21st International Workshop on Household Survey Non-response,
London (UK), 4th- 6th September 2013

David Hussey, Research Director (Statistician), NatCen Social Research,
david.hussey@natcen.ac.uk
Mark Hanly, University of Bristol.

Title: Non-response weighting: is it worth the effort?

Introduction

Survey response rates have been in steady decline over the past decades (Baruch and Holtom, 2008; Biener et al., 2004; de Leeuw and de Heer, 2002; Tolonen et al., 2006). As a result of this, non-response weighting is now an integral part of the survey process and is seen as a key tool in reducing non-response bias.

Non-response bias in survey estimates is caused by correlations between the response propensities of households (or individuals) and survey variables (Bethlehem, 2002). Non-response weights that adjust for differences in these response propensities should produce less biased survey estimates. However, this will only happen if the information used to create the weights (known as auxiliary variables) is available for both responding and non-responding households/individuals and is correlated with both response propensity and the survey variables (Beaumont, 2005; Bethlehem, 2002; Groves, 2006). Unless both these conditions hold, the adjustment will not impact on bias.

Typically, survey weighting comprises several components including selection weights, non-response adjustments for household and/or individual non-response and calibration to population estimates. The focus of this paper is on household non-response adjustments which often use auxiliary variables measured at area level, such as urbanity, geo-demographic indicators such as ACORN/OAC and interviewer observations about the area. These are assumed to be related to survey outcomes but published technical reports rarely provide any evidence of this.

Using data from two major household surveys, this paper investigates the effect of omitting the household non-response adjustment from the weighting process what impact this has on the final weighted survey estimates. In doing so we attempt to find out whether there is any evidence of a reduction in bias as a result of weighting for household non-response and, crucially, whether this reduction in bias a worthwhile trade-off for larger standard errors.

Data used

The analysis used data from two household surveys: British Social Attitudes (BSA) and the National Travel Survey (NTS).

BSA is a general population survey set up to measure the changing attitudes, values and beliefs of the British public. The survey is run annually by NatCen and is funded by a range of charitable and government sources. Data from BSA 2012 was used for analysis; this survey had just over 3,200 respondents (one per household).

The NTS is primarily designed to track the long-term development of personal travel trends in Great Britain. It is part of a continuous survey that began in July 1988, following ad hoc surveys since the mid-1960s. Data from NTS 2012 was used for analysis; this survey had just under 9,000 fully responding households and data for around 21,250 individuals.

Details of the sample design and current weighting schemes for both surveys can be found in the relevant technical reports (Taylor et al., 2012; Pickering et al., 2005; Park et al., 2012).

Aims

For each survey, analysis was undertaken to answer the following questions:

1. By how much does the inclusion of weighting for household non-response increase the variance of the weights?
2. Are the auxiliary variables correlated with key survey outcomes?
3. Does the distribution of the auxiliary variables in the final weighted sample reflect that found in the population? (*) And what does this distribution look like if the household non-response weighting is omitted?
4. Does the household non-response weighting make a difference to estimates of key survey outcomes?

(*) given that the non-response weights are trimmed and calibrated to population totals to create the final interview weights.

Methods

For each survey the weighting process was repeated excluding the household non-response adjustments. In both cases this meant that the selection weights were input directly into the final calibration step.

The following analyses, each corresponding to the questions above, were then conducted for each survey:

1. The variance of the revised weights (without the household non-response adjustment) was compared with the variance of the regular interview weights.
2. The strength of the relationship between the auxiliary variables and a set of key survey outcomes was examined using chi-squared tests on un-weighted data.
3. The distribution of the auxiliary variables (those used in the household non-response adjustment) was examined for responding households weighted by the two sets of weights. These distributions were compared with quasi-population estimates.
4. Estimates created by weighting by each of the two sets of weights were compared for the same set of key survey outcomes.

Results

1. Variance of weights

As the table shows, the inclusion of non-response weighting increases the variance of the weights for both surveys. However, the increase is relatively small, particularly for BSA where it is almost negligible. Non-response weighting therefore has only a marginal impact on the design effect of the weights in each case as evidenced by the DEFFs in the final column.

Table: Descriptive statistics for weights

	N	Mean	Std. Dev	Min	Max	DEFF
BSA final weight	3248	1	0.59	0.24	6.19	1.34
BSA revised final weight	3248	1	0.58	0.29	4.82	1.33
NTS interview weight (scaled)	8972	1	0.21	0.79	5.32	1.05
NTS revised interview weight (scaled)	8972	1	0.16	0.99	3.96	1.03

2. Relationship between auxiliary variables and key survey outcomes

Appendix A shows the results of (un-weighted) chi-squared tests of the relationship between the auxiliary variables and a set of key survey outcomes for each survey.

For BSA, three out of ten of the survey outcomes examined show no significant relationship with any of the four auxiliary variables. Only one survey outcome (proportion who "drive a car at all these days") is significantly related to all four auxiliary variables.

For NTS, all ten of the survey outcomes examined are significantly related to urbanity and ACORN. Survey month is related to four of the ten outcomes; as would be expected it is mostly related to time-variant questions.

3. Distribution of auxiliary variables

Appendix B shows the distributions of the auxiliary variables for each survey. Government Office Region (GOR) is not shown despite its inclusion in the household non-response models for both survey; this is because, in both cases, GOR is included in the subsequent calibration step and therefore, by definition, the final distribution will match the population.

For BSA, the distribution of the auxiliary variables was compared amongst: (a) respondents weighted by the pre-calibration weight (this provided quasi-population estimates for individuals); (b) respondents weighted by the regular interview weight; (c) respondents weighted by the revised weight. From comparison of (b) with (a) it is evident that the subsequent calibration step compromises the distribution of the auxiliary variables to a certain extent, although the largest difference when compared to the quasi-population estimates is only around 1% (for condition of area=good). However, omission of the household non-response adjustment would appear to result in greater bias in the auxiliary variables: a number of categories show differences of 1-2% when compared with the quasi-population estimates.

For NTS, the distribution of the auxiliary variables (and survey month) was compared amongst: (a) responding and non-responding households weighted by the selection weights (providing quasi-population estimates for these variables); (b) fully responding households weighted by the regular interview weight; (c) fully responding households weighted by the revised weight. The final distributions under the regular weighting scheme match the quasi-population estimates very closely. Furthermore, the estimates produced using the revised weights show some differences when compared to the quasi-population estimates on two out of the three variables, urbanity and for ACORN; the largest of these being a 2.6% gap between the proportion of households in Metropolitan areas in the weighted sample (29.2%) compared with that which we estimate for the population (31.8%). In summary, the non-response weighting would appear to be successful in reducing bias in the auxiliary variables. Moreover, unlike for BSA, it is not compromised by subsequent trimming or calibration steps.

4. Effect on weighted estimates

Appendix C shows the weighted estimates for ten key outcomes on each survey; in each case these are weighted using the regular weight and the revised weight for comparison.

For BSA most of the differences are fairly small. As might be expected, the largest differences are for the outcomes most strong related to the auxiliary variables; the largest (1.3%) being for proportion who "drive a car at all these days" which was found to be related to all four auxiliary variables.

For NTS, despite the strong relationship with the auxiliary variables, inclusion of the non-response adjustment makes little difference to the outcomes examined. Coincidentally perhaps, the largest difference (0.8%) is for households with regular access to a vehicle.

Discussion

Perhaps the most surprising result is the relatively small decrease in the variance of the weights as a result of omitting the household non-responses adjustments. This is obviously a positive finding as it means that inclusion of these adjustments will not inflate standard errors by much more than a negligible degree. However, this in itself is not enough to justify the inclusion of household non-response weighting. Ideally there should be evidence of a decrease in bias: as discussed we should see a clear correlation between the auxiliary variables used in the non-response adjustment on the one hand, and both response propensity and survey outcomes on the other.

The weighting process for both surveys uses logistic regression modeling to model the relationship between the auxiliary variables and household response. For BSA, only variables that are significantly related to response are included in the non-response adjustment. For NTS, the same variables are used each year but in 2012 all variables in the non-response model were highly significant. Therefore the key issue is the extent to which the auxiliary variables are related to survey outcomes. Ten outcomes were examined for each survey. The results for BSA were patchy; some survey outcomes were unrelated to any of the four auxiliary variables. For NTS, the results were more conclusive: both of the auxiliary variables (urbanity and ACORN) showed strong relationships with all ten survey outcomes. This is perhaps unsurprising: the NTS measures are high quality verified external measures whilst the BSA measures are, by contrast, low quality internal measures (all derived from interviewer observations).

In summary then, for NTS there would be appear to be a clear justification to weight for household non-response as there is clear evidence of significant relationship between the auxiliary variables on the one hand, and both response propensity and survey outcomes on the other. For BSA, on the other hand, the evidence is weaker as some survey outcomes showed no relationship with the auxiliary variables.

The third part of the analysis looked at whether the final weighted sample matched (estimates of) the population for distributions of the auxiliary variables. If no trimming or subsequent calibration took place then this would not be necessary as by definition the distributions would match very closely (the objective of the non-response adjustment being to create this correspondence). However, both surveys are calibrated to population estimates of age, sex and region which means that there exists some potential for the calibration to “upset” or compromise the distributions established by the non-response weighting.

For NTS, the subsequent calibration is at the household level and because only fully responding households (i.e. those where all individuals respond) are included, no further adjustment is required to produce estimates for individuals, in particular no person selection weight is required. For BSA on the other hand, there is only one respondent per household and therefore, the household non-response weight is multiplied by a selection weight which is a combination of a dwelling unit selection weight and a person level selection weight. This composite weight is then calibrated to population estimates for all individuals. This means there is greater potential for the calibration to upset the distributions but also that this is more difficult to assess.

For NTS, it was possible to produce quasi-population estimates for the auxiliary variables by taking the responding and non-responding households (i.e. all households after removing deadwood) and weighting them by the household selection weights. The final household level distributions could then be compared to this set of estimates. The analysis showed a good correspondence between the weighted sample and the quasi-population estimates when using the regular household weight. Furthermore, this correspondence was not as close when the household non-response adjustments were omitted. Overall then, not only was there good reason in principle for including the non-response adjustments, they clearly worked in practice and were not compromised by the subsequent calibration.

For BSA, it was not possible to produce quasi-population estimates for the auxiliary variables to compare with the weighted sample, at least not in the same way. This is because the final weighted sample is weighted by a person selection weight and calibrated to individual level population estimates and in order to produce equivalent estimates from the non-deadwood sample, we would need to know the number of persons in each responding *and* non-responding household. Unfortunately this figure is unknown for non-contacts and therefore we cannot produce quasi-population estimates in the same way. Instead, the distributions for respondents weighted by pre-calibration weight were used. This is not ideal but it did at least allow an assessment of the effect of the final calibration step.

What was clear from this comparison was that, unlike NTS, the final calibration does "upset" the distributions for the auxiliary variables to a certain extent. On the other hand, the distributions were closer to the quasi-population estimates when the final sample was weighted by the regular weight (included the non-response adjustment) than it was when using the revised weight (excluding the non-response adjustment). So the BSA weighting is arguably "better" when the non-response adjustment is included as it brings the final sample closer into line with the population for the auxiliary variables. However, the weighting could arguably be improved by including the auxiliary variables in the final calibration (this would have the effect of leaving the distributions as they were prior to calibration).

The final strand of the analysis looked at the effect on some key survey estimates of including/excluding the non-response adjustment. For BSA, the difference between the weighted estimates was less than 1% for nine out of ten of the outcomes examined. For NTS, this figure was ten out of ten. It might be tempting to conclude that the non-response adjustments are fairly inconsequential on both surveys but it worth considering some contextual issues in addition to the discussion above.

For BSA, the justification for carrying out non-response adjustments was unclear as some of the auxiliary variables were unrelated to survey outcomes. Moreover, the final weighted sample did not match quasi-population estimates for the auxiliary variables so even when these were related to survey outcomes, the non-response adjustment arguably did not do its job very well. Furthermore, it is worth considering the subjective nature of the survey questions as well as the relatively small sample size (compared to the NTS) which produces confidence intervals of +/-2-3% when stratification and clustering are taken into account in addition to the weights. Taking all these things together, the non-response adjustments are arguably of limited use.

For NTS on the other hand, there was clear justification for carrying out the non-response adjustments as the auxiliary variables were strongly related to the survey outcomes. Moreover, the final weighted sample matched quasi-population estimates for the auxiliary variables closely, therefore the non-response adjustments achieved their objective. As with BSA it is also worth considering the subject matter of the survey which is fairly factual/objective, plus the relatively large sample size, which produces confidence intervals of only around +/-1% when stratification and clustering are taken into account in addition to the weights. So despite the relative small changes observed as a result of the non-response adjustment, when taken in context they would appear to be of much greater value.

Conclusion

The original question posed by this paper asked whether (household) non-response weighting is worth the effort. The answer for the NTS, despite the small difference it makes to estimates, is yes. This is largely due to the fact that the NTS (unlike BSA) uses highly quality external measures as auxiliary variables. The answer for BSA on the other hand is much less clear; moreover, in this case there would appear to be some scope for improving the weighting process.

Appendix A: Relationship between auxiliary variables and key survey outcomes

BSA: Results of chi-squared tests between auxiliary variables and key survey outcomes

	Response	**Barriers to entry**	**Condition of area**	**Condition of address**	**Dwelling type**
Do you think of yourself as a supporter of any one political party?	*Yes*	**-**	**YY**	**-**	**YY**
How much interest do you generally have in what is going on in politics?	*A great deal / quite a lot*	Y	YY	**-**	**YY**
Would you say that most people can be trusted, or that you can t be too careful in dealing with people?	*Most people can be trusted*	YY	YY	**-**	**Y**
How much do you trust British govts of any party to place the needs of the nation above the interests of their own party?	*Only some of the time*	Y	-	**-**	**-**
How much do you trust politicians of any party in Britain to tell the truth when they are in a tight corner?	*Only some of the time*	-	-	**-**	**-**
If govt had to choose which should it choose: to cut taxes or spending	*Keep both at same level*	-	-	**YY**	**YY**
How satisfied or dissatisfied would you say you are with the way in which the National Health Service runs nowadays?	*Very satisfied /quite satisfied*	-	-	**Y**	**-**
Do you yourself drive a car at all these days?	*Yes*	YY	YY	**YY**	**YY**
Who should pay the cost of tuition fees for higher education?	*Some [not all] students or their families*	-	-	**-**	**-**
What about sexual relations between two adults of the same sex? What would your general opinion be?	*Not wrong at all*	-	-	**-**	**-**

YY = p<0.01; Y= 0.01<p<0.05

NTS: Results of chi-squared tests between auxiliary variables and key survey outcomes

	Urbanity	**ACORN**	**Month**
Households within 6mins walk of bus stop	**YY**	**YY**	**-**
Households with regular access to any vehicle	**YY**	**YY**	**YY**
Households within 15 minutes travel time to GP	**YY**	**YY**	**-**
Households with a bluebadge holder (disabled parking)	**YY**	**YY**	**-**
Persons holding a driving license	**YY**	**YY**	**YY**
Persons travelling to work by public transport	**YY**	**YY**	**YY**
Persons reporting disability affecting travel	**YY**	**YY**	**-**
Persons reporting any road accident in last 3 years	**YY**	**YY**	**-**
Persons holding a season ticket	**YY**	**YY**	**-**
Persons working from home at least once or twice a month	**YY**	**YY**	**YY**

YY = p<0.01; Y= 0.01<p<0.05

Appendix B: Distribution of auxiliary variables

BSA: distribution of auxiliary variables

		(a) Respondents weighted by pre-calibration weight	(b) Respondents weighted by final weight	(c) Respondents weighted by revised weight	Bias using regular weight (b-a)	Bias using revised weight (c-a)
Variable		%	%	%	%	%
Barriers to entry	No barriers	89.2	88.8	90.6	-0.4	+1.4
	One or more barriers to entry	10.8	11.2	9.4	+0.4	-1.4
Condition of area	Mainly good	45.8	44.7	47.7	-1.1	+1.9
	Mainly fair	50.7	51.6	49.0	+1.0	-1.7
	Mainly bad or very bad	3.5	3.7	3.3	+0.2	-0.2
Condition of add	Better	7.7	7.6	9.0	-0.1	+1.4
	About the same	85.6	85.7	84.6	+0.1	-1.0
	Worse	6.7	6.7	6.4	+0.0	-0.3
Dwelling type	Detached house	23.5	22.7	23.9	-0.8	+0.4
	Semi-detached house	33.9	33.2	32.5	-0.7	-1.4
	Terraced house	27.9	28.7	29.3	+0.9	+1.4
	Flat or maisonette purpose built	14.7	15.4	14.3	+0.7	-0.4

NTS: distribution of auxiliary variables

		(a) Responding & non-responding households weighted by selection weight	(b) Fully responding households weighted by regular weight	(c) Fully responding households weighted by revised weight	Bias using regular weight (b-a)	Bias using revised weight (c-a)
Variable		%	%	%	%	%
Urbanity	Metropolitan areas	31.8	30.7	29.2	-1.1	-2.6
	Urban (> 250k people)	13.1	13.4	13.1	+0.3	+0.0
	Urban (25-250k people)	24.9	25.3	25.2	+0.4	+0.3
	Urban (10-25k people)	9.1	9.3	9.7	+0.2	+0.6
	Urban (3-10k people)	6.5	6.6	6.9	+0.1	+0.4
	Rural (< 3k people)	14.6	14.7	15.8	+0.1	+1.3
ACORN	Wealthy Achievers	22.7	22.6	24.4	-0.1	+1.7
	Urban Prosperity	13.2	13.0	11.9	-0.2	-1.2
	Comfortably Off	28.0	28.1	28.3	+0.2	+0.4
	Moderate Means	14.3	14.5	14.0	+0.2	-0.3
	Hard-Pressed	21.8	21.7	21.4	-0.1	-0.5
Month *	January	8.3	8.1	8.1	-0.2	-0.3
	February	8.3	8.5	8.2	+0.2	-0.1
	March	8.3	8.4	8.4	+0.1	+0.1
	April	8.3	8.7	8.6	+0.3	+0.3
	May	8.3	8.3	8.1	-0.0	-0.3
	June	8.3	8.3	8.1	-0.1	-0.3
	July	8.3	8.1	8.1	-0.2	-0.2
	August	8.3	8.3	8.1	-0.0	-0.2
	September	8.3	8.2	8.8	-0.1	+0.5
	October	8.3	8.4	8.4	+0.1	+0.1
	November	8.3	8.4	8.9	+0.0	+0.5
	December	8.3	8.3	8.2	-0.0	-0.1

*** for month, the concept of quasi-population estimates doesn't apply so a comparison is made with the desired distribution i.e. an even distribution throughout the year.**

Appendix C: Comparison of weighted survey estimates

BSA: Key outcomes weighted by two sets of weights

Question	Response	Final weight	Revised final weight	Difference
		%	%	%
Do you think of yourself as a supporter of any one political party?	*Yes*	33.5	33.5	+0.1
How much interest do you generally have in what is going on in politics?	*A great deal / quite a lot*	35.6	36.3	+0.7
Would you say that most people can be trusted, or that you can t be too careful in dealing with people?	*Most people can be trusted*	38.4	39.3	+0.8
How much do you trust British govts of any party to place the needs of the nation above the interests of their own party?	*Only some of the time*	48.7	48.5	-0.2
How much do you trust politicians of any party in Britain to tell the truth when they are in a tight corner?	*Only some of the time*	39.9	40.0	+0.1
If govt had to choose which should it choose: to cut taxes or spending	*Keep both at same level*	53.5	53.3	-0.1
How satisfied or dissatisfied would you say you are with the way in which the National Health Service runs nowadays?	*Very satisfied /quite satisfied*	60.6	60.8	+0.2
Do you yourself drive a car at all these days?	*Yes*	69.7	71.0	+1.3
Who should pay the cost of tuition fees for higher education?	*Some [not all] students or their families*	67.5	67.6	+0.1
What about sexual relations between two adults of the same sex? What would your general opinion be?	*Not wrong at all*	46.6	47.0	+0.4

NTS: Key outcomes weighted by two sets of weights

	Regular interview weight	Revised interview weight	Difference
	%	%	%
Households within 6mins walk of bus stop	85.4	85.0	-0.3
Households with regular access to any vehicle	74.9	75.8	+0.8
Households within 15 minutes travel time to GP	80.1	79.5	-0.6
Households with a bluebadge holder (disabled parking)	9.7	9.7	+0.0
Persons holding a driving license	71.6	72.2	+0.5
Persons travelling to work by public transport	16.3	15.9	-0.4
Persons reporting disability affecting travel	11.8	11.7	-0.1
Persons reporting any road accident in last 3 years	11.2	11.3	+0.1
Persons holding a season ticket	30.2	30.0	-0.2
Persons working from home at least once or twice a month	12.1	12.2	+0.1

References

Baruch, Y. and Holtom, B.C. (2008). Survey response rate levels and trends in organizational research. *Human Relations*, 61(8), 1139-1160.

Beaumont, J-F. (2005). On the use of data collection process information for the treatment of unit nonresponse through weight adjustment. *Survey Methodology*, 31(2), 227-231.

Bethlehem, J. (2002). Weighting nonresponse adjustments based on auxiliary information. In *Survey Nonresponse* (eds R. Groves, D. Dillman, J. Eltinge and R. Little), pp. 275–287. New York: Wiley.

Biener, L., Garrett, C.A., Gilpin, E.A., Roman, A.M., and Currivan, D.B. (2004). Consequences of Declining Survey Response Rates for Smoking Prevalence Estimates. *American Journal of Preventative Medicine*, 27(3), 254-257.

de Leeuw, E., and de Heer, W. (2002). Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. Chapter 3 in Groves, R.M. et al., *Survey Nonresponse*. Wiley.

Groves, R.M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, 70(5), 646-675.

Park A. et al (2012) British Social Attitudes 29, Anxiety Britain: Worries on cuts and public services present big challenges for Cameron's Coalition http://www.bsa-29.natcen.ac.uk/read-the-report/technical-details/weighting.aspx

Pickering K., Tipping S. and Scholes S. (2005) Weighting the National Travel Survey: Methodology Final Report. Published on the Department for Transport website: http://www.dft.gov.uk/stellent/groups/dft_transstats/documents/downloadable/dft_transstats_041302.pdf

Taylor E., Humphrey A., Pickering K. and Tipping S. (2012) National Travel Survey 2012: Technical Report. Published on the Department for Transport website: Published on the Department for Transport website: https://www.gov.uk/government/publications/national-travel-survey-2012

Tolonen, H., Helakorpi, S., Talala, K., Helasoja, V., Martelin, T., and Prattala, R. (2006). 25-year trends and socio-demographic differences in response rates: Finnish adult health behaviour survey. *European Journal of Epidemiology*, 21, 409-415.