Appendix B Weighting the NDNS RP Wales sample

Sarah Tipping

B.1 Introduction

The NDNS RP Wales 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 and a 24-hour urine sample. Non-response weights were generated using a combination of logistic regression modelling and calibration.

The following design features were taken into account when generating the Wales 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 used in some previous reports. To allow this, a single weight has been created that can be used for both adults and children
- there are some variations in the division of addresses between 'basic' sample addresses (where both an adult and child could be interviewed) and 'child boost' addresses (where only a child could be selected) This means adults have lower selection probabilities in some years/quarters, which must be adjusted for

The weighted sample is representative of the Wales population living in private households.[2] All figures presented in the report are based on weighted data.[3]

B.2 Selection weights

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

- addresses across months
- dwelling units at multi dwelling unit[4] addresses
- catering units at multi catering unit[5] addresses
- individuals within dwelling or catering units

The first step was to generate weights to correct for differential selection probabilities of addresses. The selection probabilities varied because of changes in how the 27 addresses in each point were split between 'basic' sample addresses (where both an adult and child could be interviewed) and 'child boost' addresses (where only a child could be selected). In Years 2 and 3 there were 9 'basic' and 18 'child boost' addresses per point. In the first half of Year 4 the split was 11/16 changing to 10/17 in the second half of Year 4 to compensate for the over-achievement of adult interviews in quarters 1 and 2. The 10/16 split was retained for Year 5. This means the selection probabilities for adults vary across years.

In order to generate the selection weights, the selection probabilities of each address in each year were calculated. Whilst the addresses each year had originally been selected with equal probability, the changes detailed above mean addresses had unequal selection probabilities. The address selection weights were generated as the overall number of addresses selected per year divided by the total number of addresses. Adjustments were then made for changing the 'basic'/'child boost' address split in Year 4. These weights were then combined to create the address selection weights (w0).

The remaining selection weights for dwelling units and individuals were generated in the same way as previous years. These are outlined below.

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.[6] 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. 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. Sample points in Years 2 and 3 contained nine 'basic' sample addresses and 18 'child boost' addresses. This was changed in Year 4, to increase the number of adult interviews. Year 4 quarters 1 and 2 sample points contained 11 'basic' and 16 'child boost' addresses. The number of 'basic' addresses was later reduced, so that the split in the second half of Year 4 fieldwork (quarters 3 and 4) was 10 'basic' addresses and 17 'child boost' addresses.

At 'basic' 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. In previous NDNS RP reports 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 'child boost' addresses. However, adults and children in the Years 2 to 5 Wales data are treated as a single sample. This is because some of the age breaks used in the analysis span the two groups. (The analysis in chapter 8 requires the following age breakdown: 11 to 15 years; 16 to 24 years; 25 to 49 years and 50 to 64 years. Further, in the smoking and drinking analysis in chapter 3, 16 to 18 year olds are included with adults aged 19 years and over, rather than with younger children). As a result a single set of weights were generated that could be used for both adults and children. These 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 852 individuals who responded to the individual interview and completed three or four food diary days). 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 mid-year population totals and was run separately for adults and children. Children from 'main' sample addresses and children from 'child' 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 RP age breaks can still be used (in chapters 3[7] to 7 and chapters 9 and 10), in addition to the new age breaks (chapters 8).

The key variables used to create the individual weight were: age (grouped) and sex. The age breaks used are shown in table B.1. The population figures were taken from the mid-year population estimates.[8] As there was 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 (2009 to 2012).

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 was high (94%).

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.[9]

Response behaviour to the RPAQ was modelled using a logistic regression. A logistic regression can be used to model the relationship between a binary 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.[10]

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 UK sample (wtr_wY2345) 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 (71% of adults, 75% 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. A logistic regression can be used to model the relationship between an outcome variable (response to the nurse interview) and a set of predictor variables, 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, household size has a bigger impact on response behaviour of children than adults, hence it is not significant in the adult model but highly significant for children.

Comparisons between weighted data from the individual questionnaire and data for nurse visit respondents weighted by the weights from the non-response model show 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 UK sample (wtn_wY2345) 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; 80% of adults and 44% of children who had completed three or four diary days, had agreed to a nurse visit and were eligible to give blood had provided a blood sample.[11] Response amongst children was closely linked to age: whilst 51% of those aged 11 to 18 years provided a blood sample, only 21% 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 who were eligible to provide 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. 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 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_wY2345). 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.

The 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 para-aminobenzoic acid (PABA) excretion and whether the individual reported any missed collections. Those 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).[12] 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).

The 24-hour urine sample weights for adults 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 were modelled once using an outcome code based on the ‘standard adult criteria’. 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.

(Table B.11)

These non-response weights were combined with the final nurse weights to give the final 24-hour urine sample weights for adults (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).

Two sets of weights were generated for children : one using an outcome based on ‘standard adult criteria’ and a second 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. The child weights were generated using calibration methods. The nurse weights for responding children were calibrated to age and sex population totals. This adjusted weight therefore incorporated the selection weights, weights for non-response to the individual questionnaire and weights for non-response to the nurse visit, plus an adjustment for non-response to the urine samples.

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_Y2345v1 is based on the ‘standard adult criteria’ for all individuals aged four years and over, wtu_Y2345v2 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.12 and B.13. 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.12 and B.13)

B.8 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 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 in order to provide the required dietary breakdowns.

The total sample size for all participants (adults and children) is 852. The effective sample size is 510. This means any survey estimates from the total sample have the same level of precision as estimates from a simple random sample of 510, hence a 95% confidence interval around an estimate of 50% is (45.7%, 54.3%). Had the effective sample size been 852 and therefore equal to the actual sample size, the confidence intervals would have been (46.6%, 53.4%). 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 60%.

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 461 and the effective sample size is 345, giving a sample efficiency of 75%. The corresponding total for children (aged 1.5 to 18 years) is 391, with an effective sample size of 314 and sample efficiency of 80%.

Table B.14 shows the efficiency of the weights overall, for adults only and children only.

(Table B.14)

In addition to the weights, the precision of estimates is also affected by the degree to which the sample is clustered. The NDNS RP 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.15 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 with a poor appetite has a design factor of 1.03, this means the standard error (and therefore confidence interval) around this estimate was increased by 3% by the clustered design.

(Table B.15)

---

[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/statistics/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 18/05/15).

[2] Year 1 was not included as this first fieldwork year did not include a Wales boost sample, just a core sample which comprised 32 fully responding adults and 36 fully responding children. These 68 cases would have received very extreme weights in order to make the combined Years 1 to 5 sample evenly distributed by year (to avoid the sample being biased by year of interview) and were hence excluded.

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

[4] 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).

[5] 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.

[6] 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.

[7] 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.

[8] (Office for National Statistics. *Mid 2009, 2010, 2011 and 2012 Population Estimates*. [Online] Available: http://www.ons.gov.uk/ons/taxonomy/index.html?nscl=Population#tab-sum-pub (accessed 29/07/2014).

[9] 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

[10] 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.

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

[12] 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 (accessed 18/05/15).