Appendix B. Weighting the NDNS RP Scotland sample

Sarah Tipping

B.1 Introduction

The NDNS RP Scotland sample requires weights to adjust for differences in sample selection and response.[1] The weights adjust for:

- differential selection probabilities of addresses, households and individuals,
- non-response to the individual questionnaire,
- non-response to the nurse visit, and
- non-response of participants aged 16 years and older to the physical activity self-completion questionnaire (the RPAQ)

Weights have also been generated to adjust for non-response to providing a blood sample, a 24-hour urine sample and wearing an ActiGraph. Non-response weights were generated using a combination of logistic regression modelling and calibration.

The following design features of the combined Years 1 to 4 sample needed to be taken into account when generating the Scotland weights:

- Some analyses in this report require non-standard age breaks which cross the traditional adult (aged 19 years and over)/child (aged 1.5 to 18 years) split. To allow this, a single weight has been created that can be used for both adults and children.
- In Year 4 the division of addresses between 'main' sample addresses (where both an adult and child could be interviewed) and 'child' addresses (where only a child could be selected) varied across quarters. This means adults have lower selection probabilities in some quarters, which must be adjusted for.

- In the final quarter of Year 4 an additional sample was selected to ensure that the required numbers in the Scotland boost sample were achieved. This sample was weighted down.

The weighted sample is representative of the Scotland population living in private households. All figures presented in the report are based on weighted data.[2]

B.2 Selection weights

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

- Addresses across months
- Dwelling units at multi dwelling unit[3] addresses
- Catering units at multi catering unit[4] addresses
- Individuals within dwelling or catering units

The first step was to generate weights to correct for differential selection probabilities of addresses by country and quarter. There are a number of features of the sample design that impact on selection probabilities of addresses. These are:

- The Run In sample
- Additional Scotland boost in Year 4, quarter 4
- Changes to the number of 'main' addresses issued per point in Year 4

These are each described in more detail below.

Prior to the launch of Year 1 fieldwork, the NDNS RP 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. The Run In sample was carried out in February and March 2008, meaning there is an unequal distribution of the sample across months in Year 1.

In the final quarter of Year 4 an additional sample was selected to ensure that the required numbers in the Scotland boost sample were achieved. This sample was weighted down to ensure the data are evenly distributed across quarters.

In Year 4 the split between ‘main’ sample addresses (where both an adult and child could be interviewed) and ‘child’ addresses (where only a child could be selected) varied across quarters. This means adults have lower selection probabilities in some quarters.

In order to generate the selection weights, the selection probabilities of each address in each year were calculated. Whilst the core sample addresses each year had originally been selected with equal probability, the points above mean addresses in the combined sample (including the country boosts) have unequal selection probabilities.

The address selection probabilities were worked out for each year based on the total number of addresses selected (in that year) divided by the overall number of addresses. Adjustments were then made for selecting additional addresses in Year 4, quarter 4 and the Run In. The selection probabilities of these additional cases were calculated in the same way, and then combined with the rest of the sample. A final adjustment was made for changing the main/young person address split in Year 4. These weights were then combined to create the address selection weights (w0).

B.2.1 Selection of dwelling units and catering units

Most addresses selected from the 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.5 The selected dwelling unit was then included in the sample. The dwelling unit selection weights (w1) 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.[5] The catering unit selection weights (w2) 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.

B.2.2 Selection of one adult and one child (where present) per catering unit

The selection of individuals within catering units depended on the selection 'type' of the address. Each sample point contained nine 'main' sample addresses and 18 'child' addresses. At main 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' addresses one child was selected at random by the interviewer. In previous years adults and children were always analysed separately, hence this sample design feature was intended to reduce costs without increasing the degree of clustering in the sample. The overlap in age breaks means there will now be a small increase in clustering within the sample. This will have a small impact on precision; the standard errors will increase by a small amount.

Individual selection weights (w3) 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.

Overall selection weights (w_{sel}) were the product of the address, dwelling unit, catering unit and individual selection weights.

B.3 Non-response weights

Children have higher selection probabilities than adults because individuals aged 19 years and over are screened out at all 'young person' addresses. The weights down-weight children relative to adults, otherwise children would be over-represented in the sample.

A set of individual weights were generated for analysis of fully responding individuals (the individuals who responded to the individual interview and completed three or four food diary days). As with previous/other NDNS samples, these weights were generated using calibration methods. Calibration methods take an initial weight (in this case the selection weight) and then adjust (or *calibrate*) it. The process generates a weight that produces survey estimates that exactly match the population for the specific characteristics used in the adjustment. The aim in the NDNS RP was to reduce bias resulting from sampling error and differential non-response by age and sex to the individual interview. An iterative procedure was used to adjust an initial 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 composite selection weights (w_{sel}), which are described in Section B.2, were used as the initial weights.

The sample was calibrated to the Scotland mid-year population totals. In addition, the calibration was run separately for adults and children. Children from 'main' sample addresses and children from 'young person' addresses were weighted together. The two sub-samples of children should always be analysed together to get maximum numbers. Generating weights for adults and children separately means the sample of adults is representative of all adults in the population and children are representative of all children, once combined the adults and children are then in their correct population proportions. This means the standard NDNS age breaks can still be used (in chapters 3[6] to 7,9,10) , in addition to the new age breaks (chapter 8).

The key variables used to create the individual weight were: age (grouped) and sex. The age breaks used are shown in Table B.2. The population figures were taken from the mid-year population estimates.[7] As there are now four years' worth of NDNS RP data, the average population of the last four years was used. This was generated using the four most recent years of population data available (2008 to 2011).

Table B.1 shows the population figures used to weight adults and children with the unweighted and weighted sample distributions. It can be seen how the unweighted sample has over-sampled children relative to adults and how the selection and non-response weighting factors correct for this.

(Table B.1)

B.4 Recent Physical Activity Questionnaire (RPAQ) self-completion questionnaire weights

All individuals aged 16 years and over were asked to record their physical activity over the previous seven days in a self-completion booklet (the RPAQ). Response to the RPAQ in Scotland was high (95%).

A bivariate analysis showed that there were significant differences between responders and non-responders for some individual and household characteristics, which indicated that a non-response adjustment was required. Those completing this self-completion therefore received an RPAQ weight.[8]

Response behaviour to the RPAQ was modelled using a logistic regression. A logistic regression can be used to model the relationship between an outcome variable (response to the RPAQ) 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. Participants aged 16 to 18 years were modelled with the adult participants.[9]

The model generated a predicted probability for each participant. This is the probability the participant would complete the RPAQ, given the characteristics of the individual and the household/catering unit. Participants with characteristics associated with non-response were under-represented in the RPAQ 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 model for the RPAQ is given in Table B.2.

(Table B.2)

The RPAQ weights were re-scaled so that the sum of the combined adult and child weights equalled the number of participants who had completed the RPAQ. These are the final RPAQ weights for the Scottish sample (wtr_scY1234) and adjust for unequal selection, non-response to the household/Main Food Provider (MFP) and individual interviews and non-response to the RPAQ.

B.5 Nurse weights

Participants who completed three or four food diary days (i.e. those deemed fully productive) were asked to consent to a nurse visit. Approximately three quarters of these participants in Scotland (74% of adults, 74% 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.

The first step in creating the nurse weights was to model response behaviour using logistic regression. The same set of predictor variables used to model non-response to the RPAQ was used to model non-response to the nurse visit, namely, socio-demographic, participant and household/catering unit characteristics collected during the individual interview. Adults and children were modelled separately. The model generated a predicted probability for each participant. These predicted probabilities were 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.

(Table B.3 and Table B.4)

Although the intention was to create a single weight, adults and children were modelled separately because response behaviour can vary between the two groups. For example, housing tenure is not significant in the child model but is highly significant for adults.

Comparisons between weighted data from the individual questionnaire and data for nurse visit respondents weighted by the weights generated thus far showed the two distributions to be very close for both adults and children; these are shown in Tables B.5 and B.6.

(Table B.5 and Table B.6)

The nurse weights were re-scaled so that the sum of the combined adult and child weights equalled the number of participants who had a nurse visit. These are the final nurse weights for the Scotland sample (wtn_scY1234) and adjust for unequal selection, non-response to the household/MFP and individual interviews and non-response to the nurse visit.

B.6 Blood sample weights

A set of weights were 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. Participants were ineligible if they:

- had a clotting or bleeding disorder (e.g. conditions such as haemophilia and low platelets (thrombocytopenia))
- had ever had a fit (children), had a fit in the past five years (adults)
- were currently on anticoagulant drugs, e.g. Warfarin therapy
- had volunteered information that they are HIV or Hepatitis B or C positive

Response to the blood sample was higher for adults than for children; 78% of adults and 42% of children who had completed three or four diary days and were eligible to give blood had provided a blood sample.[10] To increase the number of cases available for modelling blood response, adults from the Scotland core and boost samples were combined with adults from the Northern Ireland core and boost samples. This was done to increase the overall sample size. Interaction variables were included in the model to take account of any additional variation in non-response behaviour between the two countries. However, this was not done for children. The only variables significant in the modelling were age and sex, hence the additional cases did not result in more information being included in the final non-response model. Age was the biggest predictor for children. Response amongst children was closely linked to age: whilst 59% of those aged 11 to 18 years provided a blood sample only 13% of the youngest children (aged 1.5 to 3 years) did so.

The ‘blood participants’ (i.e. those who provided a blood sample) were weighted to match eligible ‘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 that meet the NDNS RP eligibility requirements for 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 modelled separately. The model generated a predicted probability for each participant. These predicted probabilities were 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 small sample numbers for children meant only age and sex were included in the model. The full models for adults and children are given in Tables B.7 and B.8.

(Tables B.7 and B.8)

The non-response weights from the model were combined with the final nurse weights to give the final blood weights (wtb_scY1234). 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. The weights for adults and children were combined and then scaled, so the mean combined weight was equal to one and the weighted sample size matched the unweighted sample size.

The impact of the blood weights on key variables for adults and children are shown in Tables B.9 and B.10. These tables compare those visited by a nurse to individuals who provided a usable blood sample.

(Tables B.9 and B.10)

B.7 24-hour urine sample weights

Two sets of weights were generated to correct for differential non-response to giving a 24-hour urine sample. The 24-hour urine sample data will be biased if systematic differences between individuals that do and do not provide a complete urine sample are not corrected for. Two different definitions of completeness were agreed for the analysis of children aged 4 to 10 years, which has resulted in two sets of 24-hour urine weights.

24-hour urine samples were taken during the nurse visit. All individuals aged four years and over, with the exception of children still in nappies, were asked to provide a sample.

The analysis needed to exclude, as far as possible, all individuals with incomplete collections without introducing significant bias. Sample completeness was determined by the amount of PABA excretion and whether the respondent reported any missed collections. Individuals who provided an incomplete urine sample were counted as non-responders.

Two different definitions of completeness have been used in this report. The first definition of completeness applies the same criteria as that used for adults in the 2011 Assessment of dietary sodium in adults (aged 19 to 64 years).[11] A sample was deemed to be complete if the either the levels of PABA excretion were sufficiently high or (where the individual had declined or failed to take the full

PABA requirement) the individual claimed it to be complete. This definition is referred to as the ‘standard adult criteria’ and was used to identify responding individuals for the first set of weights and was applied to all individuals aged four years and over.

The second definition was used for children aged 4 to 10 years. By this definition, children had given a complete sample if they reported no missed collections over the required time period. This definition is referred to as ‘complete by claim’ and was used for the second set of weights to identify responding children aged 4 to 10 years. For all other participants (i.e. those aged 11 years and over) the ‘standard adult criteria’ was again used for the second set of weights. Hence, the two sets of weights are identical for participants aged 11 years and over.

The eligibility criteria meant that participants who provided a usable 24-hour urine sample were weighted to match the eligible nurse participants (i.e. those who were visited by a nurse and were eligible to provide a 24-hour urine 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 24-hour urine weights therefore make the 24-hour urine 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 24-hour urine sample).

24-hour urine sample weights were generated using logistic regression models. The predictor variables were a set of socio-demographic participant and household characteristics collected from the individual interview. Adults and children were modelled separately. Adults were modelled once using an outcome code based on the ‘standard adult criteria’. Children were modelled twice: once using an outcome based on ‘standard adult criteria’ and once using an outcome that was based on ‘complete by claim’ for children aged 4 to 10 years and the ‘standard adult criteria’ for children aged 11 to 18 years.

Predicted probabilities were generated for each participant and used to generate the non-response weights. Participants with a low predicted probability received a larger weight, increasing their representation in the sample. The full model for adults is given in Table B.11. To increase the number of cases available for modelling, adults from Scotland core and boost samples were combined with adults from the Northern Ireland core and boost samples. Interaction variables were included to take account of any additional variation in non-response behaviour between the two countries. This was not done for children because the only variables significant in the modelling were age and sex: the additional cases did not result in more variables being included in the final non-response model. The full models for children for the outcome based on ‘standard adult criteria’ and for the outcome based on ‘complete by claim’ are given in Tables B.12 and B.13, respectively. The same set of variables were significant in both child non-response models.

(Tables B.11, B.12 and B.13)

The non-response weights from the models were combined with the final nurse weights to give the final 24-hour urine sample 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). The weights for adults and children were combined and then scaled, so the mean combined weight was equal to one and the weighted sample size matched the unweighted sample size. Two sets of weights are produced: wtu_scY1234v1 is based on the ‘standard adult criteria’ for all individuals aged four years and over, wtu_scY1234v2 is based on complete by claim for individuals aged 4 to 10 years and the ‘standard adult criteria’ for individuals aged 11 to 15 years. Either weight can be used for analysis of adults; they will give the same weighted frequencies, although weighted totals will be very slightly different due to scaling.

The impact of the 24-hour urine weights on key variables for adults and children are shown in Tables B.14 and B.15. These tables compare those who gave a complete 24-hour urine sample to individuals who were eligible to give a 24-hour urine sample.

(Tables B.14 and B.15)

B.8 ActiGraph weights

A set of weights were generated to correct for differential non-response to wearing an ActiGraph. The ActiGraph data will be biased if there are systematic differences between individuals who did and did not wear an ActiGraph and these are not corrected for.

In Year 1, all children aged 4 to 10 years were asked to wear an ActiGraph. In subsequent years the eligibility criteria were extended to include children aged 11 to 15 years.

As with the 24-hour urine samples, a decision had to be made as to which individual records should be included in the analysis. For the purposes of this report, it was decided that only individuals with 24 hours of included wear time should be included in the analysis. These cases are flagged in the dataset as IncludeRecord = 1.[12]

The final ActiGraph weights are intended to make the ActiGraph sample representative of all eligible children in the population. The first step was to make an adjustment so that children aged 4 to 10 years were not over-represented in the ActiGraph sample. This adjustment weights down the 4 to 10 year olds from all four survey years and means the proportion of 4 to 10 year olds in the weighted Years 1 to 4 sample is equal to the proportion in the unweighted Years 2 to 4 sample (i.e. the survey years that included children aged 11 to 15 years).

Once this adjustment had been made, the ActiGraph weights were generated using a logistic regression model. The predictor variables tested were a set of socio-demographic participant and household characteristics collected from the individual interview, although only age and sex remained in the final models. Predicted probabilities were generated for each participant and 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 full model is given in Table B.16.

(Table B.16)

The non-response weights from the model were combined with the final interview weights to give the final ActiGraph weights (wta_scY1234). The final interview weights incorporate the selection weights and weights for non-response to the individual questionnaire. The weights were combined and then scaled, so the mean combined weight was equal to one and the weighted sample size matched the unweighted sample size.

The impact of the ActiGraph weights on key variables are shown in Tables B.17. These tables compare those who were eligible to wear an ActiGraph to those who provided useable ActiGraph data.

(Table B.17)

B.9 Sampling efficiency and 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.

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. The address selection weights in the Scotland Years 1 to 4 combined sample are relatively variable due to the additional addresses selected in Year 4, quarter 4. In addition, the requirement for age breaks that cut across the traditional adult/child sample split mean children are weighted down relative to adults. Samples that select one person per household also tend to have lower efficiency than samples that select all household members due to the selection weights required to make the sample representative. However, these aspects of the sample design were necessary and it can be shown that the overall precision of the sample is still improved by their inclusion.

The total sample size for all participants in Scotland (adults and children) is 1,695. The effective sample size is 954. This means any survey estimates from the total sample have the same level of precision as estimates from a simple random sample of 954, hence a 95% confidence interval around an estimate of 50% is (46.8%, 53.2%). Had the effective sample size been 1,695, and therefore equal to the actual sample size, the confidence intervals would have been (47.6%, 52.4%). The efficiency of a sample is given by the ratio of the effective sample size to the actual sample size. The individual sample has an efficiency of 56%.

In practice most analyses will consider adults and children separately. This improves sample efficiency because the weights are less variable when looking solely at one group or the other, since they are not affected by the need to weight down children. The total number of adults (aged 19 and over) in the interview sample is 867 and the effective sample size is 652, giving a sample efficiency of 75%. The corresponding total for children (aged 1.5 to 18 years) is 828, with an effective sample size of 627 and sample efficiency of 76%.

Table B.18 shows the efficiency of the weights overall, for adults only and children only.

(Table B.18)

In addition to the weights, the precision of estimates is also affected by the degree to which the sample is clustered. This discussed in Appendix B of the UK report.1

---

[1] Full detail of the weighting strategy for the UK sample is described in Appendix B of the UK report. https://www.gov.uk/government/publications/national-diet-and-nutrition-survey-results-from-years-1-to-4-combined-of-the-rolling-programme-for-2008-and-2009-to-2011-and-2012 (accessed 07/08/14)

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

[3] 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).

[4] 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.

[5] 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.

[6] The smoking and drinking analysis in chapter 3 does not use the standard age breaks but rather, 16 to 18 year olds are analysed with adults.

[7] (Office for National Statistics. *Mid 2008, 2009, 2010 and 2011 Population Estimates*. [Online] Available: http://www.gro-scotland.gov.uk/statistics/theme/population/estimates/special-area/ (accessed 01/07/2013).

[8] Note that response to RPAQ and the nurse visit was not hierarchical; it was possible for a participant to complete the RPAQ section but not the nurse visit, and vice versa

[9] Children under the age of 16 were not asked to fill in the self-completion booklet. This left too few children (aged 16 to 18 years) to produce a separate child non-response model.

[10] These response rates differ from the blood response rates presented in Chapter 2 as they are based on eligible individuals only.

[11] National Diet and Nutrition Survey - Assessment of dietary sodium in adults (aged 19 to 64 years) in England, 2011. Katharine Sadler et al [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/127916/Sodium-Survey-England-2011_Text_to-DH_FINAL1.pdf.pdf (accessed 06/08/2014).

[12] It would be possible to apply different inclusion criteria, for example to include individuals with incomplete records or shorter periods of wear or to use different criteria for different age groups However, a researcher using alternative cut-offs would need to apply the individual weight (wti_scY1234) during analysis.