Dataset representativeness during data collection in three UK social surveys: generalizability and the effects of auxiliary covariate choice

Jamie C. 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 B. 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 word count: 142

Main paper word count: 6572

Address for correspondence:
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

Abstract

We consider the use of representativeness indicators to monitor risks of non-response bias during survey data collection. The analysis benefits from use of a unique dataset linking call record paradata from three UK social surveys to census auxiliary attribute information on sample households. We investigate: a) the utility of census information for this purpose; b) the performance of representativeness indicators (the R indicator and the Coefficient of Variation of response propensities) in monitoring representativeness over call records; c) the extent and impacts of misspecification of auxiliary covariate sets used in indicator computation; and d) design phase capacity (PC) points in call records beyond which survey dataset improvements are minimal, and whether such points are generalizable across surveys. Given our findings, we then offer guidance to survey practitioners on the use of such methods and implications for optimising data collection and efficiency savings.

Keywords: risk of non-response bias, R indicators, Coefficient of Variation, adaptive and responsive survey designs, phase capacity, data collection efficiency savings.

1. Introduction

Survey methodologists no longer advocate maximising response rates to minimise risks of non-response bias (see Olson 2006; Kreuter 2013 for historic details). Rates have declined in the last 30 years (de Leeuw & de Heer 2002), and have also been shown to be only weakly related to biases (Groves 2006; Groves & Peytcheva 2008). Instead, monitoring risks by quantifying variation in response between sample subgroups whose attributes are correlated with survey estimates is recommended, during data collection if paradata such as call records or details of other follow up attempts are available. This can inform modifications to methods, to reduce such variation and improve dataset quality (by targeting under-represented subgroups) and / or minimise costs (adaptive and responsive collection strategies: e.g. Groves & Heeringa 2006; Wagner 2008; Peytchev et al. 2010). Survey agencies are increasingly interested in employing this more refined approach to managing non-response bias risks, but reports of its use are still few, especially concerning monitoring during data collection, and available guidance is limited.

The above described approach to managing non-response bias risks requires similarly motivated risk indicators (reviewed by Wagner 2012; see also Lundquist & Särndal 2013; Särndal & Lundquist 2014; Correa et al. 2016). One often used type are representativeness indicators, which measure risks in terms of sample response propensity variation as estimated by a statistical model given an auxiliary attribute covariate set. Low levels of variation imply representativeness and low bias risk (see Schouten et al. in press for empirical support). Indicator computation requires auxiliary information on all sample units, concerning survey estimate correlates or socio-demographic attributes, which can be obtained from administrative data, a previous wave, census or population register. The most studied is the R Indicator, which is the transformed (zero to one) standard deviation of response propensities (SD): $R = 1 - 2SD$ (Schouten et al. 2009, 2011, 2012). This form measures overall representativeness, enabling different surveys or waves to be compared given use of the same auxiliary covariate set. Partial decompositions measuring variation associated with factorial covariates also exist, enabling impacts on representativeness to be assessed, for instance to identify target subgroups when modifying methods (see Schouten & Shlomo in press). Unconditional and conditional forms can be calculated, quantifying respectively the extent to which response with respect to a covariate is representative (a random sample) or conditionally representative (a random sample given stratifying covariates). Conditional variants thus enable detection of correlated impacts, and so when modifying methods can ensure efficient targeting of (different) subgroups.

Guidance on several aspects of the use of these techniques to manage non-response bias risks is needed. To begin with, in previous reports sample information is from population registers, administrative data or previous waves (Lundquist & Särndal 2013; Luiten & Schouten 2013; Ouwehand & Schouten 2013; Kappelhof 2014; Correa et al. 2016; Schouten et al. in press). In some countries including the UK, the first two data sources do not exist, and the only non-longitudinal information available is from censuses. So far though, research on the use of such data is limited to identification of UK census derived correlates of social survey non-response (Durrant & Steele 2009; Steele & Durrant 2011; Durrant et al. 2010, 2011, 2013). Its utility for non-response bias risk monitoring is unknown.

There are also questions concerning representativeness indicator use when monitoring data collection. First, at low response rates possible response propensity variation is limited, so R indicators may suggest representativeness is highest early in call records (Schouten et al. 2009). This can, for example, cause issues if identifying when to modify methods (see also below). An indicator with potentially better properties is the Coefficient of Variation of response propensities (CV: Schouten et al. 2009). The overall CV is *SD* divided by mean propensity (low values imply representativeness), so is less likely to be similarly affected by the response rate. It also provides a link to actual non-response biases, as it quantifies the maximal absolute standardised bias of a survey estimate mean when non-response correlates maximally to the utilised auxiliary covariate set. However, the CV is less studied than the R indicator, especially partial decompositions (de Heij et al. 2015), and comparisons of indicator behaviour over call records are rare (Lundquist & Särndal 2013; Correa et al. 2016).

Second, specifying auxiliary covariate sets for use over call records is problematic. Indicators are set specific, so the same set must be fitted at each call to isolate dataset changes. Sets should include all available response propensity correlates: simulations suggest exclusions lead to overall R indicators comparatively over-estimating representativeness (Shlomo et al. 2012), and including non-correlates to under-estimation and inflated indicator errors (Schouten et al. 2009; similar is expected with CVs). However, for a given sample size model selection methods should retain fewer covariates at low response rates, again because possible propensity variation is limited (Schouten et al. 2009). Hence, any set may be correctly specified (include only correlates) over only part of a call record, with sets correct at early call(s) likely to exclude later call dataset correlates, and sets correct at later call(s) or specified without model selection likely to include (early call dataset) non-correlates. Advice on covariate set specification given these considerations is lacking. We are unaware of any published empirical work on propensity correlate changes over call records, or on the extent of set misspecification effects on indicators.

In addition, a focus when monitoring data collection is on identifying when continued use of current methods leads to minimal further increases (or even decreases) in data quality and modifications should be considered, termed reaching design phase capacity by Groves & Heeringa (2006; see also Rao et al. 2008; Wagner & Raghunathan 2009; Schouten et al. 2013). However, reports of R indicators and CVs discuss these ‘PC’ points only briefly, in the context of ending future data collection early given overall indicator stability compared to best values over (complete) call records (Correa et al. 2016). Points computed given partial indicators, and those computed given information only up to the current call (i.e. during collection), as necessary when historic data do not exist (e.g. Groves & Heeringa 2006), are not presented, and so it is unknown how they compare. As well, whether PC points are generalizable from one survey to others, which is appealing to survey agencies given frequent legislative issues relating to linking sample information and also the costs of (real time) monitoring, is unstudied.

We address these questions using a unique dataset linking details of attempts to interview households (HHs) in three UK social survey samples to HH attribute information from a concurrent census (a development of the Office for National Statistics (ONS) 2011 Census Non-Response Link Study (CNRLS)). The dataset enables monitoring of HH level response (defined as at least one interview) during data collection, which we undertake by computing R indicators and CVs at each call for each survey, considering 10 HH attribute covariates in our analyses. First, we evaluate the utility of census data for this purpose. Second, we investigate auxiliary covariate retention in sets used in indicator response propensity estimation, by conducting logistic regression model selection given datasets after: a) five interview attempts (early in data collection); and b) 20 attempts (the end of collection). Third, we compare indicator behaviour, and investigate auxiliary covariate set misspecification effects by computing indicators given sets a) and b) and also sets c) including all 10 covariates. Fourth, we identify survey overall and partial CV stability based PC points and evaluate their generalizability, both when entire call record information is available for their calculation (after collection) and when information exists only up to the current call (during collection). We then summarise our findings, and offer guidance to survey practitioners on the considered issues.

2. Methods

2.1. Datasets

The CNRLS links January to July 2011 sample HHs from six UK social surveys to their 27th March 2011 census records, providing attribute information whether they are interviewed or not (Parry-Langdon 2011). We append call records, enabling monitoring during data collection, to three surveys: a) The Labour Force Survey (LFS), covering labour market topics (ONS 2011a); b) The Life Opportunities Survey (LOS), covering local facility use and leisure and employment activity participation with a focus on the effects of impairment (ONS 2014a); and c) The Opinions Survey (OPN), covering social and health topics (ONS 2011b). The LFS and LOS randomly sample households (HHs), and seek interviews with all HH members. The OPN randomly samples HHs within areas (postcode sectors), and seeks an interview with a single HH member. Surveys are comparable both with respect to definitions of HHs and HH level response (i.e. whether an(y) interview is obtained or not). The OPN is a cross-sectional survey. The LFS and LOS are longitudinal, but to avoid sample attrition effects we consider wave one data only, so the analysed datasets are cross-sectional. Interviews are face to face in the LOS and OPN, but in LFS HHs can choose a telephone interview, a point we return to below.

In the CNRLS, the ONS link survey and census records using automated and clerical HH address matching. Linkage rates are high: 93.2% of HHs in the LFS, 94.5% in the LOS, and 93.9% in the OPN (Table 1). This means we can study the majority of samples (although without non-linked HH data we cannot completely rule out dataset selection biases), using the rich suite of attribute covariates from the census (see ONS 2014b). Only HHs sampled close to the census date are included, so this information should reflect HH attributes at the time of sampling. Hence, in this case census data are of great utility as a sample attribute information source for monitoring response. We note caveats to this in other settings in the Discussion.

We consider 10 auxiliary HH attribute covariates in analyses (Table 2), chosen because analogues impact on 2001 CNRLS individual response propensities (see Durrant & Steele 2009). ‘Tenure’, ‘Accommodation type’ and ‘Cars available’ are census HH responses. ‘HH economic status’, ‘HH structure’, ‘Ill health individual in HH’, ‘Impaired individual in HH’, ‘Retiree in HH’ and ‘English fluency in HH’ are coded from individual census responses.

‘Located in London / SE’ is a geographic identifier. The first five covariates are multi-category. ‘Unknown’ indicates no response. The others are binary, with no response coded as a negative.

The call record data detail outcomes of calls (non-contact, refusal, interview) to HHs (up to 20; Table 1). They do not exist for LFS telephone interviewed HHs (~20% of the sample), and some others (~7% of the LFS sample; <1% in the LOS and OPN; Table 1). After removing these HHs, the analysed LFS dataset includes 18997 HHs, the LOS 6469 HHs, and the OPN 6249 HHs. Final response rates are 65.7% in the LFS, 70.1% in the LOS and 64% in the OPN. Analysis using the methods in the following sections suggests HHs in houses, owner HHs, HHs with retirees and all inactive HHs are under-represented in the analysed LFS dataset compared to the all linked HH dataset (results not shown), causing differences in covariate category HH proportions compared to the OPN and LOS datasets (see Table A1 in the Appendix). We consider how these impact on results in section 3.4. We also note that in practical applications of the methods detailed here focussing on improving datasets, the impacts of non-contact and refusal on representativeness must be quantified separately. The drivers of these two forms of non-response are likely to vary, as will their correlates. Hence, the impact of collection method changes on HHs (not) responding in each way will also probably differ (Durrant & Steele 2009).

2.2. Representativeness indicators

Representativeness indicators quantify survey non-response bias risks in terms of sample response propensity variation. They are not directly related to (non-response biases in) specific estimates (Schouten et al. 2012). Weighting can be applied to enable population level inference (see Roberts et al. 1987 for an introduction to the use of survey weights in propensity modelling), but here we study the linked sample (with call records). Some HHs are not linked to census data, and are excluded from analyses, as are HHs without call records, so the supplied weights would not be useful. As well, ignoring sample design is justified because our interest is not in the population but in future data collection in the surveys (with the same designs: see Phipps & Toth 2012 for similar arguments in this context).

R indicators are described by Schouten et al. (2009, 2011, 2012), and CVs by Schouten et al. (2009) and de Heij et al. (2015). The overall R indicator is the transformed (0-1) response propensity standard deviation (*SD*): $R = 1 - 2SD$, where *SD* is $\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\hat{p}_i - \hat{\bar{p}})^2}$, *n* is the sample size, $\hat{p}_i$ the sample member *i* propensity and $\hat{\bar{p}}$ the mean propensity. Large indicators imply representativeness. The overall CV is *SD* divided by $\hat{\bar{p}}$, and quantifies survey estimate mean maximum absolute standardised bias when non-response correlates maximally to the utilised auxiliary covariate set *x* (we emphasise that indicators are specific to this covariate set). Small values imply representativeness. Partial indicator decompositions allow propensity variation associated with auxiliary covariates and their categories to be quantified. Unconditional indicators measure univariate associations. The covariate of interest Z need not be in the covariate set *x*. The unconditional partial CV (we present CVs here: equivalent partial R indicators are computed by removing the $\hat{\bar{p}}$ denominator terms) for covariate Z is:

$$\widehat{CV}_u(Z, p_x) = \frac{\sqrt{\frac{1}{n}\sum_{k=1}^{K} n_k\left(\hat{\bar{p}}_k - \hat{\bar{p}}\right)^2}}{\hat{\bar{p}}}, \tag{1}$$

where n_k is the size of covariate category k, and $\hat{\bar{p}}_k$ the mean response propensity in k. Large values suggest substantial between category propensity variability and non-representativeness associated with Z. The unconditional partial CV for category *k* of covariate *Z* is:

$$\widehat{CV}_u(Z_k, p_x) = \frac{\sqrt{\frac{n_k}{n}}(\hat{\bar{p}}_k - \hat{\bar{p}})}{\hat{\bar{p}}}. \tag{2}$$

Indicators can be positive or negative, implying respectively over- or under-representation. The further they are from zero, the greater the impact. With conditional partial indicators, covariate *Z* must be in covariate set *x*. Indicators quantify non-representativeness associated with (the category of) Z conditional on other covariates, by comparing propensities given set *x* with and without Z. The conditional partial CV for covariate Z is:

$$\widehat{CV}_c(Z, p_x) = \frac{\sqrt{\frac{1}{n}\sum_{l=1}^{L}\sum_{i\in l}\left(p_i - \hat{\bar{p}}_l\right)^2}}{\hat{\bar{p}}} \tag{3}$$

where $\hat{\bar{p}}_l$ is the mean response propensity of the *l*th of L cells resulting from cross-classification of *x* excluding Z and propensity modelling given this covariate subset. The conditional partial CV for category *k* of covariate Z is:

$$\widehat{CV}_c(Z_k, p_x) = \frac{\sqrt{\frac{1}{n}\sum_{l=1}^{L}\sum_{i\in l} h_i\left(p_i - \hat{\bar{p}}_l\right)^2}}{\hat{\bar{p}}} \tag{4}$$

where h_i is an indicator detailing whether member i is in category k. In both cases, small indicators given large unconditional equivalents suggest impacts also associated with other covariates. Large indicators imply un-correlated impacts.

Adjustments to overall and partial covariate indicators exist to account for sample size related biases caused by estimating propensities. Approximate R indicator standard errors are also available, linearizing a variance estimator for SD derived by decomposing its distribution into that due to sampling design and that due to propensity model parameter estimates (Shlomo et al. 2012). For overall indicators, propensities are estimated given set x. For both partial indicators, they are estimated given a set including only Z. In addition, De Heij et al. (2014) derive overall CV standard errors, as the square root of the linearizing approximation:

$$\widehat{Var}(CV(p_x)) \cong \frac{SD^2}{\hat{\bar{p}}^2}\left[\frac{\widehat{Var}(p)}{\hat{\bar{p}}^2} + \frac{\widehat{Var}(SD)}{SD^2} - 2\frac{\widehat{Cov}(\hat{\bar{p}}, SD)}{\hat{\bar{p}}SD}\right] \tag{5}$$

where $\widehat{Var}(p)$ is the estimated variance of the mean response propensity, $\widehat{Var}(SD)$ the estimated variance of the standard deviation of propensities, and $\widehat{Cov}(\hat{\bar{p}}, SD)$ their estimated covariance. De Heij et al. assume that $\widehat{Var}(p)$ is minimal and can be approximated by SD/n, that $\widehat{Var}(SD)$, renamed $\hat{S}^2$, can be approximated by the estimator derived by Shlomo et al. (2012), and that $\widehat{Cov}(\hat{\bar{p}}, SD)$ is negligible. Given this, they re-write (5) as:

$$\widehat{Var}\left(\widehat{CV}(p_x)\right) \cong \frac{SD^2}{\hat{\bar{p}}^2}\left[\frac{SD^2}{n\hat{\bar{p}}^2} + \frac{\hat{S}^2}{SD^2}\right] = \frac{\hat{S}^2}{\hat{\bar{p}}^2} + \frac{SD^4}{n\hat{\bar{p}}^4}. \tag{6}$$

As with R indicators, overall CV standard errors are computed with *SD* estimated given the whole auxiliary covariate set. We utilise this approach to also derive partial covariate CV standard errors, using the square root of the approximation in (6) but for both unconditional and conditional indicators calculating *SD* given only Z, as with partial R indicator errors. We extend the R code of de Heij et al. (2014) to produce partial CVs and these errors (as well as R indicators, overall CVs and their errors). Our code is available on request. We note that de Heij et al. (2015) have recently similarly updated their code (a SAS version is also available: see www.risq-project.eu). Their standard errors are derived using a linearizing approximation from partial R indicator errors. We present our errors here, as sometimes those of de Heij et al. are substantially inflated. This is because R indicator errors are large when a covariate has minimal univariate impact on propensities and $\widehat{Var}(p)$ is small, due to division of $\hat{S}^2$ by $\widehat{Var}(p)$in the derivation. Indictor point estimates are computed given a multivariate propensity model, so this can occur even if the covariate impacts non-trivially on representativeness (see Fig A1 in the Appendix for errors of this type given our datasets). Beyond this, our errors are also around an order of magnitude smaller than those of de Heij et al. (results not shown).

2.3. Statistical analyses

We conduct two sets of statistical analyses. First, we investigate auxiliary covariate retention in sets for use in indicator response propensity estimation. We identify HH attribute covariates impacting on response (a successful interview) propensities after: a) five interview attempts (early in call records, when response rates are low); and b) 20 attempts (the end of data collection). We use logistic regression to model propensities, fit main effects only, and retain only those covariates for which there is an increase in the Akaike Information Criterion (AIC) of >2 on removal from the final model (see Burnham & Anderson 2002). Survey interviews may involve multiple calls: we consider the final call as the interview in these cases.

Second, we investigate representativeness indicator use to monitor data collection. For each survey, at each call we compute overall and partial R indicators and CVs given auxiliary covariate sets a) and b) identified above, and also: c) sets including all 10 covariates. To study covariate set effects, we compare point estimates by calculating differences from no model selection set values, as percentages of the latter (these sets are common comparators as they include all 10 covariates). This includes unconditional indicators for covariates not in the sets, but not conditional variants, which are only calculable for covariates in sets. We also compare overall indicator 95% CI ranges, computing intervals as the indicator ± 1.96 times its standard error and calculating differences from no model selection set ranges as percentages of the latter. We do not compare partial indicator 95% CI ranges as they are identical. As well, we consider statistical inference, studying whether overall indicator 95% CIs overlap and for partial variants also whether intervals span zero (implying (conditional) representativeness with respect to Z).

2.4. Phase capacity (PC) point identification

We identify stability based overall CV points and partial unconditional CV points for covariates linked to substantial impacts on overall dataset representativeness. Inequalities underlying partial indicators are likely targets when modifying methods as their reduction will lead to the greatest quality increases (Schouten et al. 2012; Schouten & Shlomo in press). We study information availability effects by using two identification rules: i) if CVs are within threshold a of best values over call records ('after' collection); and ii) if CVs imply quality decreases or are within *a* of the previous call value ('during'). We identify points when threshold *a* equals 0.01, 0.02 and 0.05. We also calculate the total calls made to samples saved by ending collection at overall CV points. We note that when entire call record data exist, Schouten et al. (2013) present a framework for optimising collection given alternative methods and quality-cost trade-offs, using representativeness indicators as quality measures. Points similar to our 'after' PC points, but also incorporating cost considerations, can be identified by treating them as possible alternative methods. However, such analysis is beyond the scope of this paper: for a full representation information on call costs as well as numbers is needed, which we lack.

3. Results

3.1. Response rate development

Survey household (HH) response rates increase similarly over call records, at decreasing rates with minimal increases after calls 9 to 11 (Fig. 1). The Labour Force Survey (LFS) call one response rate is higher but later increases smaller than in the Opinions Survey (OPN) and Life Opportunities Survey (LOS) (which has the highest final response rate).

3.2. Auxiliary covariate retention at different calls

In Table 3 we detail AIC based model selection to identify HH attribute covariates correlated with response propensity in the datasets after five and 20 interview attempts (the end of data collection; we present final model parameter estimates in Table A2 in the Appendix). All 10 covariates are never retained in covariate sets. Covariates retained differ both between call five and call 20 datasets and between surveys. Concerning the hypothesis that fewer covariates are retained at low response rates, as expected in the LFS and LOS fewer covariates are retained in call five sets. However, in the OPN the reverse occurs, and some covariates are also retained only in the call five set in the LOS. Hence, the hypothesis is not always supported empirically.

3.3. Representativeness indicators and auxiliary covariate set effects

Overall indicators

In Figure 1 we present survey overall R indicators and CVs over call records given no model selection auxiliary covariate sets including all 10 covariates. Indicators given the covariate sets identified in section 3.2 are similar (CVs and 95% CIs are given as differences from no model selection set values in Table 4). R indicators are initially large, implying high representativeness, decrease to call three, then increase at decreasing rates over the remaining calls (we term this the indicator trajectory). CVs decrease, implying increased representativeness, at decreasing rates over call records. Such R indicator trajectories (equivalents are seen with partial variants; see Fig. A1 in the Appendix) can arise because possible propensity variation is limited at low response rates, which is an issue when modifying methods (see Introduction). That CVs, less likely to be similarly affected by the response rate and also quantifying maximum survey estimate mean absolute standardised bias when non-response correlates maximally to the utilised auxiliary covariate set, describe different changes, suggests this is the case here. Hence, hereafter we only report these indicators.

CVs are slightly lower in the LFS than the other surveys, and initially decrease less in the LOS than the OPN. 95% CI ranges are small (~0.002 to ~0.02). CV differences given different covariate sets reach ~10% in the OPN but are mainly <4%, with CVs mostly smaller for sets with more covariates (Table 4). To investigate set misspecification effects, we compare indicators given different sets at calls five and 20, since we identify correctly specified sets including only propensity correlates at these calls in section 3.2 and Table 3. An issue is that misspecified sets often both exclude correlates and include non-correlates. Concerning impacts of excluding correlates, comparative over-estimation of representativeness is predicted. One comparison exists where non-correlates are not also included, in the LFS at call 20. The CV given the (correlates excluded) call five set is smaller than that given the call 20 set, as expected.

Including propensity non-correlates in sets should lead to comparative under-estimation of representativeness and inflated indicator errors. Comparisons where correlates are not also excluded involve no model selection set indicators at calls five and 20 and LFS call 20 set indicators at call five. As expected, CVs and 95% CIs given these sets are larger than those given correct sets, except with LFS no model selection set CVs at call five. Differences tend to be smaller than when correlates are excluded. Hence, covariate set misspecification effects are mostly, but not always, as hypothesised. Regarding statistical inference, small CV differences given different sets mean their 95% CIs rarely fail to overlap (Table 4).

Partial indicators

Overall CV decompositions suggest similar impacts on representativeness associated with HH attribute covariates in each survey. We describe these using covariate and selected covariate category partial CVs given no model selection covariate sets (Figs. 2 & 3), though we also mention covariate indicators given the other sets identified (presented as differences from no model selection set values in Tables A3 to A8 in the Appendix). 'Ill Health individual in HH', 'Impaired individual in HH' and especially 'Retiree in HH' and 'HH Economic Status' partial unconditional CVs (CV_us) are initially high, implying substantial univariate associations with response propensity variation, then decrease at decreasing rates over call records. 'Located in London / SE' CV_us are similar, though they reach minima then increase slightly in the OPN. 'HH structure' CV_us in the LOS and OPN are also similar, but in the LFS first increase slightly then decrease over call records. 'Accommodation type' CV_us decrease slightly, after first increasing in the LOS and OPN. 'Cars available' CV_us decrease, from a high initial value in the LOS, increase, then decrease slightly again. 'Tenure' CV_us first increase (less so in the LFS) then decrease slightly. 'English fluency in HH' CV_us are minimal.

Covariate category CV_us suggest 'Ill Health individual in HH', 'Impaired individual in HH', 'Retiree in HH' and 'HH Economic Status' impacts arise because HHs categorised as no in the first three cases and all employed in the last are initially under-represented in datasets (later indicator decreases imply many of these are interviewed eventually). Partial conditional CVs (CV_cs) for these covariates (categories) are mostly much smaller than CV_us, suggesting impacts are correlated (the exception are comparable OPN 'HH Economic Status' CV_us and CV_cs given the call 20 covariate set, which may be due to it excluding 'Retiree in HH'). Named categories do to an extent identify overlapping sample subgroups (for instance, retirees are unlikely to be employed), likely differing in how contactable (any) HH members are. 'Accommodation type', 'Cars available' and 'Tenure' impacts, due respectively to flats, multi-car and non-owner HHs being under-represented (not shown), possibly reflect such differences too, with CV_cs also smaller than CV_us. In addition to this (single) impact on representativeness, two covariates have large CV_us and CV_cs, implying impacts not linked to other covariates. HHs 'Located in London / SE' are under-represented. 'HH structure' impacts are due to single LFS adult HHs being under-represented and LOS and OPN couple, no children HHs being over-represented. Both these impacts are substantial at the end of data collection.

Concerning covariate partial CVs given different covariate sets, if the covariate is in both sets CV_us differ slightly (<2.5%) in ways identical to other similar covariates (see Tables A3 to A8 in the Appendix). These differences reflect (differential) indicator bias adjustment, as equivalent unadjusted values differ negligibly (results not shown). If the covariate is not in both sets, CV_u differences are often >50%, and once in the LOS ~14000%, with signs varying between covariates and over call records. CV_cs, calculable only for set members, always differ, mostly by <50%. Indicators are mostly minimal when differences are large though (~0.00001 in the LOS example), and indeed all actual differences are mainly small (reasons for OPN 'HH Economic Status' CVs are given earlier). We again use call five and 20 indicators to study set misspecification effects. If a response propensity correlate is excluded, its CV_u is mostly, but not always, comparatively under-estimated, but if a non-correlate is included effects on its CV_u vary (relevant comparisons are identifiable in Table 3). CV_us for other covariates (correlates) in sets given such exclusions / inclusions differ due to bias adjustment only, as noted above, but effects on CV_us for those (non-correlates) not in sets vary (based on the smaller relevant comparison set described in '*Overall Indicators*'). Based also on this smaller comparison set, set member CV_cs are mostly, but not always, over-estimated when correlates are excluded, and under-estimated when non-correlates are included, due to greater conditioning with larger sets. 95% CI ranges are small (~0.001 to 0.01). Regarding statistical inference, this means that indicator 95% CIs given different sets often do not (never with CV_cs) overlap. CV_u 95% CIs rarely span zero, at times doing so given one set but not others. CV_c 95% CIs never span zero.

3.4. Phase capacity (PC) points.

We present indicator stability based overall and selected partial unconditional covariate CV PC points given 'after' and 'during' data collection identification rules and different rule thresholds a in Table 5. We illustrate results using points when a equals 0.02. Overall CV 'after' rule points are later in call records than 'during' rule points, and LOS points later than LFS and OPN points, which are similar. Ending collection at these points saves the greatest percentage of the total calls made in the LOS (also see Table 5). Call savings range from 7 to 18%.

Our earlier analyses suggest three substantial impacts on dataset representativeness (see section 3.3). We present unconditional partial CV PC points for the covariates 'Located in London / SE' and 'HH structure', which are linked to separate impacts, and 'HH Economic Status' and 'Retiree in HH', which are linked to the same impact and so should have similar points. Points mostly differ from overall CV points and from each other, being earlier for the first two covariates because CVs decrease minimally over or are near minima early in the call record (points for the last two covariates are similar, as expected). Points tend to be later given 'after' than 'during' identification rules, but exceptions include OPN 'Located in London / SE' and LFS 'HH structure' (though the latter is due to a previous call value being needed with 'during' rules). Some variability exists between covariates, but points are later in the LOS than LFS and OPN, similar to overall CV points. Both overall and partial CV points exhibit similar patterns when thresholds a equal 0.01 or 0.05. Points are earlier, and more calls are saved given overall CV points, as a is increased. A qualifier to our survey comparison results is that some differences exist between analysed LFS sample attribute category proportions and those in LOS and OPN samples (see section 2.1). However, it is LOS PC points that differ from the others: LFS points should do so if sample composition differences are important.

4. Summary and Discussion

We address questions concerning the use of representativeness indicators to monitor survey non-response bias risks. We utilise a dataset linking paradata detailing attempts to interview sample households (HHs) in three UK surveys to census HH attribute information. The surveys are the Labour Force Survey (LFS), the Life Opportunities Survey (LOS) and the Opinions Survey (OPN). Indicators quantify sample estimated response propensity variation given an attribute covariate set, with low levels implying representativeness and low non-response bias risks. They are decomposable to measure variation associated with covariates, so can inform modifications to data collection methods to improve quality and / or reduce costs. Survey agencies are increasingly interested in utilising these techniques to manage non-response bias, but guidance on their use is limited, especially concerning monitoring during data collection.

To begin with, indicators require attribute covariates for all sample units: response propensities are statistically modelled. For the first time, we use linked census data, in the UK the only information source for non-longitudinal surveys. These data are of great utility in our non-response bias analyses. HH linkage rates are around 94%, so the majority of samples can be analysed (though without non-linked HH data we cannot completely rule out selection biases). The available covariate set is rich, and samples are from within three months of the census, so information will be mostly accurate at the time of survey sampling. Concerning guidance to survey practitioners though, such timeliness is also why we advise caution before using census data more widely for this purpose. The UK census is decadal. How HH linkage rates and covariate accuracy decreases for samples further away from the census date, reducing data source utility, is unknown. To investigate this, surveys from these dates must be linked.

We also consider indicator use to monitor data collection. First, R indicators can suggest representativeness is highest early in call records because possible response propensity variation is limited at low response rates. CVs have potentially superior properties as they are less likely to be similarly affected by response rates, and also quantify maximum survey estimate mean absolute standardised bias when non-response correlates maximally to the utilised auxiliary covariate set (Schouten et al. 2009). We compare indicators in surveys. R indicators behave as described, but CVs suggest representativeness increases at decreasing rates over call records. This implies that inferences from R indicators are indeed affected by response rates, so we base further explorations on CVs. A barrier to this previously has been that they were less decomposable, but recently partial variants have been presented (de Heij et al. 2015; Correa et al. 2016). We present approximate partial covariate CV standard errors, by extending the use of overall CV error approximation of de Heij et al. (2014). Unlike similar errors derived by de Heij et al. (2015), by approximating from the partial R indicator error, our estimators are not sometimes inflated (see section 2.2 for details). More generally, comparable differences in indicator behaviour arise in other surveys (Lundquist & Särndal 2013; Correa et al. 2016). Hence, concerning guidance to survey practitioners, now that similar functionality exists we recommend that CVs are used to monitor response over call records and in other scenarios where paradata on data collection over time are available (such as mail in / mail back surveys and web surveys).

Second, there are issues specifying indicator auxiliary covariate sets for use over call records. The same set must be fitted to data at each call for indicators to be informative. Only propensity correlates should be included, else accuracy is affected, but for a given sample size model selection should also lead to reduced covariate retention at low response rates (Schouten et al. 2009; Shlomo et al. 2012). To advise on selecting covariate sets given such considerations, we study covariate retention across calls and misspecification effects (excluding available correlates, including non-correlates) on indicators. Regarding covariate retention, in the LFS and LOS fewer are retained in sets given early call datasets than end of collection datasets, as predicted. However, in the OPN the opposite occurs, and also some LOS covariates are only retained given the early dataset. These latter results occur, as covariate (category) partial CVs show (see section 3.3 and also below), because eventually HHs in under-represented categories are interviewed and category response propensities equalise. Such relationships likely often arise in surveys, and mean that correct specification of covariate sets (including only correlates) at different calls may vary due to changes in covariate impacts as well as the response rate. This makes it even more difficult to choose sets not misspecified over parts of the call record.

Regarding set misspecification, correlate exclusion should lead comparative over-estimation of representativeness, non-correlate inclusion to the opposite and inflated errors. Indicators given the sets above (and sets with all 10 covariates) at calls when sets are identified and correlates known are mostly consistent with these predictions. Differences between sets are small, and CV 95% CIs mainly overlap. Effects are larger given correlate exclusion. Partial CVs suggest substantial impacts on representativeness are under-representation of less contactable HHs (all employed HHs, no retiree, ill health and impaired individual HHs, which are overlapping groups), which declines over call records, of HHs in London / SE and of single adult HHs. Covariate set differences vary (mostly again being small, though often 95% CIs do not overlap), but excluded correlate unconditional CVs are under-estimated, and included non-correlate conditional CVs under-estimated. Concerning guidance to survey practitioners, we hence recommend that all available covariates are included in sets used to estimate response propensities. Any set is likely to be misspecified over part of the call record, but effects on indicators are mainly small and larger if correlates are excluded (overall representativeness is relatively more over-estimated, partial unconditional CVs, used to identify associations then investigated with conditional forms, are under-estimated). Therefore, there will be little gain in excluding non-correlates from sets (notwithstanding under-estimated conditional covariate impacts), and potentially costs since in the process sometime correlates may be excluded.

In addition, we study design phase capacity (PC) points, when current methods lead to minimal further quality increases (or decreases) and modifications should be considered (e.g. Groves & Heeringa 2006). We identify CV stability based points compared to best values over call records ('after' rules), and also previous call values ('during'), with rule thresholds of 0.01 to 0.05. Partial CV points for covariates linked to substantial impacts on representativeness (see earlier for details) differ from overall CV points and also between (non-correlated) covariates. This is to be expected given that they measure different inequalities. In applications, we recommend that overall CV points are used to identify when phase capacity is reached if collection is to be ended completely, as they reflect overall quality. Partial points like the ones described (and at the category level) are more of interest when modifying methods to improve quality. Impacts identified are likely targets as reducing underlying inequalities will lead to the largest improvements (for approaches to using such results to design modifications, see Schouten et al. 2012; Schouten & Shlomo in press). In this context, sometimes phase capacity decisions may best be based on these points (for example, if quality decreases), and / or targeted groups may be treatable separately (see also Groves & Heeringa 2006; Schouten et al. 2013).

Identified overall PC points range from calls 4 to 11, being earlier in call records as rule thresholds increase. This suggests that in the studied surveys collection (currently up to 20 calls) can indeed be ended early with limited increases in non-response bias risks. Of note to survey agencies interested in utilising these methods to manage risks, call savings made by ending collection at such points compared to analysed sample totals range from 7 to 18% when thresholds equal 0.02 (and increase with threshold size). As well, 'after' points, so named because they are identifiable after collection to inform future periods, tend to be later in call records than 'during' points, which are identifiable during collection as in situations when no historic information exists (e.g. Groves & Heeringa 2006). This is due to small CV decreases arising from the last responses obtained, which given CV derivation occur even if propensity variation remains similar (see also Lundquist & Särndal 2013). Practically, such a finding means 'during' rules identify points at CV values that decrease further with continued effort than 'after' rules, a detail to be considered when information availability is an issue.

Finally, we compare PC points across surveys, to provide guidance on whether they can be generalized from one survey to others. This is appealing to survey agencies given issues linking sample information and monitoring costs. We find that LOS overall CV points are one to two calls later than LFS and OPN points. Covariate partial CV points are broadly similar. This suggests that generalisation could be difficult, even when, as here, surveys are of the same sample frame (some differences between analysed samples exist, but do not affect conclusions: see section 3.4). If LFS or OPN points are used, LOS data collection will not achieve the desired CV stability. If LOS points are used, LFS and OPN collection will not be optimally efficient. As well, without complete knowledge errors cannot be identified. Consequently, though we again note the potential benefits of employing these techniques when monitoring data collection in a given survey, we end by recommending that confirmatory work is undertaken before generalising PC points from one survey to another.

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.

References

Burnham, K.P. and Anderson, D.R. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd edition, Springer-Verlag.

Correa, S, Durrant, G.B. & Smith, P.W.F. (2016) Assessing non-response bias using call record data with applications in a longitudinal study. *S3RI Technical paper, University of Southampton. Southampton, UK, 38pp.*

de Leeuw, E. and 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.

De Heij, V., Schouten, B. & Shlomo, N. (2014) RISQ Manual 2.0: Tools in SAS and R for the computation of R indicators and partial R indicators. Available from: www.risq-project.eu

De Heij, V., Schouten, B. & Shlomo, N. (2015) RISQ Manual 2.1: Tools in SAS and R for the computation of R indicators and partial R indicators. Available from: www.risq-project.eu

Durrant, G. B., & Steele, F. (2009). Multilevel modelling of refusal and non-contact in household surveys: evidence from six UK Government surveys. *J. R. Statist. Soc. A*, 172, 361-381. DOI: 10.1111/j.1467-985X.2008.00565.x

Durrant, G.B., Groves, G., Staetsky L. and Steele, F. (2010): Effects of Interviewer Attitudes and Behaviours on Refusal in Household Surveys. *Public Opinion Quarterly*, 74, 1, 1-36. DOI: 10.1093/poq/nfp098

Durrant, G.B., D'Arrigo, J. and Steele, F. (2011): Using Field Process Data to Predict Best Times of Contact Conditioning on Household and Interviewer Influences, *J. R. Statist. Soc. A*, 174, 4, 1029-1049. DOI: 10.1111/j.1467-985X.2011.00715.x

Durrant, G.B., D'Arrigo, J. and Steele, F. (2013): Analysing Interviewer Call Record Data by Using a Multilevel Discrete-Time Event History Modelling Approach, *J. R. Statist. Soc. A*, 176, 251-269. DOI: 10.1111/j.1467-985X.2012.01073.x

Groves, R. M. (2006) Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, 70, 646-675. DOI: 10.1093/poq/nfl033

Groves, R. M. and Heeringa, S. (2006) Responsive design for household surveys: tools for actively controlling survey errors and costs. *J. R. Statist. Soc. A*, 169, 439-457. DOI: 10.1111/j.1467-985X.2006.00423.x

Groves, R. M. and Peytcheva, E. (2008) The impact of non-response rates on non-response bias: a meta-analysis. *Public Opinion Quarterly*, 72, 167–189. DOI: 10.1093/poq/nfn011

Luiten, A. & Schouten, B. (2013) Tailored fieldwork design to increase representative household survey response: an experiment in the Survey of Consumer Satisfaction. *J. R. Statist. Soc. A*, 176: 169-189. DOI: 10.1111/j.1467-985X.2012.01080.x

Lundquist, P. and Särndal, C-E. (2013) Aspects of responsive designs with applications to the Swedish Living Conditions Survey. *Journal of Official Statistics*, 29, 557-582. DOI: 10.2478/jos-2013-0040

Kappelhof, J.W.S. (2013) The effect of different survey designs on nonresponse in surveys among non-western minorities in The Netherlands. *Survey Research Methods* 8: 81-98. DOI: 10.1.1.679.9232

Kreuter, F. (2013) Facing the non-response challenge. *Annals of the American Academy of Political and Social Science* 645: 23-35. DOI: 10.1177/0002716212456815

Massey, D.S. & Tourangeau, R. (2013) Where do we go from here: non-response and social measurement. *Annals of the American Academy of Political and Social Science* 645: 222-236. DOI: 10.1177/0002716212464191

Olson, K. (2006) Survey participation, nonresponse bias, measurement error bias, and total bias. *Public Opinion Quarterly*, 70, 737–758. DOI: 10.1093/poq/nfl038

Ouwehand, P. & Schouten, B. (2014) Measuring representativeness of short term business statistics. *Journal of Official Statistics* 30: 623-649. DOI: 10.2478/jos-2014-0041

ONS (2011a) LFS User Guide – Volume 1: Background and Methodology. Available at:

http://discover.ukdataservice.ac.uk/catalogue/?sn=6782&type=Data%20catalogue

ONS (2011b) Opinions Survey – Technical Report – March 2011. Available at: http://discover.ukdataservice.ac.uk/catalogue/?sn=7166&type=Data%20catalogue

ONS (2014a) Life Opportunities Survey Waves 1-2, 2009-2012: ONS report. Available at: http://discover.ukdataservice.ac.uk/catalogue/?sn=6653&type=Data%20catalogue

ONS (2014b) 2011 census variables: part 1. Available at: http://www.ons.gov.uk/census

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

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. DOI: 10.18148/srm/2010.v4i1.3037

Phipps, P. & Toth, D. (2012) Analysing establishment nonresponse using an interpretable regression tree model with linked administrative data. *Annals of Applied Statistics*, 6, 722-794. DOI: 10.1214/11-AOAS521.

Rao, R. S., Glickman, M. E., & Glynn, R. J. (2008). Stopping rules for surveys with multiple waves of nonrespondent follow-up. *Statist. Med.*, 27, 2196-2213. DOI: 10.1002/sim.3063

Roberts, G., Rao, J.N.K. & Kumar, S. (1987) Logistic regression analysis of sample survey data. *Biometrika*, 74, 1-12.

Särndal, C-E. and Lundquist, P. (2014) Balancing the response and adjusting estimates for nonresponse bias: complementary activities. *Journal de la Société Française de Statistique*, 155, 28-50.

Schouten, B., & Shlomo, N. (in press). Selecting Adaptive Survey Design Strata with Partial R-indicators. *Int. Statist. Rev.*

Schouten, B., Cobben, F. and Bethlehem, J. (2009) Indicators for the Representativeness of Survey Response. *Survey Methodology*, 35, 101-113.

Schouten, B., Shlomo, N. and Skinner, C. (2011) Indicators for monitoring and improving representativeness of response. *Journal of Official Statistics*, 27, 231-253.

Schouten, B., Bethlehem, J., Beullens, K., Kleven, Ø., Loosveldt, G., Luiten, A., Rutar, K., Shlomo, N. and Skinner, C. (2012) Evaluating, comparing, monitoring, and improving representativeness of survey response through R-indicators and partial R-indicators. *Int. Statist. Rev.*, 80, 382-399. DOI: 10.1111/j.1751-5823.2012.00189.x

Schouten, B., Calinescu, M. and Luiten, A. (2013) Optimizing quality of response through adaptive survey designs. *Survey Methodology*, 39, 29-58.

Schouten, B., Cobben, F., Lundquist, P., & Wagner, J. (in press). Does more balanced survey response imply less non-response bias? *J. R. Statist. Soc. A*

Shlomo, N., Skinner, C.J. & Schouten, B. (2012) Estimation of an indicator of the representativeness of survey response. *J. Statist. Planng. Inf.* 142, 201-211. DOI:10.1016/j.jspi.2011.07.008

Steele, F. and Durrant, G. (2011) Alternative Approaches to Multilevel Modelling of Survey Noncontact and Refusal, *Int. Statist. Rev.*, 79, 70-91. DOI: 10.1111/j.1751-5823.2011.00133.x

Wagner, J.R. (2008) *Adaptive Survey Design to Reduce Nonresponse Bias.* PhD diss., University of Michigan, Michigan.

Wagner, J.R, & Raghunathan, T. E. (2010). A new stopping rule for surveys. *Statist. Med.*, 29, 1014-1024. DOI: 10.1002/sim.3834

Wagner, J.R. (2012) A Comparison of Alternative Indicators for the Risk of Nonresponse Bias. *Public Opinion Quarterly*, 76, 555–75. DOI:10.1093/poq/nfs032

Table 1: Dataset construction and content. 'Linked to census', 'Face to face interview' and 'With call records' are the number of (remaining) HHs with such characteristics, the latter being the analytical dataset sizes. 'Interviewed', 'Refusal' and 'Non-contact' are numbers of outcomes in call 20 analytical datasets. We also present the number of calls made, as means and standard deviations (in brackets) per HH and per successful interview.

	LFS	LOS	OPN
Eligible HHs	27378	6896	6668
Linked to Census	25524	6521	6260
Face to face interview	20514	6521	6260
With call records	18997	6469	6249
Interviewed (response)	12480	4533	3997
Refusal	1902	567	672
Non-contact	4615	1369	1580
Calls Per HH	8.67 (8.34)	8.32 (7.86)	9.33 (8.22)
Calls per successful interview	2.75 (1.95)	3.33 (2.26)	3.32 (2.32)

Table 2: Auxiliary household attribute covariates considered in the analyses, and categorisations.

Covariate	Categories
HH Economic Status	1) All employed; 2) All unemployed; 3) All inactive; 4) Mixed; 5) Unknown.
HH structure	1) One adult; 2) One adult, children; 3) Couple, no children; 4) Couple, children; 5) >Two adults, children or otherwise; 6) Unknown.
Accommodation type	1) House; 2) Flat; 3) Other; 4) Unknown.
Tenure	1) Owned; 2) Rented / other; 3) Unknown.
Cars available	1) None; 2) One car; 3) Two cars; 3) Three or more cars; 4) Unknown.
Ill health individual in HH	1) No; 2) Yes.
Retiree in HH	1) No; 2) Yes.
Located in London / South East	1) No; 2) Yes.
Impaired individual in HH	1) No; 2) Yes.
Anyone fluent in English in HH	1) Yes; 2) No.

Table 3: Covariates retained in logistic regression models of response propensity following AIC based model selection on call five and call 20 datasets in each survey. AICs in underlined text indicate covariates retained in models, AICs in normal text indicate covariates not retained.

	LFS		LOS		OPN	
	Call 5	Final	Call 5	Final	Call 5	Final
HH Economic Status	25195	24110	8408.5	7736.4	8480.8	8070.8
HH structure	25348	24220	8409.8	7742.0	8472.8	8060.0
Accommodation type	25192	24117	8408.2	7731.7	8462.0	8044.2
Tenure type	25176	24112	8407.7	7736.0	8461.4	8048.5
Cars available	25176	24109	8399.4	7733.3	8472.2	8060.7
Located in London / SE	25236	24180	8421.3	7762.7	8474.8	8097.2
English Fluency in HH	25177	24109	8399.1	7730.3	8461.8	8048.4
Impaired individual in HH	25178	24109	8400.2	7734.3	8459.8	8054.4
Ill Health individual in HH	25175	24114	8399.2	7732.3	8465.1	8045.8
Retiree in HH	25199	24116	8426.9	7746.5	8465.4	8047.3
Final model AIC	25176	24107	8399.2	7731.1	8459.0	8045.7

Table 4: Percentage differences between calls five (‘CV5’) and 20 (‘CV20’) auxiliary covariate set overall CVs and no model selection set values (‘CV’) in each survey. We also present similar indicator 95% CI range differences (‘CI-‘ and ‘CI+’ detail no model selection set range, ‘CI5’ and ‘CI20’ calls 5 and 20 set differences).

	LFS							LOS							OPN						
Call	CV	CV5	CV20	CI-	CI+	CI5	CI20	CV	CV5	CV20	CI-	CI+	CI5	CI20	CV	CV5	CV20	CI-	CI+	CI5	CI20
1	0.378	-1.60	-0.18	0.373	0.383	-3.23	-0.55	0.460	-3.02	1.48	0.450	0.471	-8.56	-2.11	0.483	0.37	-4.04	0.471	0.494	-1.06	-9.81
2	0.241	-0.14	-0.05	0.237	0.244	-1.08	-0.28	0.369	0.29	0.06	0.360	0.379	-0.96	-0.72	0.310	-0.11	-5.57	0.302	0.318	-0.89	-7.94
3	0.178	-0.01	0.06	0.176	0.181	-1.04	-0.18	0.266	-0.10	0.08	0.259	0.273	-1.01	-0.58	0.225	0.06	-8.29	0.219	0.231	-0.76	-9.24
4	0.147	-0.36	0.01	0.145	0.149	-1.49	-0.25	0.199	-0.27	-1.35	0.194	0.204	-1.26	-1.94	0.179	-0.30	-6.21	0.174	0.183	-1.16	-7.30
5	0.127	-0.90	0.06	0.125	0.129	-2.12	-0.25	0.166	-0.56	-1.80	0.162	0.170	-1.66	-2.53	0.150	0.01	-5.22	0.146	0.154	-1.02	-6.66
6	0.116	-1.99	0.07	0.114	0.117	-3.18	-0.26	0.145	-1.83	-1.79	0.142	0.149	-2.90	-2.85	0.131	-0.16	-3.34	0.127	0.134	-1.28	-5.44
7	0.110	-2.14	0.17	0.108	0.111	-3.38	-0.19	0.135	-1.83	-1.93	0.131	0.138	-3.08	-3.15	0.125	-0.44	-3.37	0.122	0.129	-1.48	-5.52
8	0.105	-2.17	0.17	0.103	0.106	-3.50	-0.21	0.128	-2.23	-1.79	0.125	0.131	-3.49	-3.19	0.121	-0.19	-2.53	0.118	0.124	-1.31	-4.95
9	0.102	-2.29	0.14	0.101	0.103	-3.66	-0.25	0.123	-2.07	-1.14	0.120	0.126	-3.47	-2.69	0.118	-0.59	-2.67	0.115	0.121	-1.62	-5.13
10	0.100	-2.39	0.13	0.098	0.101	-3.80	-0.27	0.121	-1.86	-1.22	0.118	0.124	-3.32	-2.80	0.117	-0.49	-2.19	0.114	0.120	-1.52	-4.75
11	0.098	-2.37	0.18	0.097	0.100	-3.82	-0.23	0.119	-2.05	-1.00	0.116	0.122	-3.52	-2.61	0.116	-0.62	-2.08	0.113	0.119	-1.63	-4.69
12	0.099	-2.41	0.18	0.097	0.100	-3.84	-0.23	0.116	-2.21	-1.22	0.113	0.119	-3.73	-2.87	0.116	-0.57	-1.85	0.113	0.119	-1.59	-4.51
13	0.098	-2.40	0.17	0.097	0.100	-3.83	-0.24	0.116	-2.19	-1.17	0.113	0.119	-3.74	-2.83	0.115	-0.66	-1.35	0.112	0.118	-1.66	-4.11
14	0.098	-2.35	0.17	0.097	0.099	-3.80	-0.24	0.113	-2.36	-1.12	0.111	0.116	-3.96	-2.85	0.115	-0.58	-1.20	0.112	0.118	-1.60	-3.98
15	0.098	-2.38	0.17	0.097	0.099	-3.83	-0.24	0.113	-2.30	-1.09	0.111	0.116	-3.92	-2.83	0.114	-0.64	-1.25	0.111	0.117	-1.65	-4.05
16	0.098	-2.42	0.17	0.096	0.099	-3.87	-0.25	0.112	-2.40	-1.10	0.110	0.115	-4.04	-2.85	0.114	-0.64	-1.25	0.111	0.117	-1.65	-4.05
17	0.098	-2.38	0.21	0.096	0.099	-3.84	-0.21	0.112	-2.38	-1.21	0.109	0.114	-4.04	-2.98	0.114	-0.64	-1.25	0.111	0.117	-1.65	-4.05
18	0.098	-2.38	0.21	0.096	0.099	-3.84	-0.21	0.112	-2.39	-1.22	0.109	0.114	-4.05	-3.00	0.114	-0.66	-1.27	0.111	0.117	-1.67	-4.05
19	0.098	-2.38	0.21	0.096	0.099	-3.84	-0.21	0.111	-2.55	-1.20	0.108	0.114	-4.20	-3.00	0.114	-0.66	-1.27	0.111	0.117	-1.67	-4.05
20	0.098	-2.38	0.21	0.096	0.099	-3.84	-0.21	0.111	-2.62	-1.23	0.108	0.113	-4.26	-3.03	0.114	-0.56	-1.35	0.111	0.117	-1.59	-4.12

NB: Underlined ‘CV’ indicates that no model selection and call 5 set 95% CIs do not overlap, underlined ‘CV5’ that calls 5 and 20 set 95% CIs do not overlap, and underlined ‘CV20’ that no model selection and call 20 set 95% CIs do not overlap.

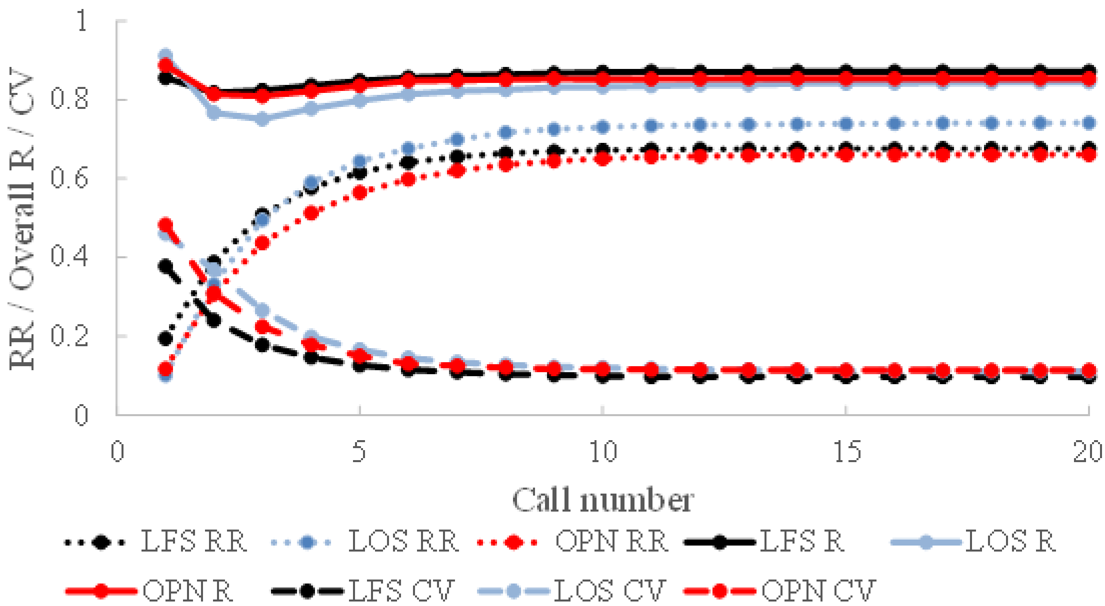
Table 5: Overall and partial unconditional covariate CV PC points in surveys given 'after' and 'during' identification rules and three rule thresholds *a*. Indicator auxiliary covariate sets include all 10 HH attribute covariates. For overall CV points, we also present (in brackets) the percentage of the total calls made to the sample saved by ending collection after the call.

	a = 0.01		*a* = 0.02		*a* = 0.05	
Survey	After	During	After	During	After	During
LFS						
Overall	8 (2.9%)	7 (4.7%)	6 (7.6%)	5 (12.2%)	4 (19.0%)	4 (19.0%)
Econ_stat	8	7	6	5	4	4
HH_struct	7	2	1	2	1	2
Retiree	8	7	6	6	5	4
LDN / SE	5	5	4	4	3	3
LOS						
Overall	11 (9.5%)	8 (15.2%)	8 (15.2%)	7 (18.2%)	6 (22.0%)	5 (27.0%)
Econ_stat	9	8	7	6	6	5
HH_struct	7	6	6	6	4	3
Retiree	10	8	8	7	6	5
LDN / SE	6	3	4	3	3	2
OPN						
Overall	8 (6.8%)	7 (9.7%)	6 (13.4%)	6 (13.4%)	5 (18.8%)	4 (26.4%)
Econ_stat	9	7	7	6	5	5
HH_struct	7	4	5	4	3	3
Retiree	10	7	8	7	6	5
LDN / SE	3	4	3	4	2	2

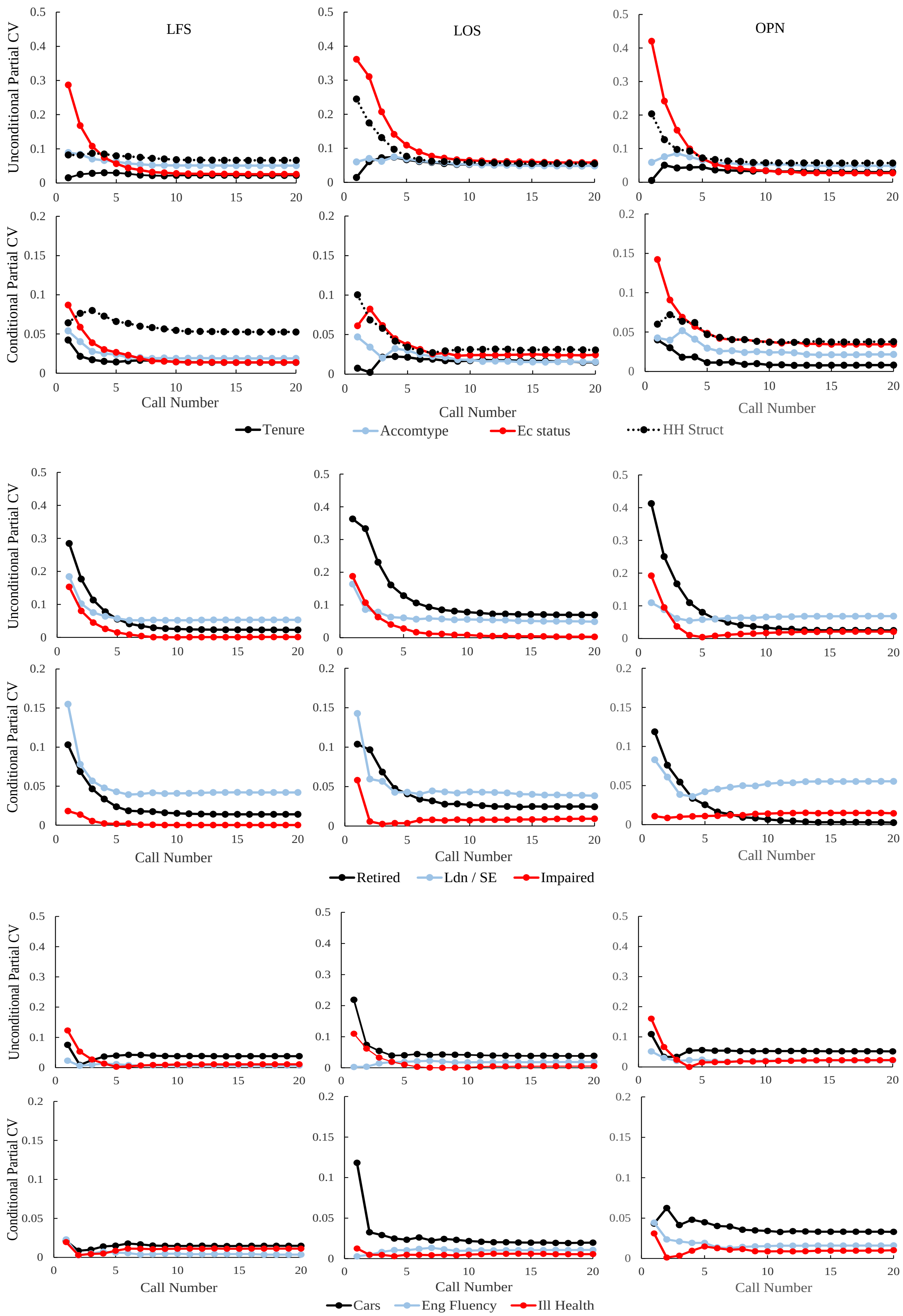
Figure 1: Cumulative response rates, and overall R indicators and CVs over call records in the three surveys given auxiliary covariate sets including all 10 HH attribute covariates.

Figure 2: Unconditional and conditional partial by covariate CVs over call records in the three surveys for HH attribute covariates given auxiliary covariate sets including all 10 HH attribute covariates. Each column of graphs details indicators for a survey.

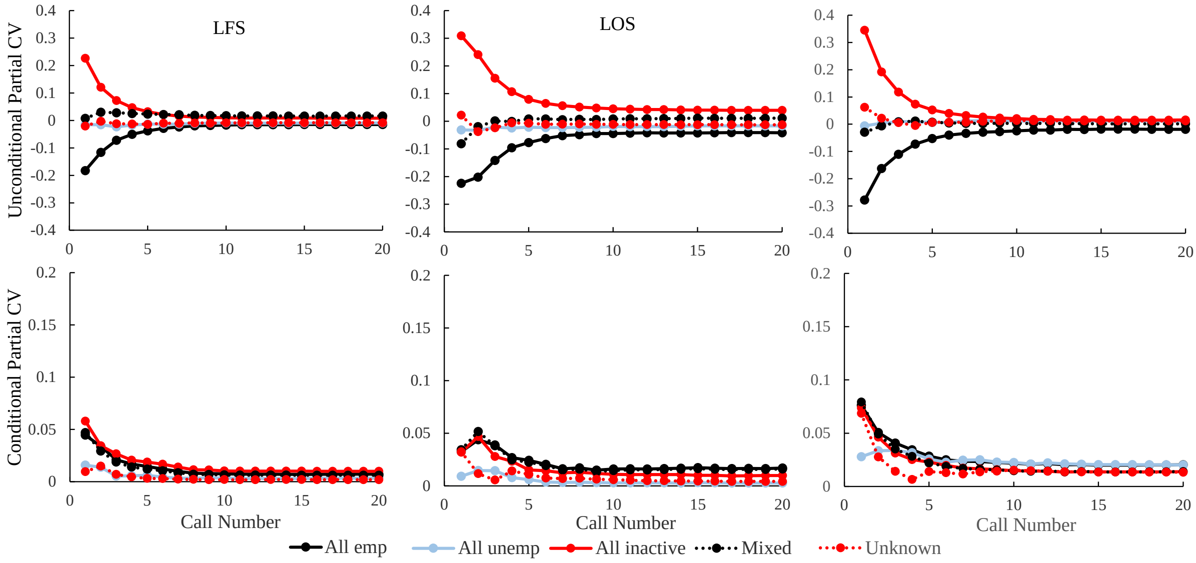
Figure 3: Unconditional and conditional covariate category CVs over call records in surveys given auxiliary covariate sets including all 10 HH attribute covariates, for : a) HH Economic Status; b) HH structure; c) Located in London / SE; d) Ill health individual in HH; e) Impaired individual in HH; and f) Retiree in HH. Each column of graphs details indicators for a survey.



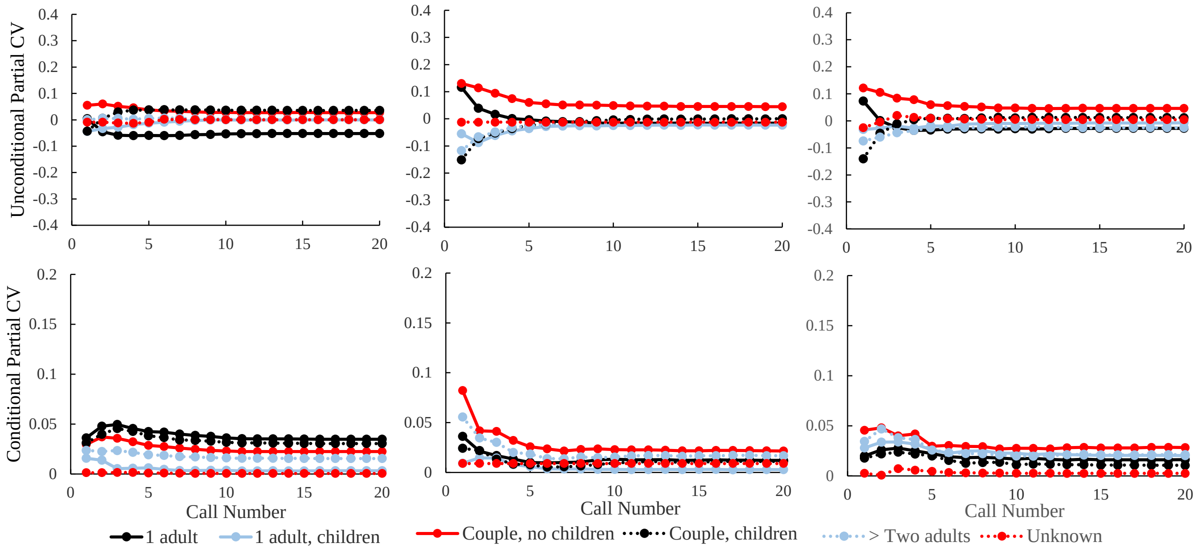
1
0.8
0.6
0.4
0.2
0
RR / Overall R / CV
0
5
10
15
20
Call number
LFS RR
LOS RR
OPN RR
LFS R
LOS R
OPN R
LFS CV
LOS CV
OPN CV


LFS
LOS
OPN
Unconditional Partial CV
Conditional Partial CV
Call Number
Tenure
Accomtype
Ec status
HH Struct
Retired
Ldn / SE
Impaired
Cars
Eng Fluency
Ill Health

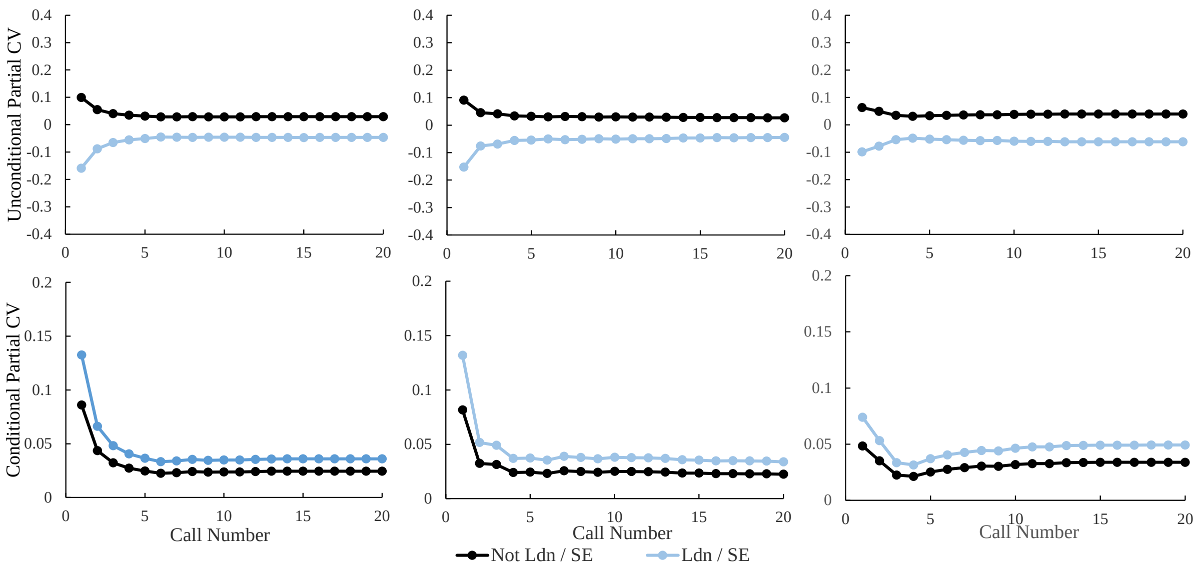
a)


LFS
LOS
Unconditional Partial CV
Conditional Partial CV
Call Number
All emp
All unemp
All inactive
Mixed
Unknown

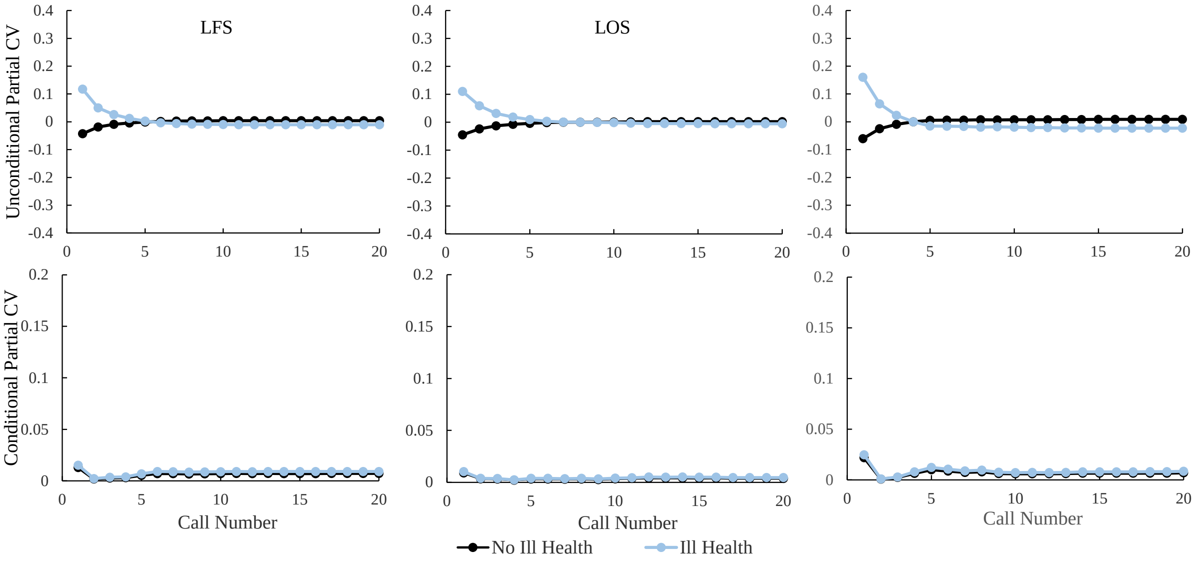
b)


Unconditional Partial CV
Conditional Partial CV
Call Number
1 adult
1 adult, children
Couple, no children
Couple, children
> Two adults
Unknown

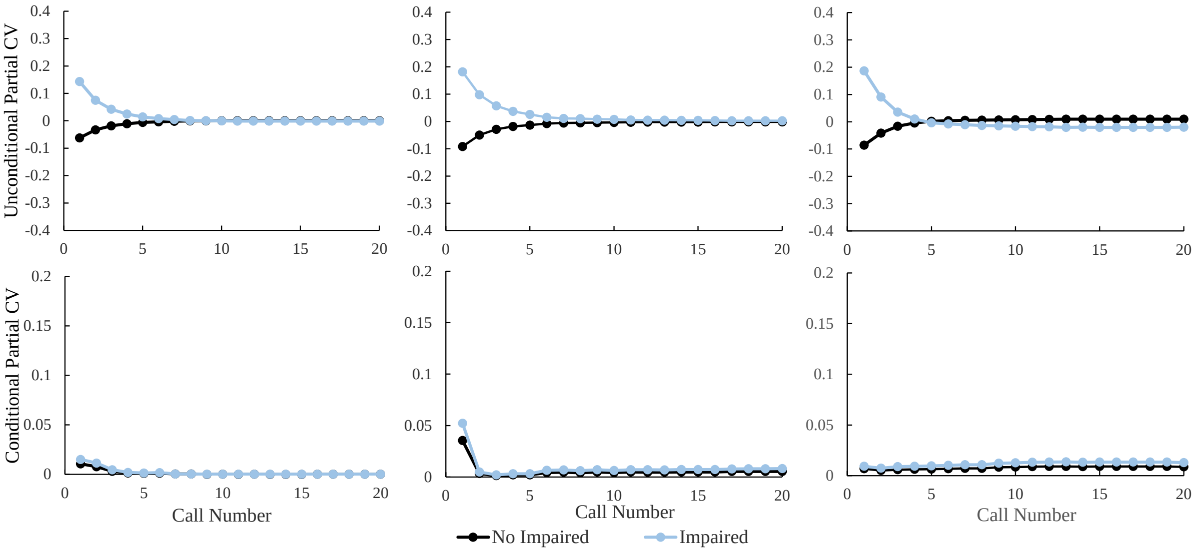
c)


Unconditional Partial CV
Conditional Partial CV
Call Number
Not Ldn / SE
Ldn / SE

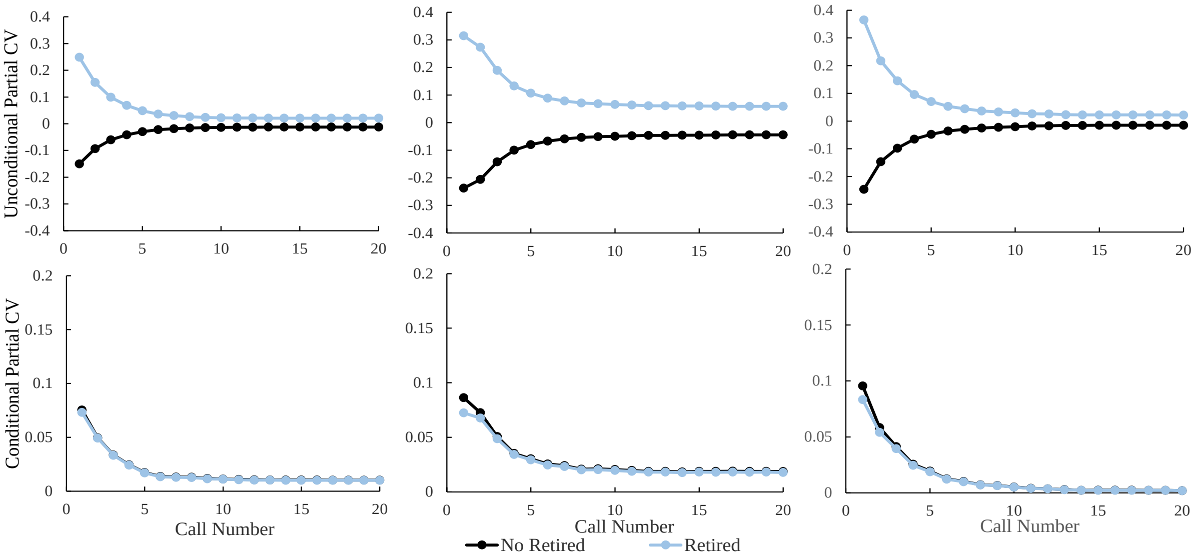
d)


LFS
LOS
Unconditional Partial CV
Conditional Partial CV
Call Number
No Ill Health
Ill Health

e)


Unconditional Partial CV
Conditional Partial CV
Call Number
No Impaired
Impaired

f)


Unconditional Partial CV
Conditional Partial CV
Call Number
No Retired
Retired