

Appendix B. Weighting the NDNS core sample

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 and non-response to the nurse visit. Non-response weights were generated using calibration weighting. All the figures presented in this report are based on weighted data.

B.2 Selection weights

Selection weights are required for

1. Unequal selection of dwelling units at multi-dwelling unit addresses¹, and
2. Unequal selection of individuals within dwelling units.

B.3 Selection of dwelling Units

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 grid². 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 units³ and selected one at random using a Kish grid. The catering unit selection

¹ A Dwelling Unit (DU) 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).

² A Kish grid is a framework to ensure that the dwelling unit is selected without interviewer bias. The number of dwelling units is listed across the top of the grid, with a random number below to indicate which dwelling unit should be selected.

³ A Catering Unit (CU) is a “group of people who eat food that is bought and prepared for them (largely) as a group”.

weights (w_2) adjust for this selection of catering units. The catering unit selection weights ensure that catering units in multi-occupied dwelling units are not under-represented in the sample.

The selection of individual within households depends on the selection 'type' of the address. Each sample point contains 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 18 months 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 are weighted separately as they will always be analysed separately, hence this sample design feature reduces costs but does not increase the degree of clustering in the sample.

Individual selection weights (w_3) are required to ensure individuals in larger households 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 household, for children this is the number of eligible children. Pregnant or breastfeeding women were not eligible for the survey.

B.4 Run-in sample

An extensive piloting stage meant the NDNS used a 'Run In' sample rather than a dress rehearsal to test field procedures. The Run In sample was used as a dress rehearsal but could be combined with the main stage data if no significant changes to procedures were to be made, as was the case. The Run In sample was selected at the same time as the main stage sample using the same methods. The Run In sample was carried out in February and March, therefore an additional set of weights were needed to correct for the unequal distribution of sample across months. The Run In sample weights (w_4) ensure the full sample (the run-in sample plus the main stage Y1 sample) 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.5 Individual non-response weight

A set of household weights were not required as all analyses will be carried out at the individual level; any information collected about the household will be reported in terms of the individual.

A weight has been generated for the analysis of fully responding individuals; individuals who responded to the individual interview and completed three diary days. The individual non-response weight was generated using calibration methods. An iterative procedure is used to adjust a starting weight until the distribution of the (weighted) sample matches that of the population for a set of key variables. The adjustment keeps the values of the final weights as close as possible to those of the initial weights, which ensures the properties of the initial weights are retained in the final calibrated weights. The composite selection weights (w_{sel}), described above, were used as the initial weights.

The key variables used to create the individual weight were age (grouped) by sex and Government Office Region. The population figures used were taken from the 2008 mid-year population estimates⁴. The calibration was run separately for adults and children, children from the general sample and child boost addresses were weighted together as the two samples of children should be analysed together. The figures used for weighting adults are shown in Table B.1; those used for weighting children are shown in Table B.2.

The aim of the calibration weighting was to reduce non-response bias resulting from differential non-response at the household and individual interview. 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 (w_{ti_adult} and w_{ti_child}). Thus the final individual weights adjust for dwelling unit, catering unit and individual selection, the run in sample, and for the age/sex and regional profiles of participating individuals.

⁴ Produced for England and Wales by ONS, for Scotland by GROS and for Northern Ireland by NISRA.

Table B.1		
2008 mid-year population estimates for adults (19 years and over)		
	Estimate	Percent
Government Office Region		
North East	2,011,569	4.2
North West	5,291,821	11.2
Yorkshire & the Humber	4,031,822	8.5
East Midlands	3,438,662	7.2
West Midlands	4,137,516	8.7
East of England	4,421,525	9.3
London	5,888,766	12.4
South East	6,457,404	13.6
South West	4,083,943	8.6
Wales	2,318,837	4.9
Scotland	4,057,785	8.6
Northern Ireland	1,317,621	2.8
Age in years (grouped) by sex		
Males 19-29	4,652,175	9.8
Males 30-39	4,112,845	8.7
Males 40-49	4,481,416	9.4
Males 50-59	3,669,712	7.7
Males 60-64	1,777,828	3.7
Males 65-69	1,323,986	2.8
Males 70+	2,996,639	6.3
Females 19-29	4,465,556	9.4
Females 30-39	4,151,552	8.7
Females 40-49	4,584,980	9.7
Females 50-59	3,769,885	7.9
Females 60-64	1,861,354	3.9
Females 65-69	1,432,422	3.0
Females 70+	4,176,921	8.8
Total	47,457,271	100

Table B.2		
2008 mid-year population estimates for children (18 months to 18 years)		
	Estimate	Percent
Government Office Region		
North East	518,980	4.1
North West	1,454,362	11.4
Yorkshire & the Humber	1,084,156	8.5
East Midlands	914,555	7.2
West Midlands	1,168,068	9.2
East of England	1,200,769	9.4
London	1,548,367	12.1
South East	1,768,154	13.9
South West	1,038,336	8.1
Wales	622,109	4.9
Scotland	1,022,437	8.0
Northern Ireland	419,677	3.3
Age in years (grouped) by sex		
Males 1.5-3	935,637	7.3
Males 4-6	1,051,976	8.2
Males 7-10	1,411,904	11.1
Males 11-14	1,508,799	11.8
Males 15-18	1,631,119	12.8
Females 1.5-3	891,641	7.0
Females 4-6	1,001,395	7.8
Females 7-10	1,350,362	10.6
Females 11-14	1,436,103	11.3
Females 15-18	1,541,031	12.1
Total	12,759,967	100

B.6 Nurse interview non-response weight

Participants who completed three or four diary days were asked for their permission for a nurse to visit. 76% of adults and 77% of children who completed an individual interview and at least three diary days 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 should be used for any analyses of nurse level data.

A number of cross-tabs were run and chi-square tests used to check which variables from the individual and household questionnaire had a significant relationship with nurse visit response, these variables were then used in the weighting. As before adults and children were weighted separately.

A second stage of calibration was used to generate the nurse weight. 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 region and match the weighted participants to the individual questionnaire in terms of household size, ethnicity of Main Food Provider, economic activity of the Household Reference Person, the participant's appetite and whether the participant avoids certain foods. The figures used for weighting adults are shown in the first column of Table B.6; those used for weighting children are shown in the first column of Table B.7. The initial weights were the individual weights; wti_adult and wti_child.

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_adult and wtn_child) and adjust for unequal selection, non-response to the household and individual interview and non-response to the nurse visit.

B.7 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 efficiency of a sample is given by the ratio of the effective sample size to 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. Table B.3 shows the effective sample size and efficiency of the final individual and nurse weights.

Variable name	Variable label	N	Min	Max	Mean	Std. Deviation	Effective sample size	Efficiency
wti_adult	Adult weight for individual and diary	548	0.25	3.13	1.00	0.50	438	80%
wtn_adult	Adult weight for nurse	418	0.12	3.57	1.00	0.53	327	78%
wti_child	Child weight for individual and diary	583	0.18	3.53	1.00	0.57	440	76%
wtn_child	Child weight for nurse	450	0.13	3.90	1.00	0.60	330	73%

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 factors⁵ 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.4 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 1.11, this means the standard error (and therefore confidence interval) around this estimate was increased by 11% by the clustered design.

Table B.4				
Estimated design factors due to the effects of clustering for a set of key survey estimates				
Survey estimates	Children		Adults	
	Mean	deft	Mean	deft
Mean % energy from saturated fatty acids	80	1.11	94	1.00
Mean fruit intake g/day	13	1.17	13	1.08
	%	deft	%	deft
How often eat meals out				
3 or more times per week	2	1.11	5	1.02
1-2 times per week	16	1.12	21	1.04
1-2 times per month	51	1.00	45	1.27
Rarely or never?	31	1.05	29	1.16
Type of appetite				
Good	68	0.88	70	1.09
Average	25	0.93	24	1.12
Poor	7	1.17	6	1.24
Use services (lunch club, meals on wheels, etc)				
Yes	-		21	1.07
No	-		79	1.07
What usually have for lunch at school				
Cooked school meal	35	1.08	-	
Cold school meal	9	1.06	-	
Packed lunch (from home)	43	1.14	-	
Buy lunch from shop/cafe	9	1.00	-	
Go home	3	1.01	-	
Do not eat lunch	2	1.00	-	
<i>Base (unweighted)</i>	<i>583</i>		<i>548</i>	

⁵ The design factors were generated as the ratio of standard errors that assume clustering to standard errors that assume simple random sampling.

B.8 Impact of the weights

The impact of the non-response and selection weights on the data can be seen in Tables B.5 to B.8. Tables B.5 and B.6 compare weighted individuals who responded to the individual interview and at least three diary days to the population. Tables B.7 and B.8 compare participants to the nurse visit to individuals who responded to the individual interview and at least three diary days weighted by the individual weight.

It must be noted that the application of non-response weights is not guaranteed to reduce bias for all of the many outcomes and behaviours measured as part of this survey. This is because weighting is equivalent to replacing members of a subgroup that failed to respond with replicates of responding members of the same subgroup.

Table B.5

Weighted and unweighted distribution of adult participants to the individual questionnaire

	Population estimates	Participants to individual questionnaire		
	%	Unweighted %	Weighted by selection wt %	Weighted by wti_adult %
Government Office Region				
North East	4.2	3.8	3.5	4.2
North West	11.2	10.6	10.6	11.2
Yorkshire & the Humber	8.5	10.0	9.5	8.5
East Midlands	7.2	9.3	8.7	7.2
West Midlands	8.7	9.5	9.7	8.7
East of England	9.3	7.8	8.7	9.3
London	12.4	9.5	9.9	12.4
South East	13.6	13.3	13.3	13.6
South West	8.6	8.4	8.9	8.6
Wales	4.9	5.8	5.4	4.9
Scotland	8.6	8.0	7.9	8.6
Northern Ireland	2.8	3.8	4.0	2.8
Age in years (grouped) by sex				
Males 19-29	9.8	6.4	8.1	9.8
Males 30-39	8.7	7.5	7.3	8.7
Males 40-49	9.4	7.5	7.6	9.4
Males 50-59	7.7	8.4	8.6	7.7
Males 60-64	3.7	3.3	3.8	3.7
Males 65-69	2.8	3.1	3.6	2.8
Males 70+	6.3	5.8	4.9	6.3
Females 19-29	9.4	9.1	9.5	9.4
Females 30-39	8.7	11.7	11.4	8.7
Females 40-49	9.7	9.3	10.1	9.7
Females 50-59	7.9	10.8	10.7	7.9
Females 60-64	3.9	5.3	4.6	3.9
Females 65-69	3.0	3.3	2.9	3.0
Females 70+	8.8	8.6	6.8	8.8
<i>Base (unweighted)</i>	<i>47,457,271</i>	<i>548</i>	<i>548</i>	<i>548</i>

Table B.6

Weighted and unweighted distribution of child participants to the individual questionnaire

	Population estimates	Participants to individual questionnaire		
	%	Unweighted %	Weighted by selection wt %	Weighted by wti_child %
Government Office Region				
North East	4.1	4.1	4.3	4.1
North West	11.4	10.8	12.0	11.4
Yorkshire & the Humber	8.5	9.4	9.1	8.5
East Midlands	7.2	10.1	10.1	7.2
West Midlands	9.2	8.7	7.7	9.2
East of England	9.4	8.4	8.0	9.4
London	12.1	10.3	10.8	12.1
South East	13.9	11.3	12.1	13.9
South West	8.1	9.1	8.3	8.1
Wales	4.9	6.2	5.8	4.9
Scotland	8.0	8.7	8.3	8.0
Northern Ireland	3.3	2.7	3.5	3.3
Age in years (grouped) by sex				
Males 1.5-3	7.3	11.1	8.9	7.3
Males 4-6	8.2	8.9	8.6	8.2
Males 7-10	11.1	11.5	12.6	11.1
Males 11-14	11.8	9.6	10.0	11.8
Males 15-18	12.8	9.9	8.8	12.8
Females 1.5-3	7.0	9.6	7.9	7.0
Females 4-6	7.8	8.4	8.9	7.8
Females 7-10	10.6	12.0	14.5	10.6
Females 11-14	11.3	8.7	9.6	11.3
Females 15-18	12.1	10.1	10.3	12.1
<i>Base (unweighted)</i>	<i>12,759,967</i>	<i>583</i>	<i>583</i>	<i>583</i>

Table B.7

Weighted and unweighted distribution of adult participants to the nurse visit

	Participants to individual questionnaire	Participants to nurse visit			
	Weighted by wti_adult %	Unweighted %	Weighted by selection wt %	Weighted by wti_adult %	Weighted by wtn_adult %
Number of people in household					
1	17.5	29.4	17.4	18.8	17.5
2	33.1	30.6	33.1	33.4	33.1
3	20.9	18.4	21.6	20.3	20.9
4	15.1	13.4	16.5	16.0	15.1
5	13.4	8.1	11.4	11.4	13.4
Work status of participant					
Working (full or part-time)	61.8	58.1	62.0	61.0	61.8
Not working	38.2	41.9	38.0	39.0	38.2
Ethnic group					
White	91.0	93.1	92.7	92.0	91.0
Not white	9.0	6.9	7.3	8.0	9.0
Tenure					
Own/buying	71.1	69.9	72.5	71.1	71.1
Rent	28.9	30.1	27.5	28.9	28.9
Type of appetite					
Good	70.3	69.6	72.0	72.5	70.3
Average	23.7	23.2	21.9	21.4	23.7
Poor	6.0	7.2	6.1	6.1	6.0
Avoid any foods					
Yes	58.9	60.0	60.6	60.2	58.9
No	41.1	40.0	39.4	39.8	41.1
Government Office Region					
North East	4.2	4.8	4.2	5.1	4.2
North West	11.2	11.0	11.2	11.9	11.2
Yorkshire & the Humber	8.5	8.9	7.5	6.7	8.5
East Midlands	7.2	9.3	9.2	7.5	7.2
West Midlands	8.7	6.7	7.3	6.5	8.7
East of England	9.3	7.4	7.9	8.6	9.3
London	12.4	9.3	10.3	12.7	12.4
South East	13.6	14.8	15.3	15.5	13.6
South West	8.6	8.9	8.6	8.5	8.6
Wales	4.9	6.0	5.8	5.2	4.9
Scotland	8.6	8.4	7.9	8.4	8.6
Northern Ireland	2.8	4.5	4.9	3.4	2.8

Table B.7 continued

**Age in years
(grouped) by sex**

Males 19-29	9.8	6.9	9.0	10.6	9.8
Males 30-39	8.7	7.4	7.1	8.4	8.7
Males 40-49	9.4	7.2	7.4	9.1	9.4
Males 50-59	7.7	8.9	9.0	7.9	7.7
Males 60-64	3.7	3.3	3.8	3.8	3.7
Males 65-69	2.8	3.1	3.7	2.9	2.8
Males 70+	6.3	6.0	5.0	6.5	6.3
Females 19-29	9.4	8.6	9.4	9.3	9.4
Females 30-39	8.7	11.5	11.4	8.8	8.7
Females 40-49	9.7	8.9	9.1	8.7	9.7
Females 50-59	7.9	10.8	10.6	7.9	7.9
Females 60-64	3.9	4.3	3.5	2.8	3.9
Females 65-69	3.0	4.1	3.7	3.8	3.0
Females 70+	8.8	9.1	7.3	9.3	8.8
<i>Base (unweighted)</i>	<i>548</i>	<i>418</i>	<i>418</i>	<i>418</i>	<i>418</i>

Table B.8

Weighted and unweighted distribution of child participants to the nurse visit

	Participants to individual questionnaire		Participants to nurse visit		
	Weighted by wti_child %	Unweighted %	Weighted by selection wt %	Weighted by wti_child %	Weighted by wtn_child %
Number of people in household					
1	0.0	0.0	0.0	0.0	0.0
2	3.9	5.1	3.0	3.0	3.9
3	22.9	32.9	22.5	23.0	22.9
4	37.3	38.9	39.1	38.6	37.3
5	35.9	23.1	35.4	35.4	35.9
Work status of HRP					
Working (full or part-time)	80.2	83.6	81.8	81.3	80.2
Not working	15.7	14.2	15.7	16.0	15.7
Ethnic group					
White	83.2	85.6	84.2	84.7	83.2
Not white	16.8	14.4	15.8	15.3	16.8
Tenure					
Own/buying	67.1	67.8	67.1	66.2	67.1
Rent	32.9	32.2	32.9	33.8	32.9
Type of appetite					
Good	68.3	67.3	68.0	69.0	68.3
Average	25.1	25.8	24.8	24.2	25.1
Poor	6.6	6.9	7.2	6.9	6.6

Table B.8 continued

Avoid any foods

Yes	73.7	70.7	73.1	74.4	73.7
No	26.3	29.3	26.9	25.6	26.3

Government Office

Region

North East	4.1	4.9	5.0	4.7	4.1
North West	11.4	10.7	11.5	11.1	11.4
Yorkshire & the Humber	8.5	8.2	8.1	7.7	8.5
East Midlands	7.2	10.7	10.9	7.9	7.2
West Midlands	9.2	7.3	6.7	7.9	9.2
East of England	9.4	7.1	6.5	7.9	9.4
London	12.1	10.0	10.3	11.4	12.1
South East	13.9	12.4	13.5	15.6	13.9
South West	8.1	10.4	9.5	9.2	8.1
Wales	4.9	6.7	6.4	5.3	4.9
Scotland	8.0	9.1	8.6	8.3	8.0
Northern Ireland	3.3	2.4	3.0	2.9	3.3

Age in years (grouped) by sex

Males 1.5-3	7.3	10.4	7.8	6.5	7.3
Males 4-6	8.2	8.2	7.9	7.8	8.2
Males 7-10	11.1	11.3	12.7	11.2	11.1
Males 11-14	11.8	8.9	9.6	11.0	11.8
Males 15-18	12.8	10.0	9.1	13.3	12.8
Females 1.5-3	7.0	10.0	8.0	7.0	7.0
Females 4-6	7.8	8.4	9.4	8.2	7.8
Females 7-10	10.6	13.3	15.4	11.2	10.6
Females 11-14	11.3	9.6	10.0	11.9	11.3
Females 15-18	12.1	9.8	10.1	12.0	12.1

<i>Base (unweighted)</i>	583	450	450	450	450
--------------------------	-----	-----	-----	-----	-----

B.9 Alcohol and smoking weights

An additional weight was required for a specific analysis of smoking behaviour and alcohol use of individuals aged 16 years and over. This weight was generated for all core sample participants at general sample addresses⁶, who were aged 16 years and over and had competed an individual interview and at least three 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 composite selection weight (w_{sel}), described above. This weight was adjusted using an iterative procedure to give a final weight that made the age, sex and regional and distribution of the weighted sample match that of the population aged 16 years and over. The final weight ($w_{ti_adult16}$) makes the weighted sample of individuals

⁶ It was more efficient to exclude the child boost sample than to include them but weight them down.

representative of the UK population age years 16 and over. Table B.9 shows the weighted and unweighted distribution of the participants aged 16 years and over.

Table B.9				
Weighted and unweighted distribution of participants aged 16 years and over				
	Population estimates	Participants to individual questionnaire		
	%	Unweighted %	Weighted by selection wt %	Weighted by wti_adult16 %
Government Office Region				
North East	4.2	3.8	3.4	4.2
North West	11.2	10.6	11.1	11.2
Yorks & Humber	8.5	9.6	9.1	8.5
East Midlands	7.3	9.6	9.0	7.3
West Midlands	8.7	9.3	9.5	8.7
East of England	9.3	7.4	8.2	9.3
London	12.3	9.3	9.8	12.3
South East	13.6	13.4	13.1	13.6
South West	8.6	9.4	9.5	8.6
Wales	4.9	5.7	5.2	4.9
Scotland	8.5	8.1	7.8	8.5
Northern Ireland	2.8	3.9	4.3	2.8
Age in years (grouped) by sex				
Males 16-18	2.5	3.3	2.3	2.5
Males 19-24	5.2	2.9	4.5	5.2
Males 25-39	12.4	10.1	10.2	12.4
Males 40-49	9.0	7.0	7.2	9.0
Males 50-59	7.4	7.9	8.2	7.4
Males 60-64	3.6	3.1	3.6	3.6
Males 65-69	2.7	2.9	3.4	2.7
Males 70+	6.0	5.5	4.7	6.0
Females 16-18	2.3	2.7	2.9	2.3
Females 19-24	4.9	4.6	5.1	4.9
Females 25-39	12.3	14.9	14.7	12.3
Females 40-49	9.2	8.7	9.6	9.2
Females 50-59	7.6	10.1	10.1	7.6
Females 60-64	3.7	5.0	4.3	3.7
Females 65-69	2.9	3.1	2.7	2.9
Females 70+	8.4	8.1	6.5	8.4
<i>Base (unweighted)</i>	<i>49,865,937</i>	<i>583</i>	<i>583</i>	<i>583</i>