.1	18.7	13.4	16.0	21.8	19.7
Base (unweighted)	*701*	*2665*	*626*	*2133*	*701*	*2665*	*626*	*2133*	*701*	*2665*	*626*	*2133*
Base (weighted)	*677*	*2448*	*670*	*2374*	*729*	*2803*	*651*	*2227*	*2323*	*2665*	*1881*	*2133*
P-value		*0.045*		*0.022*		*0.015*		*0.044*		*0.036*		*0.223*

Apx. Table D5 Proportion of respondents reporting to have eaten different fruit and vegetable categories

	WEIGHTED FOR NON-RESPONSE				MATCHED DATA (UNWEIGHTED)			
	Women		Men		Women		Men	
	Core	Boost	Core	Boost	Core	Boost	Core	Boost
	%	%	%	%	%	%	%	%
Tablespoons of vegetables	66	72	59	66	67	72	61	67
Tablespoons of vegetable dishes	11	33	13	35	11	32	12	34
Tablespoons of pulses	35	31	34	33	35	30	35	33
Bowls of salad	44	54	38	48	44	54	38	48
Handfuls of very small fruit	13	47	9	41	13	47	9	41
Small fruit	16	35	10	29	18	35	10	29
Medium sized fruit	64	67	57	64	63	68	58	64
Half a large fruit	7	7	6	6	7	7	5	6
Slices of very large fruit	3	12	5	10	4	12	4	10
Other large fruit	5	16	6	15	5	16	6	16
Tablespoons of dried fruit	17	24	10	18	18	24	10	19
Small glass fruit juice	50	62	54	60	49	61	51	59
Base (unweighted)	*833*	*2736*	*736*	*2206*	*833*	*2736*	*736*	*2206*
Base (weighted)	*811*	*2517*	*783*	*2456*	*2736*	*2736*	*2206*	*2206*

Apx. Table D6 Mean portions of fruit and vegetables consumed by sample type: women

		WEIGHTED FOR NON-RESPONSE			MATCHED DATA (UNWEIGHTED)		
	Sample type	Mean number of portions	Weighted base	Unweighted base	Mean number of portions	Weighted base	Unweighted base
Tablespoons of vegetables	Core	4.0	*547*	*533*	4.0	*547*	*1823*
	Boost	4.3*	*1977*	*1810*	4.4*	*1977*	*1977*
Tablespoons of vegetable dishes	Core	4.1	*96*	*89*	4.1	*96*	*288*
	Boost	4.6*	*877*	*833*	4.6*	*877*	*877*
Tablespoons of pulses	Core	3.6	*297*	*285*	3.4	*297*	*963*
	Boost	1.0*	*828*	*781*	1.0*	*828*	*828*
Bowls of salad	Core	1.2	*362*	*357*	1.2	*362*	*1194*
	Boost	1.8*	*1482*	*1351*	1.8*	*1482*	*1482*
Handfuls of very small fruit	Core	1.8	*111*	*104*	1.9	*111*	*359*
	Boost	2.1*	*1290*	*1181*	2.1*	*1290*	*1290*
Small fruit	Core	2.3	*142*	*133*	2.3	*142*	*481*
	Boost	1.8*	*948*	*883*	1.9*	*948*	*948*
Medium sized fruit	Core	2.1	*529*	*516*	2.0	*529*	*1730*
	Boost	1.8	*1850*	*1696*	1.8	*1850*	*1850*
Half a large fruit	Core	1.4	*56*	*56*	1.5	*56*	*185*
	Boost	1.7	*183*	*164*	1.7	*183*	*183*
Slices of very large fruit	Core	3.3	*29*	*27*	3.4	*29*	*98*
	Boost	1.9*	*331*	*297*	1.9*	*331*	*331*
Other large fruit	Core	3.0	*40*	*39*	3.1	*40*	*145*
	Boost	2.7*	*444*	*408*	2.7*	*444*	*444*
Tablespoons of dried fruit	Core	2.0	*144*	*139*	1.9	*144*	*484*
	Boost	1.9	*660*	*593*	1.8	*660*	*660*
Small glass fruit juice	Core	2.1	*412*	*407*	2.1	*412*	*1335*
	Boost	2.2*	*1656*	*1552*	2.2*	*1656*	*1656*

* difference in means is significant ($p<0.05$)

Apx. Table D7 Mean portions of fruit and vegetables consumed by sample type: men

		WEIGHTED FOR NON-RESPONSE			MATCHED DATA (UNWEIGHTED)		
	Sample type	Mean number of portions	Weighted base	Unweighted base	Mean number of portions	Weighted base	Unweighted base
Tablespoons of vegetables	Core	4.4	*435*	*459*	4.3	*435*	*1342*
	Boost	5.0*	*1477*	*1619*	4.9*	*1477*	*1477*
Tablespoons of vegetable dishes	Core	4.5	*94*	*98*	4.5	*94*	*273*
	Boost	5.6*	*743*	*853*	5.5*	*743*	*743*
Tablespoons of pulses	Core	4.2	*257*	*266*	4.2	*257*	*776*
	Boost	1.0*	*722*	*808*	1.0*	*722*	*722*
Bowls of salad	Core	1.2	*274*	*298*	1.2	*274*	*840*
	Boost	1.7*	*1059*	*1172*	1.7*	*1059*	*1059*
Handfuls of very small fruit	Core	1.8	*68*	*73*	1.8	*68*	*209*
	Boost	2.3*	*904*	*1010*	2.3*	*904*	*904*
Small fruit	Core	2.3	*72*	*75*	2.4	*72*	*227*
	Boost	2.2*	*641*	*706*	2.2*	*641*	*641*
Medium sized fruit	Core	2.1	*418*	*447*	2.1	*418*	*1281*
	Boost	2.0*	*1405*	*1561*	2.0*	*1405*	*1405*
Half a large fruit	Core	1.3	*41*	*46*	1.3	*41*	*115*
	Boost	1.8*	*141*	*149*	1.8*	*141*	*141*
Slices of very large fruit	Core	3.7	*30*	*37*	3.6	*30*	*95*
	Boost	2.2*	*218*	*244*	2.2*	*218*	*218*
Other large fruit	Core	3.7	*45*	*48*	3.6	*45*	*140*
	Boost	2.9	*344*	*369*	2.8	*344*	*344*
Tablespoons of dried fruit	Core	1.9	*74*	*78*	1.9	*74*	*227*
	Boost	2.1*	*418*	*449*	2.1*	*418*	*418*
Small glass fruit juice	Core	2.2	*379*	*424*	2.1	*379*	*1130*
	Boost	2.5*	*1295*	*1469*	2.4*	*1295*	*1295*

* difference in means is significant (p<0.05)

Apx. Table D8 Participation in sports by sample type and sex

	Women			Men		
	Core %	Boost %	sig	Core %	Boost %	sig
Weighted by non-response weight						
Aerobics/ keep fit/ gymnastics/ dance for fitness	10.2	12.3		2.7	5.1	*
Any other type of dancing	9.1	13.9	*	3.2	7.2	*
Cycling	5.0	7.7	*	11.3	15.7	*
Exercises (e.g. press ups, sit ups)	12.5	17.3	*	14.0	20.0	*
Football/ rugby	0.7	1.1		14	11.1	
Running/ jogging	7.5	11.5	*	12.1	19.4	*
Squash	0.4	0.4		1.3	1.6	
Swimming	12.1	13.7		10.6	11.1	
Tennis/ badminton	2.4	3.5		4.2	4.9	
Weight training/ gym/ exercise bike	12.5	14.3		16.6	21.8	*
Base(unweighted)	*833*	*2735*		*736*	*2206*	
Base(weighted)	*811*	*2517*		*783*	*2456*	
Matched data (unweighted)						
Aerobics/ keep fit/ gymnastics/ dance for fitness	10.7	12.6		3.1	4.7	
Any other type of dancing	8.8	13.2	*	3.1	6.4	*
Cycling	5.7	7.4		11.6	15.7	*
Exercises (e.g. press ups, sit ups)	12.1	16.9	*	12.0	18.7	*
Football/ rugby	0.9	1.0		11.3	9.6	
Running/ jogging	7.1	11.2	*	10.4	17.2	*
Squash	0.5	0.4		1.4	1.4	
Swimming	13.2	13.6		10.9	10.5	
Tennis/ badminton	2.6	3.3		3.2	4.6	
Weight training/ gym/ exercise bike	12.2	14.3		15.1	20.2	*
Base(unweighted)	*833*	*2735*		*736*	*2206*	
Base(weighted)	*2736*	*2736*		*2206*	*2206*	

* difference in proportions is significant (p<0.05)

Apx. Table D9 Body Mass Index by sample type and sex

	WEIGHTED BY NON-RESPONSE WEIGHT				MATCHED DATA (UNWEIGHTED)			
	WOMEN		MEN		WOMEN		MEN	
	Core	Boost	Core	Boost	Core	Boost	Core	Boost
	%	%	%	%	%	%	%	%
BMI based on all available measurements								
Less than 18.5	9.0	10.5	4.2	5.6	7.6	10.3	3.4	5.1
18.5 to less than 25	44.8	45.6	37.1	39.7	44.9	45.3	36.4	37.6
25 to less than 30	26.3	27.6	41.3	40.4	27.1	27.5	41.9	41.9
30 or more	19.9	16.4	17.4	14.2	20.4	16.9	18.2	15.4
Base (unweighted)	*661*	*2308*	*611*	*1913*	*661*	*2308*	*611*	*1913*
Base (weighted)	*641*	*2132*	*654*	*2136*	*2166*	*2308*	*1833*	*1913*
P-value for difference in distribution		*0.064*		*0.238*		0.064		0.238
Mean BMI	25.9	25.4	26.4	26.0	26.0	25.4	26.6	26.2
P-value for difference in means		***0.040***		*0.105*		***0.017***		*0.130*
BMI based on interviewer measurements								
Less than 18.5	9.0	8.2	4.2	5.9	7.6	8.2	3.4	5.1
18.5 to less than 25	44.8	42.7	37.1	37.3	44.9	42.2	36.4	35.2
25 to less than 30	26.3	29.9	41.3	41.7	27.1	29.8	41.9	43.6
30 or more	19.9	19.1	17.4	15.0	20.4	19.8	18.2	16.2
Base (unweighted)	*661*	*1529*	*611*	*1226*	*661*	*1529*	*611*	*1226*
Base (weighted)	*641*	*1388*	*654*	*1363*	*2166*	*1529*	*1833*	*1226*
P-value for difference in distribution		*0.418*		*0.389*		*0.535*		*0.361*
Mean BMI	25.9	26.0	26.4	26.1	26.0	26.1	26.6	26.4
P-value for difference in means		*0.613*		*0.271*		*0.790*		*0.436*