Are trajectories of dataset representativeness during survey data collection generalizable? Evidence from the 2011 Census Non-Response Link Study.

Jamie Moore*
Administrative Data Research Centre for England and Department of Social Statistics and Demography, University of Southampton, Southampton, UK. Email: j.c.moore@soton.ac.uk
Gabriele Durrant
Administrative Data Research Centre for England and Department of Social Statistics and Demography, University of Southampton, Southampton, UK. Email: g.durrant@soton.ac.uk
Peter W.F. Smith
Administrative Data Research Centre for England and Department of Social Statistics and Demography, University of Southampton, Southampton, UK. Email: p.w.smith@soton.ac.uk

Abstract

Methodologists have moved away from using survey response rates to indicate risks of non-response bias. Instead, indicators estimating dataset representativeness in terms of variation in sample frame response propensities have been developed. Incorporating statistical models predicting response propensities, and so requiring attribute information on the entire sample frame, these indicators also allow variation to be decomposed to assess the impact of respondent response propensity variation associated with multiple attribute variables. Coupled with call record paradata, they can also be used to monitor data collection over time, and so can inform adaptive strategies in which methods may be modified between fieldwork periods with the aim of maximizing data quality / minimizing costs. Use of these indicators is growing, but gaps in our knowledge remain, especially regarding adaptive strategies. For example, the generalizability of trajectories of dataset representativeness over data collection across different surveys of the same sample frame is unstudied. Hence, it is unknown whether rules for ending data collection with minimal effect on representativeness derived for one survey of a population are applicable to another. We address this question using sample frame census attribute data and call record paradata for two UK social surveys. We use R indicators and Coefficients of Variation of response propensities to describe dataset representativeness trajectories over collection for each survey, and consider whether it is possible to end collection early without affecting data quality. We then compare trajectories in each survey, and evaluate whether rules are generalizable.

Keywords: Non-response bias, representativeness indicators, adaptive data collection, paradata.

1. Introduction

For many years, methodologists sought to maximize survey response rates, utilizing them as indicators of the risk of bias in estimated quantities caused by non-response (e.g. Olson 2006). Recently though, response rates have fallen (de Leeuw & de Heer 2002). In addition, empirical work has indicated they are poor predictors of non-response bias (see, for example, Groves & Peytcheva 2008). Consequently, researchers have begun to develop alternative methods to assess the risk of non-response bias.

In particular, these methods include indicators estimating dataset representativeness in terms of variation in the propensity of survey sample frame members to respond, such as R indicators and Coefficients of Variation of response propensities (hereafter CVs: see Schouten et al. 2012; Wagner 2012 for details). Based on statistical models predicting response propensities, indicators of this type require attribute information on the entire sample frame, from a previous wave, linked administrative data, population register or census. Another advantage is that they can be decomposed to assess impacts on representativeness of response propensity variation associated with multiple attribute variables. Moreover, when coupled with call record paradata and visualisation of their trajectories, such indicators can be used to monitor data collection over time (Lundquist & Särndal 2013; Correa et al. 2015). Hence, they can inform adaptive strategies in which data collection is modified between fieldwork periods to maximize data quality / minimize collection costs.

Use of the above indicators is growing, but gaps in our knowledge remain, especially regarding adaptive strategies. For example, one reason for monitoring representativeness trajectories is to derive stopping rules for ending data collection without affecting data quality. Given monitoring costs, it would be useful if once derived these rules could be generalized to other surveys. In the limited work so far though, it has been found in some surveys that representativeness increases at a decreasing rate over the call record, reaching a point where further increases are minimal (e.g. Correa et al. 2015), whereas in others it does not increase at all (e.g. Lundquist & Särndal 2013). However, the studied surveys took place in different countries. Consequently, it is still unknown whether representativeness trajectories are similar for surveys of the same population.

We address this question using, for the UK, a unique dataset linking outcomes and call records from two social surveys to census attribute data on both responding and non-responding households (part of the Office for National Statistics (ONS) 2011 Census Non-response Link Study, or CNRLS). We utilise R indicators and CVs to quantify dataset representativeness over the repeated attempts to interview sample frames. We consider whether data collection for each survey can be ended early without affecting quality, and then compare patterns to evaluate whether such rules are generalizable.

2. Methods

We study: a) The Life Opportunities Survey (LoS), which considers local facility use and participation in leisure activities and employment opportunities, focussing on the effects of impairment; and b) The Labour Force Survey (LFS), which considers labour market related topics. Surveys employ simple random sampling of households (HHs), with face to face interviews with all HH members sought.

The CNRLS links HHs in the survey sample frames from January to July 2011 to the 27th March 2011 census records of their occupants, providing HH attribute data whether survey responses were obtained or not (Parry-Langdon 2011). It is a rare resource for studying dataset representativeness as it is, but we have further appended details of calls made (up to 20) to HHs until interview, refusal or fieldwork ending. Hence, we can also investigate representativeness trajectories over data collection. Both the LoS and the LFS are longitudinal surveys, but as the CNRLS includes cases around the census HHs appear only once. Hence, our datasets are cross-sectional. We consider wave one survey data only, to avoid sample attrition effects on representativeness. The LoS dataset includes 5870 HHs and the LFS dataset 16931 HHs. The final response rate is 74.2% in the LoS and 67.6% in the LFS.

For each dataset, we quantify representativeness at each call (1 to 20) using R indicators and CVs (Schouten et al 2012; Wagner 2012). Given a set of attribute variables, overall R indicators are the transformed (to 0 to 1) standard deviation of sample predicted response propensities. Higher values indicate greater representativeness. Overall CVs are these standard deviations divided by response propensity means. Smaller values indicate greater representativeness. We compute overall indicators, and unconditional partial by variable variants to assess impacts of variation associated with different HH attribute variables. With partial indicators, smaller values indicate greater representativeness.

We consider a range of HH attribute variables relevant to surveys in analyses (Table 1), but assess impacts on representativeness of (and include in propensity models) only those with, as main effects, a substantial effect in either of two logistic regressions predicting response propensities in the two final survey datasets. We defined a substantial effect as an increase in the Akaike Information Criterion (AIC) of >2 when removed from / added to the final model, and dropped 'Impaired individual in HH' and 'Anyone Fluent in English in HH' for the analyses presented here (details not shown). We use graphical methods developed by Correa et al (2015) to visualise indicator trajectories.

3. Results

We visualise response rate development in the two datasets in Figure 1. Rates develop similarly, at first increasing substantially but at a decreasing rate with minimal further increases for later call numbers. The LFS call one rate is higher but rates of increase after lower than in the LoS. Rates increase by <1% after call 11 in the LoS and call 9 in the LFS. The mean number of calls per HH is 7.64 (SD = 7.55) for the LoS and 7.97 (SD = 8.08) for the LFS, and the mean number of calls to a HH per successful interview is 3.33 (SD = 2.22, range = 1 to 19) for the LoS and 2.75 (SD = 1.94, range = 1 to 17) for the LFS.

We visualise overall representativeness indicators in the two datasets given the attribute variables in Figure 1. The two sets of trajectories are qualitatively similar, but differ quantitatively. R indicators at call one are both high, indicating high representativeness, decrease at calls two and three, then increase at decreasing rates, with minimal (final value minus call *n* value <0.01) further increases after call 11 for the LoS and call 8 for the LFS. CVs both decrease, indicating increased representativeness, at decreasing rates, with minimal (call *n* value minus final value <0.01) further decreases after call 11 for the LoS and call 8 for the LFS. The R indicator early trajectories likely reflect a shortcoming: at low response rates they can over-estimate representativeness because population sub-group response propensities are less able to diverge (Schouten et al. 2012). CVs account for this by standardising, though they also indicate increased representativeness if response rates rise but response propensity variation stays the same (Lundquist & Särndal 2013). However, here R indicator increases after initial reductions suggest representativeness increases over call records. The LoS R indicator starts larger and decreases for longer than for the LFS, and its final value is slightly smaller. CVs for the two datasets also indicate similar representativeness trajectories (see also below).

We visualise LOS dataset partial indicators in Figures 2a, 2b, 3a & 3b. 'HH Economic Status', 'Retiree in HH', 'HH Structure' (the variables impacting most on representativeness), and 'Individual in Ill Health in HH' indicators exhibit equivalent trajectories to overall indicators, with the same explanations. R indicators at call one are low (indicating high representativeness), increase at calls two and three then decrease at decreasing rates over the record. CVs decrease at a decreasing rate. 'Located in London / SE' R indicators increase then level off and 'Cars available' R indicators remain similar over the record, with CVs decreasing at decreasing rates. These trajectories are analogous to and have the same explanation as those found by Lundquist & Särndal (2013): that response propensity variation associated with variables plateaus / remains similar over the record and decreases in CVs reflect higher response rates. For 'Tenure' and 'Accommodation type', both indicators increase to call four then decrease slightly. The CV trajectories suggest variation associated with these variables does initially increase, though they could also occur if enough subgroup response propensity divergence accompanies higher response rates. For variables for which they are relevant, points after which there are minimal indicator decreases tend to occur earlier in the record than overall indicators suggest (details not shown).

We visualise LFS dataset partial indicators in Figures 4a, 4b, 5a & 5b. Some are qualitatively similar to those for the LoS. 'HH Economic Status' and 'Retiree in HH' R indicators increase then decrease at decreasing rates and CVs decrease at decreasing rates over the record. 'Tenure' R indicators and CVs both increase then decrease slightly. For most other variables, only slight differences are seen from those for the LoS. 'Individual in Ill Health in HH' R indicators decrease at a decreasing rate and 'Located in London / SE' R indicators are similar over the record, instead of increasing then decreasing, but CVs decrease at decreasing rates as for the LoS. 'Accommodation type' R indicators increase then decrease slightly, as for the LoS, but CVs decrease slightly instead of increasing then decreasing. 'HH structure' R indicators increase then decrease at a decreasing rate and CVs decrease at a decreasing rate, as for the LoS but to a lesser extent. These differences seem linked to response rate development. The lack of initial increases in 'Individual in Ill Health in HH' and 'Located in London / SE' R indicators is correlated with higher LFS response rates at that point, which likely enable more response propensity divergence than in the LoS (this is also probably why LoS overall R indicators initially indicate greater representativeness). Moreover, higher response rates can explain the lack of initial increase in LFS 'Accommodation type' CVs: they are denominators in statistics. Generally, indicator changes are also smaller in the LFS, in which response rates rise less sharply.

For 'Cars available', a substantively different representativeness trajectory to that for the LoS is suggested. Both R indicators and CVs decrease then increase at decreasing rates over the record, instead of respectively remaining similar and decreasing at a decreasing rate. In our talk, we further consider this difference using by variable category indicators (we do not present these here due to lack of space). We mention as well that even when LFS indicator trajectories are qualitatively similar to in the LoS, there are differences in call one values, rates of change and points at which values asymptote or plateau. Final indicator values for variables are mostly similar (differences <0.02), though within dataset rankings differ, with for the LFS 'Located in London / SE', 'Accommodation type' and 'HH

Structure' impacting most on representativeness. For variables for which they are relevant, points after which there are minimal indicator decreases (again these are mostly earlier than the overall indicators suggest) tend to occur earlier in the call record for the LFS (details not shown).

4. Conclusions and further work

We quantify dataset representativeness trajectories for two UK social surveys using R indicators and CVs, to study whether rules for ending data collection early without affecting quality are generalizable across surveys of the same population. Our results indicate many qualitative similarities in trajectories between the two datasets, both overall and with partial indicators assessing variation associated with different HH attribute variables. Generally (i.e. with a few variable level exceptions), representativeness increases at a decreasing rate, with minimal increases later on in call records. Moreover, differences between datasets often seem to be linked to response rate development patterns.

Representativeness trajectories like those detailed above, with minimal increases later on in call records, suggest derivation of rules for ending data collection early is feasible for the two datasets. However, would rules be similar? Notwithstanding qualitative similarities in trajectories, actual points when representativeness increases became minimal (defined as a difference between an indicator and its final value <0.01) differ between datasets. For example, points for overall indicators occur at call 11 for the LoS and at call 8 for the LFS (we consider variable level patterns in our talk). Given this, and as we are not aware of other work comparing dataset representativeness over call records in surveys of the same sample frame, we must therefore advise caution when generalising stopping rules, both across the studied surveys and across groups of surveys considering other populations.

We recognise though, that a two survey comparison has limited power. The CNRLS includes two more UK survey datasets with relevant paradata, the ONS Opinions Survey and the Living Costs and Food Survey. We are currently analysing these datasets, and plan to include findings in our talk.

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

Correa, S, Durrant, G. & Smith, P.W. (2015) Assessing non-response bias using call record data with applications in a longitudinal study. *Submitted to Journal of the Royal Statistical Society Series A.*

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

Lundquist, P. & Särndal, C-E. (2013) Aspects of responsive designs with applications to the Swedish Living Conditions Survey. *Journal of Official Statistics*, 29, 557-582.

Olson, K. (2006) Survey participation, nonresponse bias, measurement error bias, and total bias. *Public Opinion Quarterly*, 70, 737–758.

Parry-Langdon, N. (2011) *Social Survey Non-Response Update*. ONS Technical Report. Available at: http://www.ons.gov.uk/ons/dcp171766_240879.pdf.

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: Household attribute variables considered in this study. Variables in italics not included in representativeness analyses. See text for explanation.

Variable	Categories
HH Economic Status	1) All employed; 2) all unemployed; 3) all inactive; 4) mixed.
HH structure	1) One adult; 2) One adult, children; 3) couple, no children; 4) couple, children; 5) >Two adults, children or otherwise.
HH type	1) House; 2) Flat; 3) Other.
Tenure type	1) Owned; 2) Rented / other.
Cars available	1) None; 2) One car; 3) Two cars; 3) Three or more cars.
Individual in ill health in HH	1) Yes; 2) No.
Retiree in HH	1) Yes; 2) No.
Located in London / South East	1) Yes; 2) No.
Impaired individual in HH	*1) Yes; 2) No.*
Anyone fluent in English in HH	*1) Yes; 2) No.*

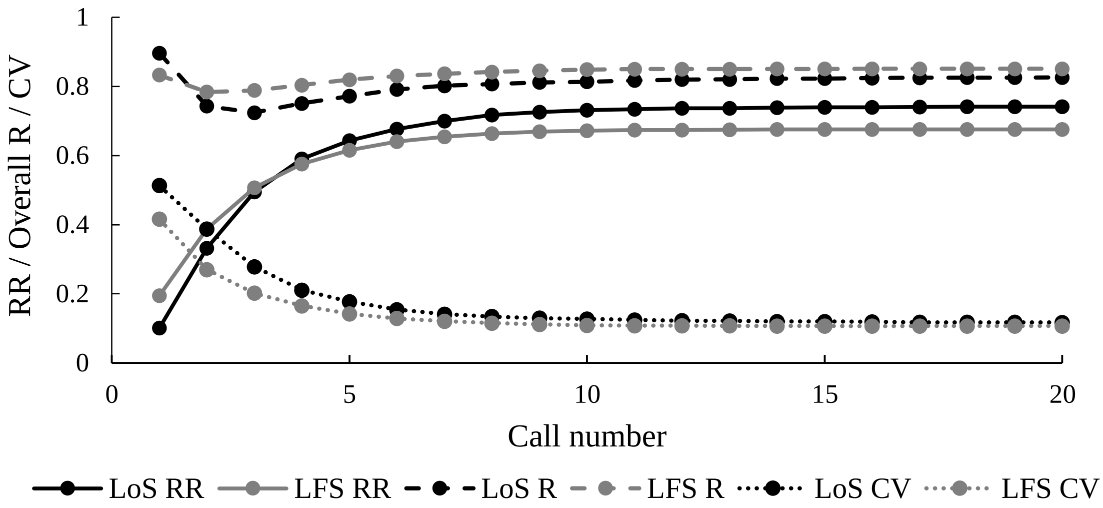

Figure 1: Response rates (RR), overall R indicators (R), and overall Coefficients of Variation of response propensities (CVs) over the call record for the LoS and LFS datasets. See text for explanation.

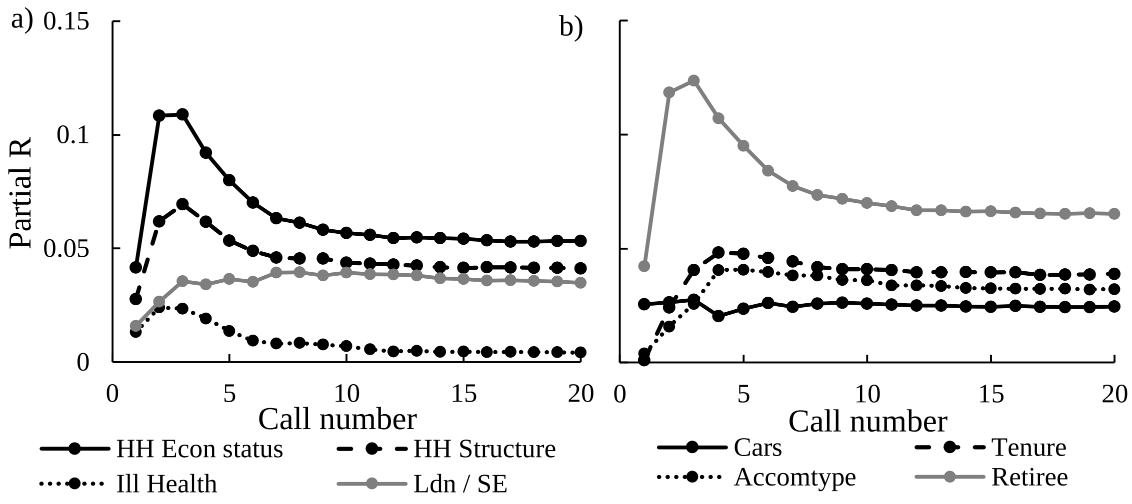

Figure 2: a) and b) Unconditional partial R indicators over the LoS call record for attribute variables included in the representativeness analyses. See text for explanation.

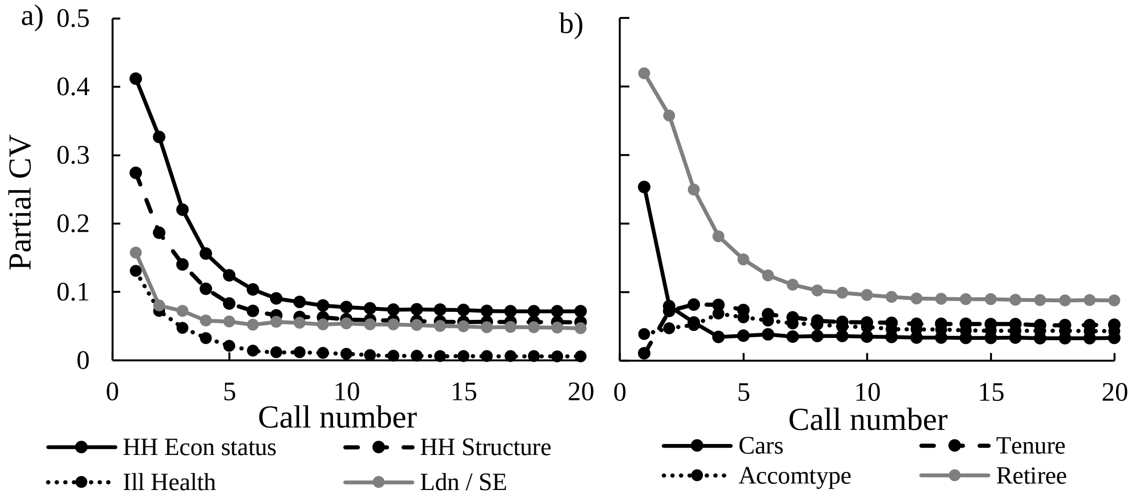

Figure 3: a) and b) Unconditional partial CVs over the LoS call record for attribute variables included in the representativeness analyses. See text for explanation.

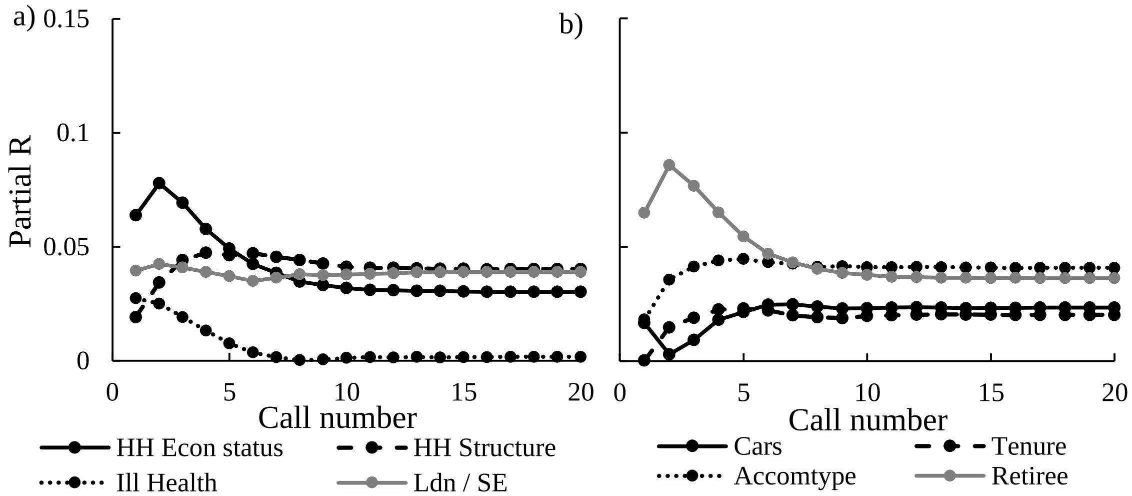

Figure 4: a) and b) Unconditional partial R indicators over the LFS call record for attribute variables included in the representativeness analyses. See text for explanation.

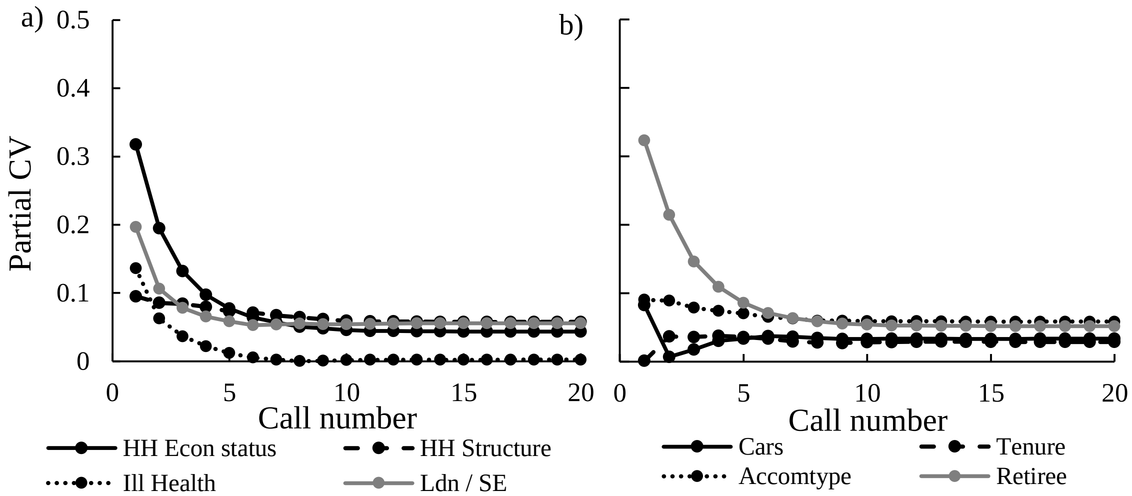

Figure 5: a) and b) Unconditional partial CVs over the LFS call record for attribute variables included in the representativeness analyses. See text for explanation.