Dadansoddi ar gyfer Polisi

YMCHWIL GYMDEITHASOL Y LLYWODRAETH

GOVERNMENT SOCIAL RESEARCH

Analysis for Policy

Llywodraeth Cymru
Welsh Government

Ymchwil gymdeithasol
Social research

www.cymru.gov.uk

Number: 36/2013

An assessment of the utility of new interviewer observation variables for non-response weighting on the National Survey for Wales

June 2013

A report prepared for the Welsh Government by

Professor Patrick Sturgis

Professor of Research Methodology and Director of the ESRC National Centre for Research Methods.

Views expressed in this report are those of the researcher and not necessarily those of the Welsh Government.

For further information please contact:

Chris McGowan
Knowledge and Analytical Services
Welsh Government
Cathays Park
Cardiff
CF10 3NQ
Tel: 029 2080 1067
Email: chris.mcgowan@wales.gsi.gov.uk

Welsh Government Social Research, 2013
ISBN 978-0-7504-9627-8

Contents

Executive summary

Given the substantial cost of household surveys and their centrality to understanding the attitudes and behaviours of populations for the purpose of policy-making and evaluation, it is essential that every effort is made to ensure that the estimates produced from them are accurate.

The objective of this small research project was to evaluate whether a more strategic approach to the selection of interviewer observation variables might result in more powerful weighting adjustments for the National Survey for Wales. While most surveys select interviewer observations in order to maximize the prediction of response propensity, the objective here was to also maximize the prediction of specific survey outcomes. If weighting variables are correlated with both response propensity and with survey outcomes, they become more effective in reducing bias in survey estimates.

Three new variables were selected from a list of candidate measures and included in the survey between April and June 2012. These required interviewers to systematically assess how safe they would feel in the area at night and how much vandalism, litter and graffiti was visible in the local area. It was hypothesized that these measures would be predictive of responses to the questions in the National Survey which ask respondents to rate various different aspects of their own local areas.

A number of analyses were undertaken which assessed the predictive power of the new observation variables with respect to both response propensity and survey outcomes. These estimates were compared to equivalent ones derived from the standard interviewer observation variables included in the survey from January to March 2012 and to an area classification derived from 2001 Census and administrative data.

The results showed that the new variables were significant predictors of response propensity whether considered on their own, or alongside the standard observation variables or the area classification variable. However, the strength of the correlation was weak, with only approximately 1% of the variability in response propensity explained by the new variables. This did not represent an improvement over the existing interviewer observation variables.

Although the new observation variables were stronger predictors of survey outcomes than the standard variables, this was only by a small amount. And the explanatory strength of the new observation variables was marginally lower than for the area classification variable. This was even the case for survey outcomes that were obviously and directly related to the observation variables.

Non-response weights were derived using the new interviewer observation variables. Application of these new weights for univariate estimates across a range of survey outcomes had very little effect compared to the unweighted estimates. By way of contrast, application of a standard post-stratification weight had a considerably greater impact on estimates these survey outcomes.

The overall conclusion of the report is that the resources currently being expended in the collection of interviewer observations could be more cost-effectively deployed to a different area of fieldwork activity.

Introduction

A serious threat to the accuracy of survey estimates is that non-respondents might be different to respondents on the substantive variables of interest in the survey. Because non-response constitutes a large (and growing) proportion of the issued sample for the vast majority of household surveys, the potential for non-response bias can be substantial. The direction and magnitude of non-response bias is, however, generally unknown because, by definition, the substantive variables of interest are observed only for the respondents. This is a vexing issue for analysts who rely on survey data for the development of academic theory and evidence-based policy alike.

The primary means by which survey analysts seek to counteract the biasing effect of non-response is through the application of weighting adjustments. Non-response weighting can reduce bias in survey estimates by increasing the contribution to the estimate of under-represented groups in the sample by a factor which is proportional to the inverse of the probability of responding.

A basic requirement of variables that are to be used in the derivation of non-response weights is, therefore, that they must be observed for both respondents and non-respondents. This is a demanding requirement which only relatively few variables meet in practice. Variables which do meet this criterion can be divided into two broad categories:

1. Variables for which the population distribution is available from another source (such as the Census, another large survey, or administrative data) and which are also measured in the survey.

2. Variables which are observed for all issued cases in the sample, both respondents and non-respondents.

Variables of type 1, typically age, sex, and housing tenure, can be used to produce what are referred to as post-stratification weights. Post-stratification weights ensure that the sample matches the population according to these known distributions and can be effective in reducing bias as well as increasing the precision of estimates.

Variables of type 2 can be used to identify sub-groups which are under-represented in the sample and to up-weight these groups by a factor equal to the inverse of their probability of selection in the sample. It is with variables of type 2 that this report is concerned.

In the UK, variables of type 2, which are observed for both respondents and non-respondents in the issued sample, are generally not available at the individual or household level because the UK does not maintain a population register. Thus, these variables are predominantly defined as area-level characteristics which are derived as aggregations at different levels of geography, from sources such as the decennial Census or from administrative data.

Because of their aggregate nature and also because they tend to be measured rather a long time before the survey takes place, such variables are generally only weakly predictive of individual and household level outcomes. For this reason, survey designers have sought to obtain additional leverage on non-response bias by requiring interviewers to record systematic observations about all addresses in their issued sample. Because these types of observations are quite straightforward and inexpensive to record as a routine part of fieldwork, they represent a potentially cost-effective way of mitigating non-response bias through weighting adjustment. Over the past ten years or so, a large number of different observation variables have been included on many different household surveys in the UK.

However, in the companion publication to this report, 'An assessment of the potential utility of interviewer observation variables for reducing non-response error in the National Survey for Wales' (Sturgis and Brunton-Smith, 2012)[1], a key conclusion was that the standard interviewer observation variables employed in most UK and international surveys are unlikely to be of notable utility in reducing survey error through non-response weighting. This is because, while they are generally (weakly) correlated with response propensity, they are hardly correlated at all with most survey outcomes. This conclusion is consistent with other recent investigations of interviewer observation variables, notably Kreuter et al (2010).

In the Sturgis and Brunton-Smith (2012) report, it was recommended that the National Survey for Wales take the opportunity to innovate in the kinds of observations that interviewers are asked to make, with the particular intention of developing measures which have a stronger association with key variables in the questionnaire. As a preliminary stage of this research, a review of candidate variables was undertaken and the following questions were identified for inclusion in the survey from April 2012:

- OBS5 - In the immediate area, how common is litter or rubbish lying around? *Very common, fairly common, not very common, not at all common.*

- OBS6 – How common is vandalism, graffiti or deliberate damage to property? *Very common, fairly common, not very common, not at all common*.

- OBS7 - How safe or unsafe would you feel if you were walking in this area after dark? *Very safe, fairly safe, fairly unsafe, very unsafe*.

Because (as with most surveys) the National Survey for Wales contains questions on a wide variety of topics, any given observation variable will differ in the extent to which it is correlated with any specific survey variable. In determining a variable selection strategy, therefore, one can either seek to identify observation variables which are correlated with as many survey

[1] This report can be downloaded here: www.wales.gov.uk/NationalSurvey/Non-response research

variables as possible, or to maximize the correlation with a specific survey variable or subset of variables. Because the primary objective of this research is to evaluate whether interviewer observations with stronger correlations with survey variables can be more effective than those which correlate with response propensity, the latter strategy was pursued here. That is to say, the new observation variables were selected on the grounds of having potentially high correlations with specific survey variables.

The survey variables with which these questions are expected to be correlated are those relating to the respondent's satisfaction with their local area. In particular, these observations should be anticipated to have high correlations with the survey questions which ask respondents to rate how safe they feel in their area after dark and how much litter, graffiti and vandalism there is in the area. Given the specific focus of the National Survey on perceptions of and satisfaction with local areas, the potential for bias in these variables is of particular concern.

In addition to these three new items, four standard TNS-BMRB interviewer observation variables were also included. These were:

- OBS1 - Which of these best describes the condition of residential properties in the area? *Mainly good, mainly fair, mainly bad, mainly very bad*.
- OBS2 - Is the outside of the sampled house/flat in a better or worse condition than the others in this area? *Better, worse, about the same*.
- OBS3 - Type of house (*e.g. detached, semi-detached*)
- OBS4 - Are any of these physical barriers to entry present at the house/flat? (*locked gates, entry phone* etc.) *Yes/No*[2]

[2] The answer categories for this question include a full list of different barriers but in the analysis a yes/no dichotomy is used.

Data and analysis strategy

Data for this analysis come from the quarter 1 (April – June 2012) issued sample of the National Survey for Wales. Although the interviewer observation variables were also included in quarter 2, July – September 2012, there were a large number of cases from quarter 2 without final disposition codes at the time this analysis was conducted. Because the probability of having a final disposition code was associated with the interviewer observation variables, it was decided not to include cases from quarter 2 in the analysis.

The issued sample file for quarter 1 contained 5899 cases. For 305 of these, the interviewer observation variables were coded 'data not available' due to an error in retrieving the data from the electronic contact sheets used by interviewers. These cases were, therefore, deleted from the analysis. As they represent only 5% of the total number of cases, it is unlikely that their omission will have any notable effect on the results presented here. Of the remaining 5594 cases, 569 were deadwood, yielding an analysis sample of 5025 for the response propensity models. Of these 5025 cases, 3541 (70.5%) were successful interviews, 295 (5.9%) were non-contacts, 933 (18.6%) were refusals, and 256 (5.1%) were 'other unproductive'. Models predicting survey outcomes are, therefore, based on the responding sample of 3541 households.

The analysis proceeded in three stages. In stage 1, response propensity models were fitted to the issued sample data, using different sets of predictor variables. In order for the new observation variables to be of any use in the derivation of weights, they must be predictive of whether households respond to the survey request. These models therefore determine whether the new variables meet the minimal requirement for use in the derivation of non-response weights.

As well as a response propensity model containing only the new observation variables, models were also fitted which included the new variables alongside a) the standard TNS-BMRB observation variables and b) the variables that were used by TNS-BMRB to derive the non-response weight for the quarter 0[3] and quarter 1 data. This is a categorical variable (LSOAcode) at the lower super-output areas (LSOAs)[4] level which allocates LSOAs to one of 15 mutually exclusive categories, as a function of their similarity on a range of Census and administrative variables. The purpose of including these models was to determine whether the new observation variables have any predictive utility over and above that which is already available through the existing observational and auxiliary data.

[3] The annual cycle of the National Survey runs from the beginning of April to the end of March the following year. This means that Quarter 1 runs from April to June. Because the first quarter in year 1 of the survey covered January to March, this is referred to as Quarter 0.

[4] LSOA are spatial units comprised of aggregations of census output areas (Martin, 2008).

In stage 2, models were fitted to a range of outcome variables from the survey. Again, comparisons are made between the predictive power of the new observation variables relative to the standard TNS-BMRB variables and the LSOAcode variable used to produce the non-response weight for quarters 0 and 1. Models are fitted for 19 different outcomes: 13 relating to attitudes toward the local area; 5 relating to subjective well-being; and one measure of self-rated health. The a priori expectation is that the new observation variables should have the greatest explanatory power for the local area variables, the remaining variables are included to provide a contrast for this expectation.

The third stage of the analysis was to compare weighted to unweighted estimates using three different weights. The first weight (Weight1) was produced using only the LSOAcode variable. It was derived by taking the inverse of the case-level predicted probability from a logistic regression model, where the dependent variable was a binary indicator of whether the household responded or not and the predictor was the LSOAcode variable.

The second weight (Weight2) was derived in the same manner but replacing the LSOAcode variable with the three new observation variables.

The third weight (Weight3) was the scaled respondent calibration weight produced by TNS-BMRB. It combines the design weight (which corrects for different selection probabilities due to multiple households at addresses and multiple adults within households), the household non-response weight based on the LSOAcode variable and a post-stratification weight, which adjusts the sample distribution for age, sex and housing tenure status to match the 2011 mid-year Census estimate projected forward to 2012.

Results

Table 1 shows the coefficient estimates, standard errors and associated p values, and odds ratios for the model predicting response propensity using the three new interviewer observation variables only. Two of the variables, obs5 and obs7 are significant predictors ($p<0.05$) of response propensity while the third, obs6, is marginally non-significant when included alongside the other two variables (it is significant when included on its own).

For the variable obs5, households at which the interviewer rates the amount of litter in the surrounding area as being ‘not common at all’ have a 56% greater odds of responding compared to the reference category (‘litter fairly common’). For the variable obs7, none of the substantive categories differ in the odds of responding but, for households where the interviewer records that he or she was unable to provide a response to this question, the odds of responding are 73% higher than for the reference category (‘not very safe after dark’).

Although the new observation variables are significantly related to response propensity, the predictive power of the model is weak. The pseudo R-squared is just 0.009 which means that these variables account for less than 1% of the variability in response propensity across households.

Table 1 Response Propensity model, New Observation Variables

Predictor	Category	Beta	S.E.	p value	Odds ratio
obs5	litter very common	0.130	0.305	0.669	1.139
	litter not very common	0.152	0.119	0.200	1.164
	litter not at all common	0.448	0.132	0.001	1.565
	unable to code	-0.367	1.003	0.715	0.693
obs6	vandalism common	-0.820	0.566	0.148	0.440
	vandalism fairly common	0.219	0.215	0.310	1.244
	vandalism not common at all	-0.166	0.093	0.075	0.847
	unable to code	-0.120	0.951	0.900	0.887
obs7	very safe after dark	0.132	0.134	0.327	1.141
	fairly safe after dark	0.047	0.128	0.710	1.049
	very unsafe after dark	0.203	0.282	0.471	1.225
	unable to code	0.551	0.250	0.028	1.734
	Constant	0.603	0.141	0.000	1.827

Source = National Survey for Wales, quarter 1; n=5025; Nagelkerke R squared=0.009

Table 2 presents the estimates for the model with the standard TNS-BMRB observation variables included alongside the new variables. All but one of the TNS-BMRB variables (obs1) are significantly related to the propensity to respond. The explanatory power of the model, at 0.022, is still weak in absolute terms but is more than double the magnitude of the model containing the new variables on their own. However, it is also evident that the new variables are able to account independently for some of the variability in response propensity because both the significant predictors from table 1 remain significant in table 2. Indeed, obs6 is also now significant, with households reported as being in areas where vandalism is not at all common having 17% lower odds of responding compared to areas where vandalism is reported to be very common.

Table 2 Response Propensity model, Old and New Observation Variables

Predictor	Category	Beta	S.E.	p value	Odds ratio
obs5	litter very common	0.101	0.306	0.741	1.107
	litter not very common	0.124	0.121	0.305	1.132
	litter not at all common	0.356	0.140	0.011	1.427
	unable to code	-0.251	1.084	0.817	0.778
obs6	vandalism common	-0.681	0.580	0.241	0.506
	vandalism fairly common	0.239	0.221	0.281	1.269
	vandalism not common at all	-0.189	0.094	0.045	0.828
	unable to code	-0.096	1.031	0.925	0.908
obs7	very safe after dark	0.103	0.138	0.456	1.108
	fairly safe after dark	0.063	0.129	0.627	1.065
	very unsafe after dark	0.250	0.289	0.387	1.284
	unable to code	0.649	0.256	0.011	1.914
obs1	mainly good	1.285	0.805	0.110	3.616
	mainly fair	1.178	0.806	0.144	3.247
	mainly bad	1.263	0.830	0.128	3.537
obs2	better	0.250	0.114	0.028	1.284
	worse	-0.055	0.125	0.663	0.947
	unable to code	1.187	0.793	0.134	3.276
obs3	detached	-0.043	0.087	0.621	0.958
	mid-terrace	-0.244	0.089	0.006	0.783
	end of terrace	-0.061	0.117	0.601	0.941
	maisonette	-0.294	0.419	0.483	0.745
	purpose-built flat	0.072	0.161	0.652	1.075
	flat-converted	-0.339	0.257	0.187	0.712
	unable to code	-1.192	0.521	0.022	0.304
obs4	no physical barriers to entry	0.526	0.143	0.000	1.691
	Constant	-0.982	0.833	0.239	0.375

Source = National Survey for Wales, quarter 1; n=5025; Nagelkerke R squared=0.022

Table 3 shows the parameter estimates for the model which includes the new interviewer observation variables alongside the LSOAcode variable. Only one of the LSOAcode categories (areas which are described as 'farming and forestry') is significantly related to response propensity. Of the new observation variables, obs5 and obs7 remain significant, though as with the model in table 1, obs6 is not significant. The pseudo R-squared of the model is again weak and only marginally superior to that observed for the model containing only the new observation variables as predictors.

In summary, the new observation variables are significant predictors of response propensity whether considered on their own, or alongside the standard observation variables, or the variable used by TNS-BMRB to produce the non-response weight for quarters 0 and 1. However, although their predictive power is statistically non-zero, its overall magnitude is weak – explaining only around 1% of the variability in the propensity to respond to the survey across households.

Table 3 Response Propensity model, New Observation and LSOA clusters

Predictor	Category	Beta	S.E.	p value	Odds ratio
obs5	litter very common	0.124	0.305	0.684	1.132
	litter not very common	0.136	0.120	0.256	1.146
	litter not at all common	0.400	0.135	0.003	1.491
	unable to code	-0.289	1.040	0.781	0.749
obs6	vandalism common	-0.848	0.570	0.137	0.428
	vandalism fairly common	0.200	0.217	0.357	1.221
	vandalism not common at all	-0.159	0.094	0.091	0.853
	unable to code	-0.268	0.987	0.786	0.765
obs7	very safe after dark	0.121	0.138	0.379	1.129
	fairly safe after dark	0.045	0.129	0.725	1.046
	very unsafe after dark	0.202	0.282	0.474	1.224
	unable to code	0.546	0.253	0.031	1.725
LSOAcode	Affluent Urban Commuter	0.034	0.293	0.909	1.034
	Blue Collar Urban Families	0.219	0.157	0.164	1.245
	Countryside Communities	0.347	0.181	0.056	1.414
	Educational Centres	-0.166	0.386	0.666	0.847
	Farming and Forestry	0.522	0.190	0.006	1.685
	Mature City Professionals/Suburbia	-0.043	0.375	0.908	0.958
	Mature Urban Households/Young City Professionals	0.034	0.160	0.834	1.034
	Multicultural/Struggling Urban Families	0.102	0.216	0.637	1.107
	Resorts and Retirement	0.208	0.218	0.341	1.231
	Rural Economies	0.106	0.168	0.529	1.112
	Small Town Communities	0.257	0.173	0.139	1.293
	Urban Commuter	0.066	0.189	0.725	1.069
	Urban Terracing	-0.051	0.200	0.798	0.950
	Well off Mature Households	0.321	0.197	0.102	1.379
	Constant	0.462	0.200	0.021	1.588

Source = National Survey for Wales, quarter 1; n=5025; Nagelkerke R squared=0.016

Next, we turn to the key question of how strongly the new variables are correlated with a range of survey outcomes. Table 4 shows the R-squared values for five different models fitted to each of the nineteen survey outcomes. R-squared is used here as a summary measure of the power of the prediction equation, full model parameter estimates for the 19*5=95 models are presented in Appendix 1 and Appendix 2.

Column 1 of table 4 shows that the mean R-squared across the 19 variables for the standard TNS-BMRB observation variables is 0.036. This compares with a corresponding figure of 0.045 for the new observation variables which shows that, for these survey outcomes, the explanatory power of the new observation variables is twenty five per cent greater than the standard measures. However, the average R-squared for the LSOAcode variables is actually slightly higher than the new observation variables, at 0.048. So, while the new variables represent an improvement compared to the existing interviewer observations, they do not improve on the predictive power of the variables which are currently used to produce the non-response weight for the National Survey.

Looking across the individual items, the expected pattern is generally confirmed. The largest R-squared values for the new variables are found for the local area attitude variables, with all but two models showing a higher magnitude compared to the standard observation variables. However, somewhat surprisingly, the explanatory power of the LSOAcode variable is actually higher for the local area attitude topic area on all but four of the questions. Even for the three survey questions of most direct relevance to the new observation variables – safety in local area after dark, graffiti in the local area and litter in the area, the LSOAcode variable produces a marginally higher R-squared value.

Combining the new variables with a) the standard observation variables and b) the LSOAcode variable in columns five and six respectively produces notable increases in the explanatory power of the model. In particular, the combination of the new variables with the LSOAcode measure results in a mean R-squared across the nineteen questions of close to 7%. This is still quite low in absolute terms but represents more than a 50% increase in explanatory power compared to any of the variables considered on their own.

In summary, the new observation variables offer an improvement in the magnitude of explanatory power on the survey outcomes considered, here relative to the standard TNS-BMRB measures included at quarter 0. Additionally, the pattern of the correlations across the different survey outcomes fits with the theoretical expectation that these new variables would be most strongly associated with the questions relating to the characteristics of the local area, which are also the subject of the new observation variables.

However, in absolute terms, the predictive strength of these variables is still quite weak, with an average R-squared of just 4.5% and lower than the LSOAcode variable, which is a geographical aggregation derived from administrative data and variables from the 2001 Census.

Table 4 Explanatory Power of Interviewer Observation and other Auxiliary variables on a range of Survey Outcomes

Outcome	Old variables	LSOAcode	New variables	Old + new variables	LSOAcode+new variables
Local area					
belonging to local area	0.018	0.031	0.015	0.025	0.038
people willing to help neighbours	0.038	0.047	0.042	0.054	0.065
safety at home after dark	0.039	0.045	0.044	0.059	0.063
safety walking in local area after dark	0.043	0.070	0.058	0.070	0.090
Local area - safety walking in nearest town/city centre after dark	0.016	0.026	0.016	0.025	0.037
safety travelling by public transport after dark	0.017	0.021	0.028	0.035	0.041
trusting people in the neighbourhood	0.073	0.090	0.052	0.085	0.102
well maintained	0.030	0.050	0.054	0.060	0.079
free from litter and rubbish	0.053	0.066	0.067	0.084	0.096
free from graffiti and vandalism	0.017	0.108	0.107	0.125	0.150
safe for children to play outside	0.052	0.070	0.057	0.075	0.092
free from heavy traffic	0.027	0.023	0.017	0.034	0.034
Health in general	0.080	0.091	0.086	0.111	0.125
treating with respect/consideration	0.037	0.020	0.021	0.042	0.030
Well-being					
satisfaction with life	0.020	0.020	0.024	0.033	0.036
feeling things done in life are worthwhile	0.028	0.027	0.027	0.039	0.040
happiness yesterday	0.015	0.015	0.026	0.031	0.033
anxiety yesterday	0.009	0.009	0.012	0.016	0.019
overall satisfaction with area lived in	0.078	0.086	0.097	0.118	0.131
Mean	**0.036**	**0.048**	**0.045**	**0.059**	**0.068**

Source = National Survey for Wales, quarter 1; n=5025; Coefficients are R squared values from regression models reported fully in Appendix 2.

The final stage of the analysis is to compare unweighted estimates to weighted estimates for the nineteen different survey outcomes using three different weights. This enables an assessment of the effect of different variables on the bias in survey estimates. It must be noted, however, that this is a far from perfect strategy for assessing non-response bias, because there is no criterion measure by which bias can be measured directly. Instead, we must assume that changes in estimates after the application of weights reflect *reductions* in bias. In practice, however it is possible for the application of weights to increase the bias in estimators and this needs to be borne in mind when considering the implications of table 5.

Table 5 Unweighted and Weighted estimates for 19 Survey Outcomes

Outcome	Unweighted	Weight1	Weight2	Weight3
Local area				
belonging to local area	78.14	78.01	77.98	76.05
people willing to help neighbours	77.14	76.95	76.81	75.26
safety at home after dark	68.28	68.00	67.87	68.10
safety walking in local area after dark	36.75	36.34	36.23	35.70
safety walking in nearest town/city centre after dark	15.52	15.37	15.36	15.75
safety travelling by public transport after dark	20.76	20.60	20.55	20.89
trusting people in the neighbourhood	50.88	50.52	50.31	48.06
well maintained	20.48	20.40	20.28	18.16
free from litter and rubbish	21.24	21.05	20.89	19.07
free from graffiti and vandalism	34.23	33.85	33.69	30.54
safe for children to play outside	25.29	25.04	24.87	24.24
free from heavy traffic	19.92	19.79	19.70	19.25
Health in general 'very good'	28.07	27.99	27.91	30.80
treating with respect and consideration	29.19	28.95	28.74	26.15
Well-being				
satisfaction with life	8.65	8.65	8.65	8.79
feeling that the things done in life are worthwhile	8.89	8.88	8.88	8.95
happiness yesterday	8.49	8.49	8.48	8.54
anxiety yesterday	3.79	3.80	3.80	3.74
overall satisfaction with area lived in	9.16	9.14	9.13	9.07

Source = National Survey for Wales, quarter 1; n=5025; Weight1=LSOAcode variables only; weight2=new observation variables; Weight3 = LSOAcode + post-stratification weights.

The estimates in Table 5 show that the first two weighting adjustments have very little influence compared to the unweighted estimate. For nearly every variable, weight1 and weight2 move the estimate very slightly away from its unweighted value. Additionally, for the majority of questions the new observation variables lead to a slightly larger shift away from the unweighted estimate compared to the LSOAcode variable. One might conclude from this pattern that there is little evidence of non-response bias in these estimates and that the new observation variables have a somewhat stronger effect in shifting the unweighted estimates toward the true population value. However, the change in the estimates across all variables after application of the weights is so small that weighting would, in practice, make no difference to the substantive conclusions that might be drawn from the data.

However, such a conclusion is rendered problematic by the estimates produced when the post-stratification weight (weight3) is applied. For a number of questions, application of weight3 results in notable shifts away from the unweighted estimate. Moreover, these are sometimes in the opposite direction to the changes engendered by application of the first two weights. For example, the proportion of the public strongly agreeing that people in their local area treat each other with respect and consideration drops by three percentage points from 29.2% to 26.2% when the post-stratification weight is applied. Similarly, the proportion reporting their health to be ‘very good’ increases by nearly three percentage points when weight3 is applied while weight1 and weight2 result in very modest decreases in the proportion reporting very good health. This can be taken as evidence that the unweighted estimates are subject to non-trivial levels of non-response bias which the first two non-response weights do not correct. This suggests that the variables used to derive the post-stratification weights (age, sex and housing tenure) are more powerful predictors of the covariance between response propensity and these survey outcomes than the variables used to derive the response propensity weights, including the new interviewer observation variables.

Conclusion

Given the substantial cost of household surveys and their centrality to understanding the attitudes and behaviours of populations for the purpose of policy-making and evaluation, it is essential that every effort is made to ensure that the estimates produced from them are accurate. One of the primary threats to the validity of survey estimates is non-response; when those who agree to participate in a survey differ from those who do not on the survey variable of interest, the survey estimate will be different from the true population value.

In addition to the various measures that are implemented during fieldwork in order to ensure that a survey achieves a high response rate, survey designers routinely apply weighted estimators to survey data in an effort to mitigate over- and under-representation of population sub-groups in the interviewed sample. However, the effectiveness of weighting strategies in reducing bias is hampered by the lack of available variables that are observed for both respondents and non-respondents. Over the past decade or so, many household surveys in the UK and overseas have adopted a strategy of requiring interviewers to make systematic observations about the characteristics of addresses and local areas for every address in their issued sample. Such variables are inexpensive and relatively straightforward to collect and therefore represent a potentially useful and cost-effective antidote to differential non-response.

However, despite their evident promise, recent assessments have concluded that interviewer observations are of limited utility in reducing bias in survey estimates. This is primarily because they are only very weakly correlated, if at all, with key survey variables (Kreuter et al 2010; Sturgis and Brunton-Smith 2012). The objective of this small research project was to evaluate whether a more strategic approach to the selection of interviewer observation variables might result in more powerful weighting adjustments. Most surveys appear to select interviewer observation variables on the basis of maximising the prediction of response propensity, or without any clear underlying rationale at all. In this study, observation variables were selected with the intention of maximising their ability to predict particular survey outcomes. These related to perceptions of the characteristics of local areas.

The results of the various models and comparisons undertaken do not result in a favourable evaluation of the new interviewer observation variables, or of interviewer observations in general. Although the new variables have a non-zero (conditional and unconditional) association with response propensity, the strength of the correlation is very weak, explaining only around 1 to 2% of the household variability in propensity to respond to the survey request. Furthermore, although the new observation variables selected for this study performed better than the standard variables implemented at the outset of the survey, this was by only a small amount. And the explanatory strength of the new observation variables was marginally lower than was observed for an auxiliary variable based on 2001 Census and administrative data, even for survey outcomes that were obviously and directly related to the observation variables.

Given these two findings – weak prediction of both response propensity and survey outcomes – it is perhaps unsurprising that the application of weights derived from models using the new variables had little effect compared to the unweighted estimates. Although the difference between unweighted and weighted estimates was slightly larger for the new observation variables compared to the standard ones, the magnitude of this difference was very small indeed and would be very unlikely to make any difference at all to substantive conclusions. In contrast, application of the post-stratification weight resulted in some more sizeable shifts relative to the unweighted data. This suggests that the variables used in the construction of the post-stratification weight – age, sex and housing tenure – are more powerful predictors of the covariance between response propensity and survey outcomes.

Given their weak association with both response propensity and survey outcomes (even when specifically selected in order to maximize the latter association), combined with the more substantial effect of the post-stratification weight, it is hard to escape the conclusion that the resources currently being expended in the collection of interviewer observations could be more usefully deployed on another area of fieldwork activity.

References

Kreuter F, Olson K, Wagner J, Yan T, Ezzati-Rice T, Casas-Cordero C, Lemay M, Peytchev A, Groves R, Raghunathan T (2010): Using Proxy Measures of Survey Outcomes to Adjust for Survey Non-response, *Journal of the Royal Statistical Society: Series A*, 173, 2, 389-407.

Martin, David (2001) *Geography for the 2001 Census in England and Wales*, London: Office for National Statistics.

Sturgis, Patrick and Brunton-Smith, Ian (2012) *An assessment of the potential utility of interviewer observation variables for reducing non-response error in the National Survey for Wales*. Cardiff, GB, Welsh Assembly Government, 48pp.