**Using linked census data to estimate survey non-response biases**

Jamie Moore, Peter. W. Smith & Gabriele B. Durrant

ADRCE & Dept. Social Statistics & Demography, University of Southampton, UK

**1. Introduction**

Methodologists no longer advocate only maximising response rates to minimise risks of non-response biases in survey dataset estimates. Such rates are an appealing indirect quality measure as information about non-respondent answers rarely exists, but have fallen recently (de Leeuw & de Heer 2002), and also work has shown a weak correlation with measured biases (Groves 2006; Groves & Peytcheva 2008). Instead, monitoring of variation in response between sample subgroups defined by attributes correlated with responses is recommended, during data collection if call record paradata detailing interview attempts is available. This can inform adaptive collection strategies in which methods may be modified to optimise quality – cost trade-offs (see Groves & Heeringa 2006; Peytchev et al. 2010).

Interest in using these methods to manage survey dataset non-response bias risks is increasing, but guidance for practitioners is limited, especially concerning monitoring response. One approach is to use representativeness indicators that quantify bias risk in terms of variation in sample estimated response propensities given an attribute covariate set, such as R indicators and Coefficients of Variation of response propensities (CVs: see Schouten et al. 2012; Wagner 2012). Low levels of variation (representativeness) imply low bias risk. These indicators require attribute information on the entire sample as they use logistic regression to estimate response propensities, but as well as measuring overall dataset quality are decomposable so that impacts of propensity variation associated with attribute covariates can be assessed. In support of their use as quality measures, recent work shows that increased dataset representativeness is associated with reduced non-response bias (Schouten et al. 2016).

An area that has received less attention is monitoring response during data collection. This often informs modifications to methods in adaptive collection strategies. A focus in this situation is on identifying design phase capacity (PC) points in call records after which further dataset quality increases are minimal and modifications, including ending collection completely, should be considered (e.g. Groves & Heeringa 2006; Moore et al. 2016). However, how representativeness indicator changes over call records and derived PC points compare to similar changes in survey estimate non-response biases has not been studied, so it is unknown whether indicator based inference is accurate.

We address this question using, for the UK, a unique dataset linking social survey responses and call record paradata to contemporaneous census sample member attribute information (a development of the Office for National Statistics (ONS) 2011 Census Non-Response Link Study (CNRLS)). Here, we focus on the Labour Force Survey (LFS). First, we utilise (overall) CVs and their (partial) decompositions to quantify dataset representativeness over the call record, and compute derived PC points. Second, we similarly quantify and derive PC points given changes in estimates of (likely) non-response biases, computed from the census attribute covariates used in CV response propensity estimation, from survey responses with analogues among these covariates, and from survey responses without such analogues. We then compare sets of findings to evaluate CV based inference.

**2. Methods**

The LFS is a longitudinal survey on labour market related topics. Simple random sampling of UK households (HHs) is used. Face to face interviews with all HH occupants are sought. We consider wave one data only, to avoid sample attrition effects. The CNRLS links HHs in the January to July 2011 survey sample to the 27th March 2011 occupant census records, providing individual level attribute data whether responses are obtained or not. We have appended details of calls (up to 20) to HHs until interview, refusal or fieldwork ending, enabling data collection to be studied (see Moore et al. 2016 for more details). Our analytical dataset includes 33811 individuals. The final response rate is 67.3%.

We use CVs to quantify dataset representativeness at each call (Schouten et al 2012; Wagner 2012). Given an attribute covariate set, the overall CV is the standard deviation of sample estimated response propensities divided by their mean (small values imply representativeness). It also defines survey estimate mean maximal absolute relative bias. Partial CV decompositions partition propensity variation across covariate categories, enabling those associated with non-representativeness to be identified. Values near zero suggest representativeness with respect to covariate or category (category CVs can be positive or negative, respectively implying over- or under-representation). We compute overall CVs, and partial unconditional CVs, which quantify univariate impacts, for attribute covariates (categories) in the response propensity model (we list these in Table 1). As well, we identify derived PC points. Following Moore et al. (2016), we begin by identifying points in terms of stability compared to call record best values, given an absolute threshold of 0.02 (call *n* CV – call record minima < 0.02).

We compare a range of census and survey data based estimates to our CV results. For census covariates in the CV response propensity model entire sample information is available, so we compute as estimates of category ‘non-response biases’ respondent – sample dataset proportion differences relative to the latter (((call *n* respondent dataset proportion – sample proportion) / sample proportion) \*100). We identify derived PC points using an absolute threshold of 2% (abs(call n value – call record minima) < 2)), which, given indicator definition is comparable to the CV analysis 0.02 threshold. For survey covariates though, non-respondent information is unavailable and sample values are unknown, so only changes in respondent dataset estimates are quantifiable. We address this issue by first for CVs and census covariate category relative respondent – sample dataset differences also comparing PC points given percentage thresholds of within 2% of final dataset values (((call *n* CV or difference – final dataset CV or difference) / final dataset CV or difference) \* 100 < 2). Second, for the census covariates we compute category respondent – final respondent dataset differences relative to the latter (((call n dataset proportion – final dataset proportion) / final dataset proportion) \* 100), and identify derived PC points given thresholds of within 2% of final dataset values. We then compare these to similar computed for survey responses with census covariate analogues, and also for three survey responses without such analogues (see Table 1). We interpret CV based inference as accurate for census covariate survey analogues if it is similar to covariate respondent-sample dataset difference based inference, and covariate and survey analogue respondent – final respondent dataset difference based inference is similar. This can be extended to survey non-census covariate analogues if respondent – final respondent dataset difference based inference is similar to the equivalent for census covariate analogues.

**3. Results**

The LFS survey response rate increases over the call record, at first substantially but then at a decreasing rate, with minimal further increases after calls 9 to 11 and none at all after call 16 (Fig. 1a). Overall CVs decrease, suggesting increased dataset representativeness, at a decreasing rate over the call record (also see Fig. 1a). To quantify sources of non-representativeness, in Figure 1b we present attribute covariate CVs. Sample member gender CVs are negligible, implying a minimal impact on representativeness. HH tenure CVs increase from a minima at call one to a maxima at call five, then decrease slightly afterwards, due to under-representation of rented HH sample members in the dataset (results not shown; we concentrate on under-represented categories as they are likely targets in adaptive collection strategies: see Schouten et al. 2012). Sample member age, highest qualification, and activity last week CVs all decrease at a decreasing rate over the call record. We present category CVs for these three covariates, which we focus on in the following work (though see Discussion), in Figure 2a. Age CVs are due to (initial) under-representation of 16 to 34 and 35 to 54 year olds, highest qualification CVs due to under-representation of those with no qualifications, GCSEs and A levels, and activity last week CVs due to under-representation of those employed. Decreases in indicators over the call record imply that many of these sample members are interviewed eventually.

Given a 0.02 absolute difference threshold, the derived overall CV PC point is at call six. Under-represented age, highest qualification and activity last week category CV PC points are all earlier in the call record (Table 2), the earliest being for the categories no qualifications and A levels (call two).

Relative census covariate category respondent – sample dataset differences (‘non-response biases’) for age, highest qualification and activity last week exhibit analogous patterns to equivalent CVs (Fig. 2b). Under-represented category values are negative, and reduce at a decreasing rate over the call record. PC points given a 2% absolute difference threshold are also similar to the CV derived points, being the same for two of the six categories and one call later for the other four (Table 2).

Category CV and respondent – sample dataset differences relative to final respondent dataset values are identical (not shown). Changes over call records are as expected given earlier results. PC points for the under-represented categories given 2% of final respondent dataset value difference thresholds are much later in the call record than absolute difference threshold points (Table 2). The earliest are for those aged 16 to 34 (call 9) and for several categories metrics are outside thresholds until the last responses are obtained (call 16). The overall CV PC points are at call 9.

Relative census covariate category respondent – final respondent dataset differences over call records for age, highest qualification and activity last week are also as expected given earlier results (Fig. 2c). PC points for under-represented categories given a 2% of final respondent dataset value difference threshold are much earlier than CV and respondent – sample dataset difference equivalents, in fact being only one to two calls later than points given the two metrics and absolute thresholds (Table 2). Similar patterns are also found when such differences are computed for survey response analogues of the census covariates (Fig. 2d). Derived PC points are the same for four of the under-represented categories, and one call later with survey responses for the other two (Table 2).

We also compute relative respondent – final respondent dataset differences for the survey items on benefits, marital status and health, which lack census covariate analogues. Some category proportions are initially lower than final dataset values, namely not on benefits, single, separated, in a civil partnership, and in very good health (Fig. 3). Category PC points also identified as above are the same or slightly later in call records than census covariate survey analogue points (Table 2).

**4. Discussion**

We assess representativeness indicator based inference concerning survey dataset quality during data collection. Studying the UK Labour Force Survey, we compare indicators (CVs) over call records and derived phase capacity (PC) points after which quality improvements are minimal to relevant estimates computed from linked census sample member attribute information (also used in CV estimation) and survey responses. Considered census covariate category CV changes are similar to category relative respondent-sample dataset difference (‘non-response bias’) changes. Derived PC points given absolute difference thresholds tend to be slightly later in call records for category differences, but points given percentage of final dataset value thresholds (as necessary for survey responses, for which sample values are unknown) are identical (and later than absolute threshold points; we use comparable thresholds of 0.02 for CVs and 2% for differences). Census covariate respondent-final respondent dataset difference derived PC points, identified given percentage difference of final dataset value thresholds (which are the same or slightly later than absolute difference threshold points) are similar to equivalent points derived given survey response analogues of the census covariates, and also to points derived given responses to three survey items without such analogues covering other topics.

These findings could be taken to imply that CV based inference concerning when in call records further dataset quality improvements are minimal is accurate. Category CV and census covariate respondent-sample dataset difference derived PC points are similar. PC points (instead) derived given census covariate respondent – final respondent dataset differences are comparable to points derived similarly from survey responses, both those that are analogues of the census covariates and those on other topics. However, a caveat to this with regard to non-census covariate analogues concerns the type of category estimate changes that we focus on in the above work. With each, census covariate respondent – sample dataset differences (‘non-response biases’) reduce over the call record. This may not always be the case, an example from our larger analysis being the impact of the HH tenure category Rented / Other (which we mention only briefly earlier). This category is least under-represented in the call one dataset. Respondent-sample dataset differences are also minimal at this call, but due to subsequent increases in such differences census and survey respondent – final respondent dataset value difference derived PC points are later in call records (results not shown). Consequently, it must be stated that our non-census covariate analogue survey response results still require confirmation: presented differences may even be computed relative to data collection maximal non-response biases.

Given this, our questions to workshop colleagues concern alternative methods of evaluating non-response biases, particularly for covariates for which there is no information about non-respondent answers. Is there anything out there? Can the information from respondents be incorporated?

**5. Acknowledgements**

This research was funded by the ESRC National Centre for Research Methods, Workpackage1 (grant reference number ES/L008351/1) and the ESRC Administrative Research Centre for England (ADRCE) (grant reference number ES/L007517/1). This work contains statistical data from ONS which is Crown Copyright. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

**6. References**

de Leeuw, E. & de Heer, W (2002) Trends in Household Survey Nonresponse: A Longitudinal and International Perspective. In: *Survey Nonresponse* (eds. R.M. Groves, D.A. Dillman, J.L. Eltinge and R.J.A. Little). pp. 41-54. New York: Wiley.

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

Groves, R. M. and Heeringa, S. (2006) Responsive design for household surveys: tools for actively controlling survey errors and costs*. J. Roy. Stat. Soc. Ser. A.*, 169, 439-457.

Groves, R. M. & Peytcheva, E. (2008) The impact of non-response rates on non-response bias: a meta-analysis. *Public Opinion Quarterly*, 72, 167–189.

Moore, J.C, Durrant, G.B. & Smith, P.W.F (submitted) Dataset representativeness during data collection in three UK social surveys: generalizability and the effects of auxiliary covariate choice. *J. Roy. Stat. Soc. Ser. A.*

Peytchev, A., Riley, S., Rosen, J., Murphy, J., & Lindblad, M. (2010). Reduction of nonresponse bias in surveys through case prioritization. *Survey Research Methods,* 4, 21-29.

Schouten, B., Bethlehem, J., Beullens, K., Kleven, Ø., Loosveldt, G., Luiten, A., Rutar, K., Shlomo, N. & Skinner, C. (2012) Evaluating, comparing, monitoring, and improving representativeness of survey response through R-indicators and partial R-indicators. *Int. Stat. Review*, 80, 382-399.

Wagner, J. (2012) A Comparison of Alternative Indicators for the Risk of Nonresponse Bias. *Public Opinion Quarterly*, 76, 555–75.

Table 1: Census attribute covariates, their survey response analogues, and other survey response covariates considered in this study. See text for explanation.

|  |  |
| --- | --- |
| Covariate | Categories |
| *Census and survey attribute:* |  |
| Gender | 1) Male; 2) Female; 3) Not Recorded (NR). |
| Age | 1) 16 to 34; 2) 35 to 54; 3) 55 to 74; 4) 75 & over; 5) NR. |
| Highest Qualification | 1) None; 2) Other; 3) GCSEs; 4) A levels; 5) Higher Ed.; 6) Degree; 7) NR. |
| Activity last week | 1) Employed; 2) Unemployed; 3) Economically inactive (EI): Student; 4) EI: Retired; 5) EI: Ill / impaired; 6) EI: at home; 7) EI: Other; 8) NR. |
| Household Tenure | 1) Owned; 2) Rented / Other; 3) NR. |
| *Other survey:* |  |
| Benefits | 1) No; 2) Yes; 3) NR. |
| Marital status | 1) Single; 2) Married; 3) Separated; 4) Divorced; 5) Widowed; 6) Civil partnership; 7) NR. |
| Health | 1) No; 2) Yes; 6) NR. |

Table 2: Design phase capacity (PC) points for selected census and survey (census covariate analogue and otherwise) covariate categories given different metrics and stability thresholds (see text).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Covariate | CV | Census % ‘ NR bias’ | CV % / Census % ‘NR bias’ diff. | Census % diff. | Survey % diff. |
| Age: 16 to 34 | 4 | 5 | 9 | 5 | 5 |
| Age: 35 to 55 | 4 | 4 | 12 | 4 | 5 |
| H. Qual.: None | 2 | 3 | 11 | 5 | 5 |
| H. Qual.: GCSEs | 3 | 4 | 16 | 4 | 4 |
| H. Qual.: A levels | 2 | 3 | 16 | 3 | 4 |
| Act. L. Wk.: Employed | 5 | 5 | 14 | 5 | 5 |
| Benefits: No | NA | NA | NA | NA | 5 |
| Mar. stat.: Single | NA | NA | NA | NA | 4 |
| Mar. stat.: Separated | NA | NA | NA | NA | 7 |
| Mar. stat.: Civil partner | NA | NA | NA | NA | 6 |
| Health: Very good | NA | NA | NA | NA | 5 |



Figure 1: a) Survey response rate and overall CVs over the call record; b) similar partial unconditional CVs for attribute covariates in the propensity model.

 





Figure 2: a) Partial unconditional covariate category CVs over the call record; b) similar census covariate category relative respondent – sample dataset differences (‘non-response biases’); c) similar census covariate category relative respondent – final respondent dataset proportion differences; and d) similar census covariate survey covariate analogue category relative respondent – final respondent dataset proportion differences, for the covariates age, highest qualification and activity last week.



Figure 3: Category relative respondent-final respondent dataset proportion differences over the call record for the (non-census covariate analogue) survey covariates benefits, marital status, and health.