Appendix B. Weighting the NDNS core sample

Sarah Tipping

B.1 Introduction

NDNS requires weights to adjust for differences in sample selection and response. The weights adjust for differential selection probabilities of households and individuals, non-response to the individual questionnaire, non-response to the nurse visit and non-response to providing a blood sample. Non-response weights were generated using logistic regression modelling and calibration. Figures presented in this report are based on weighted data.[1]

B.2 Selection weights

Selection weights are required to correct for the unequal selection of:

1. dwelling units at multi dwelling unit[2] addresses,
2. catering units at multi catering unit[3] addresses,
3. individuals within dwelling or catering units, and
4. the sample across months.

Most addresses selected from the Postcode Address File (PAF) contain a single dwelling unit. However, a small number of addresses contain multiple dwelling units. At these addresses the interviewer selected one dwelling unit at random using a Kish grid.[4] The selected dwelling unit was then included in the sample. The dwelling unit selection weights (w_1) adjust for this selection. The weights are equivalent to the number of dwelling units identified at the address and were trimmed at three to avoid any large values. The dwelling unit selection weights ensure dwelling units at addresses containing more than one are not under-represented in the issued sample.

At each selected dwelling unit the interviewer enumerated the number of catering units and selected one at random using a Kish grid.[4] The catering unit selection weights (w_2) adjust for this selection of catering units. The catering unit selection weights ensure that catering units in multi-occupied dwelling units or addresses are not under-represented in the sample.

The selection of individuals within catering units depended on the selection ‘type’ of the address. Each sample point contained nine general sample addresses and 18 child boost addresses. At general sample addresses one adult (aged 19 years and over) and, where available, one child (aged 1.5 to 18 years) were selected at random from each responding catering unit by the interviewer. At child boost addresses one child was selected at random by the interviewer. Adults and children were weighted separately as they will always be analysed separately, hence this sample design feature has reduced costs without increasing the degree of clustering in the sample.

Individual selection weights (w_3) are required to ensure individuals in larger catering units are not under-represented in the sample. The individual selection weight is the inverse of the individual selection probabilities. For adults this is equivalent to the number of eligible adults in the catering unit, for children it is the number of eligible children in the catering unit. Pregnant or breastfeeding women were not eligible for the survey and were excluded from selection.

Prior to the launch of Year 1 fieldwork, NDNS used a ‘Run In’ sample to test field procedures. The Run In sample was selected alongside the Year 1 sample using the same methods and was subsequently incorporated into the Year 1 data. An additional set of weights were needed to correct for the unequal distribution of the sample across months in Year 1. The Run In sample weights (w_4) ensure the full sample (Years 1, 2 and 3 plus the Run In) is seasonally representative.

The combined selection weight (w_{sel}) is the product of the dwelling unit, catering unit and individual selection weights and the Run In weight.

B.3 Individual calibration weight

A set of household weights were not required as all analyses have been carried out at the individual level; any information collected about the household or catering unit has been reported in terms of the individual.

A set of individual weights was generated for the analysis of fully responding individuals (the 3,073 individuals who responded to the individual interview and completed three or four food diary days). These weights were generated using calibration methods. The aim was to reduce bias resulting from sampling error and differential non-response by age and sex and Government Office Region to the individual interview. An iterative procedure was used to adjust a starting weight until the distribution of the (weighted) sample matched that of the population for a set of key variables. The adjustment kept the values of the final weights as close as possible to those of the initial weights, which ensured the properties of the initial weights were retained in the final calibrated weights. The combined selection weights (w_{sel}), described in section B.2, were used as the initial weights.

The key variables used to create the individual weights were: age (grouped); by sex; and Government Office Region. The population figures used were taken from the mid-year population estimates.[5] As there are now three years worth of NDNS data, the average population of the three most recent years of population data (2008, 2009 and 2010) was used. The calibration was run separately for adults and children; the figures used for weighting adults are shown in Table B.1; those used for weighting children are shown in Table B.2. **(Tables B.1, B.2)**

The calibration weights generated were re-scaled so that the sum of the weights equalled the number of participating individuals. These are the final individual weights for the core sample (wti_adY123 and wti_chY123). Thus the final individual weights adjust for dwelling unit, catering unit and individual selection, the Year 1 Run In sample, and for the age/sex and regional profiles of participating individuals.

B.4 Nurse interview non-response weight

Participants who completed three or four food diary days (i.e. those deemed fully productive) were asked for their consent for a nurse to contact them to arrange a visit. Approximately three quarters of these participants (75% of adults, 76% of children) went on to do a nurse interview. Non-response weights were generated to adjust for differences between participants and non-participants to the nurse visit. These weights have been used for all analyses of nurse level data.

A number of cross-tabulations were run and chi-square tests used to check which variables from the individual and household questionnaires had a significant relationship with nurse visit response. These variables were then used in the weighting.

The first step in creating the nurse weights was to model response behaviour using logistic regression.[6] A logistic regression models the relationship between an outcome variable (response to the nurse interview) and a set of predictor variables. The predictor variables were a set of socio-demographic, participant and household/catering unit characteristics collected during the interview. Adults and children were modelled separately. The model generated a predicted probability for each participant. This is the probability the participant would take part in the nurse interview, given the characteristics of the individual and the household/catering unit. Participants with characteristics associated with non-response were under-represented in the nurse sample and therefore receive a low predicted probability. These predicted probabilities were then used to generate a set of non-response weights; participants with a low predicted probability got a larger weight, increasing their representation in the sample. The full non-response models for adults and children are given in Tables B.3 and B.4.

(Tables B.3, B.4)

The final stage of designing the nurse weights was to calibrate the weights produced by the non-response model. The weighting totals were estimates based on weighted data from the individual questionnaire. This stage of weighting makes participants to the nurse visit match the population distribution in terms of age, sex and Government Office

Region and match the weighted participants to the individual questionnaire in terms of household size, ethnicity of Main Food Provider (MFP) and economic activity of the Household Reference Person (HRP). The figures used for weighting adults are shown in the first column of Table B.9; those used for weighting children are shown in the first column of Table B.10. The initial weights were the weights from the non-response model. **(Tables B.9, B.10)**

As before, the calibration weights were re-scaled so that the sum of the weights equalled the number of participants who had a nurse visit. These are the final nurse weights for the core sample (wtn_adY123 and wtn_chY123) and adjust for unequal selection, non-response to the household/MFP and individual interviews and non-response to the nurse visit.

B.5 Effective sample size

The effect of the sample design on the precision of survey estimates is indicated by the effective sample size (neff). The effective sample size measures the size of an (unweighted) simple random sample that would achieve the same precision (standard error) as the design being implemented. If the effective sample size is close to the actual sample size then the design is efficient and has a good level of precision. The lower the effective sample size, the lower the level of precision. The actual sample size for adult interviews/participants is 1,491 and the effective sample size is 1,218. This means the individual sample has the same level of precision as a simple random sample of 1,218 hence a 95% confidence interval around an estimate of 50% is (47.2%, 52.8%). Had the effective sample size been 1,491, and therefore equal to the actual sample size, the confidence intervals would have been (47.5%, 52.5%). Large fluctuations in the size of the selection probabilities (and therefore large fluctuations in the size of the selection weights) will cause the effective sample size to be low compared with the actual sample size. Samples that select one person per household tend to have lower efficiency than samples that select all household members due to the selection weights required to make the sample representative. However, this aspect of the sample design was unavoidable in NDNS as eating habits are so highly correlated within households.

The efficiency of a sample is given by the ratio of the effective sample size to the actual sample size. The individual adult sample has an efficiency of 82%. Table B.5 shows the effective sample size and efficiency of the final individual and nurse weights for adults and children. **(Table B.5)**

In addition to the weights, the precision of estimates is also affected by the degree to which the sample is clustered. The NDNS sample was clustered within geographical areas to reduce fieldwork costs. A high degree of clustering can have a negative impact on the precision of the survey estimates, since individuals within a cluster tend to be more alike. Design factors (defts) show the extent to which the sample design has increased the standard error and can be used to assess the impact of clustering. The effects of clustering vary; it impacts more on some survey estimates than others. Table B.6 shows the design factors due to clustering for a number of estimates. Other elements of the sample design have been ignored to enable the impact of clustering to be isolated. Whilst the impact on some estimates is relatively large, the overall effects are small. For example, the estimate for children of the mean proportion of food energy taken from saturated fat has a design factor of 1.09, this means the standard error (and therefore confidence interval) around this estimate was increased by 9% by the clustered design. **(Table B.6)**

B.6 Impact of the weights

The impact of the non-response and selection weights on the data can be seen in Tables B.7 to B.10. Tables B.7 and B.8 compare weighted fully productive individuals (those completing three or four food diary days) to the UK population. Tables B.9 and B.10 compare those visited by a nurse to individuals who responded to the individual interview and completed at least three food diary days.

(Tables B.7 to B.10)

B.7 Alcohol and smoking weight

An additional weight was required for a specific analysis of smoking behaviour and alcohol consumption of individuals aged 16 years and over. Questions about smoking and drinking were asked of all adults and all young persons aged 16 to 18 years. The alcohol and smoking weight allows the information collected from both age groups to be analysed together. This weight was generated for all core sample participants at general sample addresses[7] who were aged 16 years and over and had completed an individual interview and at least three food diary days. A nurse weight was not required for this specific sample. As before, calibration methods were used to generate the weights. The initial weight was the combined selection weight (w_{sel}), described in section B.2. This weight was adjusted using an iterative procedure to give a final weight (wti_adY12316) that made the age, sex and regional distribution of the weighted sample representative of the UK population aged 16 years and over. Table B.11 shows the weighted and unweighted distribution of the participants aged 16 years and over.

(Table B.11)

B.8 Blood weights

An additional set of weights was generated to correct for differential non-response to giving a blood sample. Non-response, whether due to refusal or inability to give a blood sample, will cause the blood data to be biased if there are systematic differences between individuals that provide a blood sample and individuals that do not.

- Blood samples were taken during the nurse visit. Only participants who fulfilled certain eligibility criteria were asked whether they would be prepared to give a blood sample (see Appendix M, section M.4).

Response to the blood sample was higher for adults than for children; 74% of adults and 42% of children who had been visited by a nurse and were eligible to give blood had provided a blood sample. Response amongst children visited by a nurse was closely linked to age: whilst 58% of those aged 11 to 18 years provided a blood sample, only 20% of younger children (aged 1.5 to 3 years) did so.

The ‘blood participants’ (i.e. those who provided a blood sample) were weighted to match all ‘nurse participants’ (i.e. those who were visited by a nurse and were eligible to provide a blood sample). It can be assumed that the eligible nurse participants (weighted by the nurse weight) are representative of all eligible persons in the population, since the nurse weights make the full nurse sample representative of the population. The final blood weights should therefore make the blood sample participants representative of all eligible persons in the population. This assumption is made because there are no available estimates of the actual eligible population (i.e. the population providing a blood sample).

The methods used to generate the blood weights were similar to those used to generate the nurse weights. Cross-tabs and chi-square tests were used to check which variables from the individual and household questionnaires were significantly associated with a participant giving blood. These variables were then entered into a logistic regression model.

A logistic regression models the relationship between a binary outcome variable (whether or not a participant gave blood) and a set of predictor variables. The predictor variables were a set of socio-demographic participant and household characteristics collected from the individual interview. Adults and children were modeled separately. The model generated a predicted probability for each participant. This is the probability the participant would give blood during the nurse interview, given the characteristics of the participant and the household. Participants with characteristics associated with non-response were under-represented in the blood sample and therefore received a low predicted probability. These predicted probabilities were then used to generate a set of non-response weights; participants with a low predicted probability received a larger weight, increasing their representation in the sample.

The same variables were entered into the two non-response models. These were; age and sex, region, household size, ethnicity, work status (for children this was work status of the HRP), general health and response to the weight and height measurements taken by the interviewer. The full models for adults and children are given in Tables B.12 and B.13. **(Tables B.12, B.13)**

The non-response weights from the model were combined with the final nurse weights to give the final blood weights (the final nurse weights incorporate the selection weights, weights for non-response to the individual questionnaire and weights for non-response to the nurse visit). These weights were scaled, so the mean weight equalled one and the weighted sample size matched the unweighted sample size.

The final stage of the nurse weights was to calibrate the weights produced by the non-response model. This step was not done for the blood weights. The weighted blood sample was already very close to the weighted eligible nurse sample and hence this additional step was not necessary. The impact of the weights on key variables for adults and children are shown in Tables B.14 and B.15.

(Tables B.14, B.15)

---

[1] Chapter 2 which covers response rates uses unweighted data.

[2] A Dwelling Unit is an address or part of an address, which has its own front door. The front door does not have to be at street level, but it must separate one part of the address from other parts (i.e. only those who live behind the door have access to the area, it is not a communal part of the address).

[3] A Catering Unit is a "group of people who eat food that is bought and prepared for them (largely) as a group". A household will consist of more than one catering unit if any of its members generally buy and prepare food separately from other members. For example, a household of students may share a living space but they all cook and prepare food independently and hence would form separate catering units within the household.

[4] A Kish grid is a framework to ensure that the unit is selected without interviewer bias. The number of units is listed across the top of the grid, with a random number below to indicate which unit should be selected.

[5] (Office for National Statistics. *Mid 2008, 2009 and 2010 Population Estimates*. [Online] Available: http://www.statistics.gov.uk/statbase/Product.asp?vlnk=15106 (accessed 17/01/2011).

[6] This step was not carried out on Year 1 data. The small sample sizes resulted in weak non-response models and this stage was left out.

[7] It was more efficient to exclude those aged 16 to 18 years from the child boost than include them and weight them down.