

Appendix B. Weighting methods in DNSIYCS

Sarah Tipping

Table of contents

B.1. Background	1
B.2. Interview weighting factors.....	1

List of tables

Table B.1	4
Table B.2	5
Table B.3	6

List of figures

Figure B.1: Overview of methods used for the interview weighting factor	3
---	---



Appendix B. Weighting methods in DNSIYCS

B.1. Background

The Diet and Nutrition Survey of Infants and Young Children (DNSIYC) required a set of weighting factors to adjust the sample for differences in sample selection and response. The weighting factors adjust for differential selection probabilities of boost sample members, non-response to the individual questionnaire, non-response to the clinic visit and non-response to providing a blood sample.

DNSIYC incorporated a boost sample of parents resident in Scotland. The analysis in this report is based on data collected from parents resident in Scotland, taken from both the core sample and Scottish boost sample. This combined Scottish sample uses the DNSIYC weighting factors for non-response to the individual questionnaire but not for non-response to the clinic visit or blood samples, since this information was not collected for boost sample members. The weighting factors used in the analysis are the same as those used for the combined UK sample as the weights were designed for this dual purpose. The design of these weights is outlined below.

B.2. Interview weighting factors

An interview weighting factor was required for the 2,683 parents (616 in Scotland) who responded to the individual interview and completed three or four food diary days for their child. This weighting factor was generated using a combination of logistic regression modelling and calibration. The aim was to reduce bias resulting from sampling error and differential non-response.

The first step was to model response behaviour to the interview using logistic regression. The DNSIYC sample contained two boost samples; a boost of Healthy Start recipients and a boost sample of individuals living in Scotland. The modelling was therefore carried out separately for the core sample from England, Wales and Northern Ireland and for the core and boost sample from Scotland. The Healthy Start recipients were excluded at this step. The weighting factors from the Scottish non-response model fed into both the overall weighting factors and into a separate weighting factor for analysis of the Scottish sample only. Having a separate model for Scotland ensures any estimates for Scotland in the overall sample match those produced for the Scottish-only sample.

A logistic regression models the relationship between an outcome variable (response to the interview) and a set of predictor variables. The predictor variables were a set of address, household and individual characteristics taken from the Child Benefit register. The model was used to generate predicted probabilities of response. For each participant, this is the predicted probability of

them taking part in the interview, given their individual characteristics and the characteristics of their household. Participants with characteristics associated with non-response receive a low predicted probability. The predicted probabilities were then used to generate a set of non-response weighting factors; a larger weighting factor was applied to participants with a low predicted probability, increasing their representation in the sample. The full non-response models for core England, Wales and Northern Ireland and for Scotland (core plus boost) are given in Tables B1 and B2.

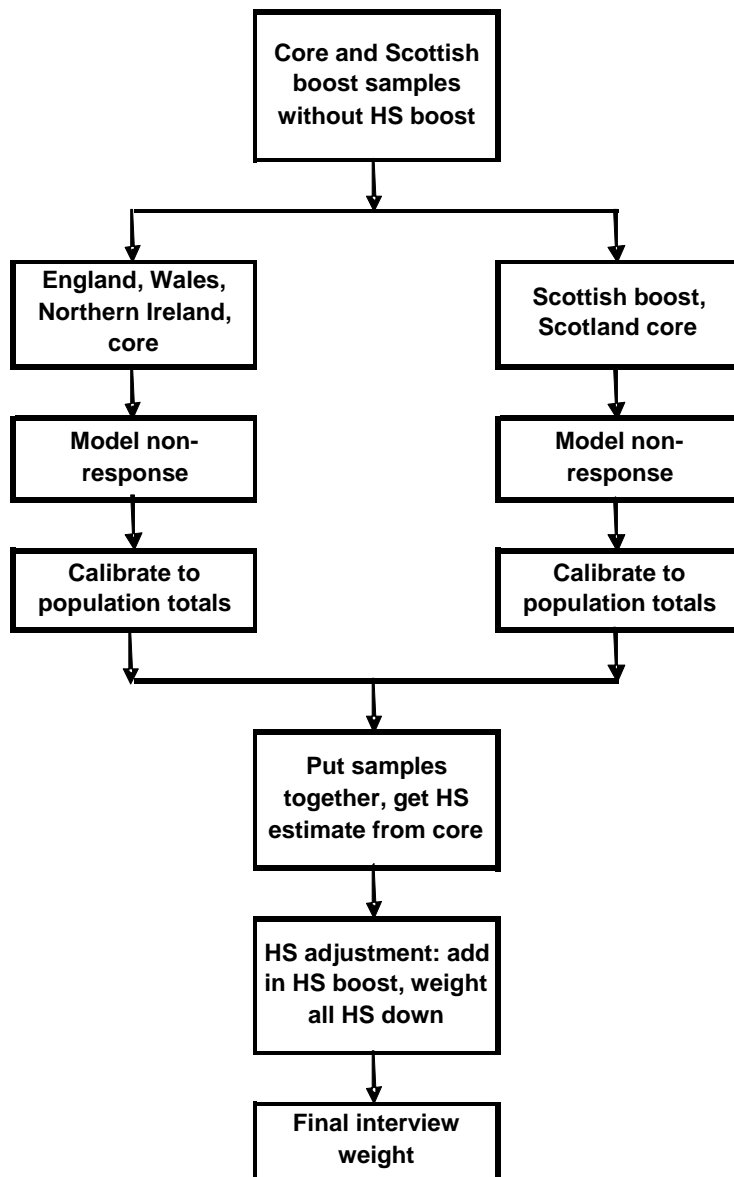
Tables B.1 and B.2

The next step was to calibrate the non-response weighting factors. Again, this was carried out separately for core England, Wales and Northern Ireland sample and for the combined Scottish core and boost samples. An iterative procedure was used to adjust the non-response weighting factor until the distribution of the (weighted) sample matched that of the population for a set of key variables. The adjustment keeps the values of the final weighting factors as close as possible to those of the initial weighting factors, which ensures the properties of the initial non-response weighting factors are retained in the final calibrated weights. This step made the profile of the sample match the population for child's age and gender, age of mother at the time of the child's birth and region. The population figures for calibration were taken from birth counts¹. The profiles of the population and the final weighted sample are shown in Table B3.

Table B.3

The calibrated weighting factors were combined into a single weighting factor. All core UK and Scottish boost cases had a calibrated non-response weighting factor. The final stage of the interview weighting factors was to incorporate the Healthy Start boost into the sample. The core UK and Scottish boost data (excluding the Healthy Start boost) were used to get a weighted estimate of the proportion of Healthy Start recipients. It was found that 20.6% of the weighted core and Scottish boost sample were in receipt of Healthy Start vouchers. The Healthy Start boost recipients were then added into the sample, increasing the proportion of Healthy Start recipients to 23.3%. All Healthy Start recipients were therefore weighted down by a factor of 20.6/23.3, so that the combined proportion of Healthy Start recipients matched the weighted estimate from the core data. This weighting factor was scaled, meaning the weighted sample size matched the unweighted sample size and the weights had a mean of one. This scaled weighting factor is the final interview weight. Figure B.1 gives an overview of this process.

Figure B.1: Overview of methods used for the interview weighting factor



References and endnotes

¹ This means the weights also account for any non-take up of Child Benefit, although it can be seen in Table B.3 that the distributions were very similar for each.

Table B.1

Non-response model for DNSIYC interview weighting factors – core England, Wales and NI

	B	S.E.	Wald	df	Sig.	Exp(B)
Wave of sampling	0.16	0.075	4.7	1	0.030	1.18
Recipient's title			14.9	3	0.002	
Miss					(baseline)	
Mr	-0.35	0.150	5.3	1	0.021	0.71
Mrs	0.17	0.091	3.5	1	0.062	1.19
Ms	0.12	0.199	0.3	1	0.561	1.12
Recipient's age in years at sampling			24.8	4	0.000	
<25					(baseline)	
25-29	0.40	0.113	12.6	1	0.000	1.49
30-34	0.50	0.119	17.4	1	0.000	1.64
35-39	0.62	0.135	21.0	1	0.000	1.86
40+	0.54	0.184	8.6	1	0.003	1.71
Government Office Region			26.0	10	0.004	
North East					(baseline)	
North West	-0.14	0.214	0.4	1	0.508	0.87
Yorkshire and The Humber	-0.02	0.226	0.0	1	0.937	0.98
East Midlands	-0.02	0.234	0.0	1	0.950	0.99
West Midlands	0.02	0.222	0.0	1	0.922	1.02
East of England	-0.16	0.222	0.5	1	0.462	0.85
London	-0.57	0.215	7.0	1	0.008	0.57
South East	-0.32	0.210	2.3	1	0.132	0.73
South West	-0.04	0.229	0.0	1	0.867	0.96
Northern Ireland	0.26	0.285	0.9	1	0.354	1.30
Wales	-0.07	0.257	0.1	1	0.794	0.94
Number of children aged 0-16 yrs in household			14.7	4	0.005	
1					(baseline)	
2	-0.04	0.087	0.2	1	0.682	0.97
3	-0.25	0.120	4.4	1	0.036	0.78
4	-0.40	0.186	4.6	1	0.031	0.67
5+	-0.79	0.279	8.0	1	0.005	0.45
Population density - quintiles			18.2	4	0.001	
1 - least dense					(baseline)	
2	0.08	0.131	0.3	1	0.556	1.08
3	-0.04	0.122	0.1	1	0.748	0.96
4	-0.28	0.117	5.9	1	0.016	0.75
5 - most dense	-0.42	0.139	8.9	1	0.003	0.66
Age in months at sampling			18.7	6	0.005	
4-5 mth					(baseline)	
6-7 mth	-0.35	0.175	4.1	1	0.043	0.70
8-9 mth	-0.36	0.173	4.4	1	0.036	0.70
10-11 mth	-0.23	0.173	1.7	1	0.194	0.80
12-13 mth	-0.53	0.174	9.4	1	0.002	0.59
14-15 mth	-0.27	0.175	2.5	1	0.116	0.76
16-17 mth	-0.66	0.190	11.9	1	0.001	0.52
Constant	0.71	0.257	7.6	1	0.006	2.03

The response is 1 = individual responded to the interview, 0 = non response

Only variables that are significant at the 0.05 level are included in the model

The model R2 is 0.048 (Cox and Snell)

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

The Wald-test measures the impact of the categorical variable on the model with the appropriate number of degrees of freedom df.

If the test is significant (sig. < 0.05) then the categorical variable is considered to be 'significantly associated' with the response variable and therefore included in the model

Table B.2

Non-response model for DNSIYC interview weighting factors – Scotland core and boost

	B	S.E.	Wald	df	Sig.	Exp(B)
Wave of sampling	-0.13	0.146	0.8	1	0.379	0.88
Recipient's title			4.2	3	0.241	
Miss					(baseline)	
Mr	-0.25	0.302	0.7	1	0.401	0.78
Mrs	0.24	0.172	1.9	1	0.166	1.27
Ms	-0.11	0.357	0.1	1	0.749	0.89
Recipient's age in years at sampling			20.3	4	0.000	
<25					(baseline)	
25-29	0.39	0.206	3.5	1	0.062	1.47
30-34	0.79	0.214	13.7	1	0.000	2.21
35-39	0.91	0.263	11.9	1	0.001	2.48
40+	1.27	0.382	11.0	1	0.001	3.56
Number of children aged 0-16 yrs in household			0.2	2	0.913	
1					(baseline)	
2	0.07	0.160	0.2	1	0.670	1.07
3+	0.04	0.219	0.0	1	0.874	1.04
Population density - quintiles			5.0	4	0.285	
1 - least dense					(baseline)	
2	0.46	0.230	4.0	1	0.045	1.58
3	0.09	0.187	0.2	1	0.635	1.09
4	-0.07	0.228	0.1	1	0.761	0.93
5 - most dense	0.00	0.231	0.0	1	0.987	1.00
Age in months at sampling			2.2	6	0.897	
4-5 mth					(baseline)	
6-7 mth	-0.28	0.325	0.8	1	0.384	0.75
8-9 mth	-0.37	0.322	1.3	1	0.253	0.69
10-11 mth	-0.41	0.325	1.6	1	0.205	0.66
12-13 mth	-0.23	0.322	0.5	1	0.485	0.80
14-15 mth	-0.26	0.321	0.6	1	0.424	0.77
16-17 mth	-0.41	0.364	1.3	1	0.261	0.67
Constant	0.25	0.323	0.6	1	0.447	1.28

The response is 1 = individual responded to the interview, 0 = non response

Only variables that are significant at the 0.05 level are included in the model

The model R2 is 0.049 (Cox and Snell)

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

The Wald-test measures the impact of the categorical variable on the model with the appropriate number of degrees of freedom df.

If the test is significant (sig. < 0.05) then the categorical variable is considered to be 'significantly associated' with the response variable and therefore included in the model

Table B.3

Distribution of the weighted interview sample, issued sample and population figures

	Population	Achieved sample1	Selected sample2
		Weighted by interview weighting factor	Scottish boost selection weighting factors only
	%	%	%
GOR			
North East	4	4	4
North West	11	11	11
Yorks and Humber	8	8	8
East Midlands	7	7	7
West Midlands	9	9	9
East of England	9	9	9
London	17	17	15
South East	13	13	14
South West	8	8	8
Northern Ireland	3	3	4
Scotland	7	7	8
Wales	4	4	5
IMD quintiles (from sampling frame)			
1 - least deprived	20	21	22
2	20	18	19
3	20	18	18
4	20	25	23
5 - most deprived	20	18	18
Population density quintiles (from sampling frame)			
1 - least dense	20	23	23
2	20	16	16
3	20	20	20
4	20	21	21
5 - most dense	20	21	20
Recipient's age - grouped (HMRC population counts)			
<25	22	22	22
25-29	25	25	25
30-34	28	28	28
35-39	18	18	19
40+	6	6	7
Recipient's gender (HMRC population counts)			
Male	10	11	10
Female	91	89	90
Mother's age at child's birth (birth records)			
<25	25	25	Not known for selected sample
25-29	28	28	
30-34	28	28	
35-39	16	16	
40+	4	4	

Child's gender (HMRC population counts)			
Male	51	51	50
Female	49	49	50
Child's age at sampling (HMRC population counts)			
0-2 months	16	16	16
3-4 months	16	17	17
5-6 months	16	16	17
7-8 months	17	17	18
9-10 months	18	17	16
11-12 months	17	17	17
Child's gender (birth records)			
Male	51	51	50
Female	49	49	50
Total number of children in household (HMRC population counts)			
1	49	51	50
2	33	32	32
3	13	12	13
4+	6	6	6
<i>Unweighted bases</i>	<i>748,480</i>	<i>2,586</i>	<i>4,276</i>

¹*Healthy Start boost not included, Scottish boost included but weighted down*

²*Healthy Start and Scottish boosts included*