Aspects of Responsive Design with Applications to the Swedish Living Conditions Survey

Peter Lundquist[1] and Carl-Erik Särndal[2]

In recent literature on survey nonresponse, new indicators of the quality of the data collection have been proposed. These include indicators of balance and representativity (of the set of respondents) and distance (between respondents and nonrespondents), computed on available auxiliary variables. We use such indicators in conjunction with paradata from the Swedish CATI system to examine the inflow of data (as a function of the call attempt number) for the 2009 Swedish Living Conditions Survey (LCS). We then use the LCS 2009 data file to conduct several "experiments in retrospect". They consist in interventions, at suitable chosen points and driven by the prospects of improved balance and reduced distance. The survey estimates computed on the resulting final response set are likely to be less biased. Cost savings realized by fewer calls can be redirected to enhance quality of other aspects of the survey design.

Key words: Household surveys; nonresponse; auxiliary vector; register variables; stopping rules; balance indicators; representativeness; *R*-indicator.

1. Introduction

Large nonresponse is typical of many sample surveys today. This can be a serious detriment to survey quality. Nonresponse causes systematic error, called bias, in the survey estimates. The purpose of this article is to define and apply new tools, in the spirit of responsive design, to the Swedish Survey of Living Conditions (LCS), so as to improve the data collection for this important survey, which has become affected by high nonresponse in recent years.

An extensive literature is devoted to survey nonresponse and its consequences. In dealing with the problem, statisticians need to consider (a) measures to be taken at the data collection stage, and (b) measures to be taken at the estimation stage.

With the data collection completed, the estimation stage begins, and the statistician's task is to produce estimates that are properly adjusted for the nonresponse bias still remaining, despite efforts to achieve balance or representativity at the data collection stage. The objective at the estimation stage is to achieve the best possible reduction of a nonresponse bias that can never be completely eliminated. One way to do this is by adjustment weighting, through calibration on selected auxiliary variables. Nonresponse

[1] Senior Methodologist, Statistics Sweden, Karlavägen 100, 104 51 Stockholm, Sweden. Email: peter.lundquist@scb.se
[2] Visiting Professor, Statistics Sweden, 70189 Örebro, Sweden and Örebro University. Email: carl.sarndal@scb.se

weighting adjustment has been studied in several publications, including Särndal and Lundström (2005, 2008, 2010), Särndal (2011b).

The focus in this article lies on the data collection. The nonresponse rate measures one aspect of the data collection. It has become increasingly clear that the nonresponse rate by itself is not suitable, or at least not sufficient, for effective monitoring of the data collection. For example, it may be wasteful to continue a data collection according to an unchanging scenario driven primarily by the desire to obtain the highest possible response rate in the end, or to reach, by a costly and unrelenting effort, a predefined rate of response, such as 70% for example.

Wagner (2012 p. 557) expresses the dilemma as follows: "To the extent that response rates are not a good indicator for nonresponse bias, decisions about data-collection activities or post-survey adjustments that are made based on the response rate will be inefficient, biasing or both. Something is needed to fill this gap between response rates – which are known – and nonresponse biases – which are unknown, but are the thing about which we are really concerned." In the typology of data sources in Wagner (2012), the type that describes our approach is "the response indicator and frame data/paradata."

Two important recent concepts with implications for this article are *adaptive design* and *responsive design*. Bethlehem et al. (2011) regard responsive design as a special case of adaptive design.

At the present stage of development, adaptive design appears to refer mainly to situations where treatments applied to sampled elements are identified prior to the start of the data collection, although they may also be revised or modified during the data collection.

Responsive design is an adaptive approach where available information is used to modify the data collection for the remaining cases. The data collection may thus involve two or more phases, with decisions taken underway about subsequent phases. The general objectives of responsive design are formulated in Groves and Heeringa (2006). A number of applications of related approaches have subsequently appeared. Options for responsive design in a Canadian setting are discussed in Mohl and Laflamme (2007) and Laflamme (2009). Work on the development of adaptive designs has been presented for example in Wagner (2008). The present article draws mainly on the ideas of responsive design. Groves and Heeringa (2006) use the term "phase capacity" for "the stable condition of an estimate in a specific design phase". When phase capacity has been reached in a given phase, it is no longer effective to continue data collection in the same mode or phase; there is an incentive to modify the design, if data collection is to be continued at all.

Several directions have emerged in recent years in research on adaptive designs. The question whether a definite relationship exists between nonresponse rates and bias in the estimates is reviewed in Groves (2006). A meta-analysis on nonresponse studies is reported in Peytcheva and Groves (2009). The conclusion, somewhat pessimistic about the bias-reducing effect of demographic auxiliary variables, is that there is no strong evidence that variation in response rates across sample groups can help reduce biases in the study variables.

In the Scandinavian countries, the choice of auxiliary variables is much broader. Indicators for the data collection were developed in Schouten et al. (2009) and in Särndal (2011a). We apply the indicators to an existing data set: that of the 2009 Swedish Living

Conditions Survey (LCS). Our objective is to demonstrate how the indicators work and to suggest improvements for the data collection in future versions on the Swedish LCS.

Stopping rules for the data collection have been studied in Rao et al. (2008) and in Wagner and Raghunathan (2010).

An approach with obvious appeal is to observe changes in survey estimates, for variables that allow this, as a function of making additional contact attempts. It is one of the techniques used in this article. Related to this is the question of whether respondents interviewed early (say, in the first five attempts) differ considerably from those brought in later with respect to important measurable variables. These questions are studied in Peytchev et al. (2009) and in Peytchev et al. (2010). They conclude that focusing on groups with low response probability may not be efficient in some surveys; it may be better to identify those units with the greatest potential to induce bias in the survey variable estimates.

Responsive design may take different forms. One option is to strive for an ultimate set of respondents with measurable and favorable characteristics for the set of respondents. Especially in the later stages of the data collection intervention is permissible in order to realize an ultimate response set that is better balanced, or more representative of the total sample, than if no special effort is made. Recently proposed indicators for balance and representativity are important in this process; they are used in this article to monitor the data collection and to implement changes. Both concepts build on a specified auxiliary vector with values known for the full sample.

The 7th EU Framework Programme funded a project called RISQ, which stands for Representativity Indicators for Survey Quality; on it, see, for example Schouten and Bethlehem (2009). One of its objectives was to develop and study indicators for the *representativity* of survey response. The *R*-indicator (with *R* for Representativity) was proposed by Schouten et al. (2009) and further developed in Schouten et al. (2011). One of its uses is in comparing surveys – the same survey in different countries, or different surveys within the same country – with respect to the representativity of the final set of respondents. The statistical concept behind the *R*-indicator is the variance of the response probabilities, estimated with the aid of auxiliary variables. The motivation is that a small variability of such estimates would suggest a "representative set of respondents".

Indicators based on the concept of a *balanced response set* were developed in Särndal (2011a). The response set is said to be balanced if the means for specified important auxiliary variables are the same or almost the same for the set of respondents as for all those selected in the probability sample. That respondents should be on average like all those sampled is an attractive notion. The balance indicators are computable from the auxiliary variable values available for responding as well as for nonresponding units.

The present article presents general concepts for monitoring the data collection, and they are applied to the 2009 LCS. We describe the survey in Section 3, and we analyze the LCS 2009 data in Section 4. The concepts of balance (of the response set), distance (between respondents and nonrespondents), and representativity are reviewed in Section 5, then applied to the LCS 2009 data. In Section 6 we conduct several "experiments in retrospect" with the LCS 2009 data. These experiments show that balance and distance can be improved by interventions in the data collection with the aid of paradata from Statistics

Sweden's WinDATI system explained below. Implications for the future are discussed in Section 7. The theoretical framework presented in the article is general in scope, applicable to any probability sampling design.

The access to ample auxiliary information is of crucial importance. Statistics Sweden operates in a data-rich survey environment, where high quality administrative registers allow access to many auxiliary variables, particularly for surveys on individuals and households. This also applies to the other Scandinavian countries and the Netherlands. The whole issue of nonresponse adjustment will necessarily present itself in quite a different light in countries where only highly limited auxiliary information is available, say at best a few demographic variables. However, a trend towards increased availability and use of high quality administrative data is evident in many countries.

2.   Earlier Experiences at Statistics Sweden

Several earlier studies at Statistics Sweden illustrate that a data collection motivated principally by a desire to achieve the best possible ultimate rate of response is inefficient. They suggest that scarce resources are being spent with little effect on the estimates and little improvement in representativity. Hörngren et al. (2008) and Lundquist and Särndal (2012) summarize several studies of surveys with telephone interviewing of individuals drawn by probability sampling from the Swedish Total Population Register (TPR). We mention them briefly here.

A study of the November 2002 edition of the Swedish Labour Force Survey (LFS) had found that the estimates change very little after the fifth contact attempt. It was concluded that a less elaborate fieldwork strategy, with say four call attempts instead of twelve, could considerably reduce the monthly cost for calls in the LFS.

A study along similar lines was carried out in 2007 for the Household Finances (HF) survey. Estimates were computed successively over the data collection period, as a function of the number of call attempts identified by "WD-events", which are events registered by the data collection instrument WinDATI. The simple expansion estimator (the mean for units having responded up until a given number of attempts) stabilizes at an early stage in the data collection: After about ten call attempts, the estimates change very little. Since the total number of call attempts for a sampled person may exceed 20, there is strong indication that resources are not effectively used. The calibration estimator based on selected auxiliary variables stabilizes even sooner, at around five call attempts.

In a later project, the effect of a follow-up strategy for the HF survey was studied. Low response rates had been observed in the primary data collection for several groups expected to have a high impact on the nonresponse error. However, it was found that the follow-up (the field work following the ordinary data collection) had little effect on the estimates, and that follow-up respondents are not the ones that influence the nonresponse error the most. In the end, the response rate remains disappointingly low for groups already underrepresented in the ordinary data collection.

An earlier study of the LCS had found that the representativity (measured by indicators) changes very little after an early point in the data collection, suggesting that the response set fails to become more similar to the selected sample. In addition, the follow-up appears to have little effect on the estimates. The present article examines the LCS in more depth.

3. The Swedish Living Conditions Survey

The Swedish LCS is a yearly sample survey. It has a long tradition of providing important information about social welfare in Sweden, in particular among different subgroups of Swedish society. It has become increasingly affected by nonresponse. The sample consists of individuals with an age of 16 or above drawn from the Swedish Total Population Register. The data set used in the analysis in this article is a subsample of $n = 8,220$ individuals, taken from the actual LCS 2009 sample. This subsample can be regarded as a simple random sample from the population of individuals.

Telephone interviews were conducted by a staff of interviewers using the Swedish CATI-system, WinDATI. All attempts by interviewers to establish contact with a sampled person are registered by WinDATI. Those paradata are important for this article. For every sampled individual, the WinDATI system records a series of events which we refer to as "call attempts". They play an important role in our analysis. The WinDATI events include "call without reply", "busy line", "contact with household member other than the sampled person", and "appointment booking for later contact". When contact and data delivery has occurred, the data collection effort is completed for the sampled person in question. All registered WinDATI events are taken into account in the analysis that follows.

The LCS 2009 ordinary field work lasted five weeks, at the end of which the response rate was 60.4%; for some sampled persons, 30 or more call attempts had then been recorded. This was followed by a three week break during which characteristics of non-interviewed individuals were examined in order to prepare the three week follow-up period, which concluded the data collection. All individuals considered by the survey managers to be potential respondents were included in the follow-up effort, which brought the response rate up to an ultimate 67.4%. However, there was no separate strategy or revised procedure for the follow-up. It followed the same routines as the ordinary field work. Hence, there were no attempts at responsive design where, for example, a follow-up would focus on underrepresented groups, in an objective to improve balance and reduce nonresponse bias.

In addition to these paradata, the information recorded in the LCS 2009 data set includes the response obtained on the survey target variables. In addition, it contains for all 8,220 individuals the values of a number of register variables, some of which we use as auxiliary variables. Three other register variables are used as study variables (*y*-variables), as explained in Section 4. For these we can compute unbiased estimates, based on the full sample, and compare them with estimates made under nonresponse.

We have chosen here to regard data inflow as a function of the attempt number rather than as a function of time evolved (Day 1, Day 2, and so on) since the start of data collection. Our analysis could have been conducted under the time-evolved perspective instead, with somewhat different results. In a CATI data collection, the attempt number concept is practical and natural.

4. An Analysis of the LCS 2009 Data

Results in this section reinforce the impression from earlier studies at Statistics Sweden that a data collection (including a follow-up) that proceeds according to an essentially unchanging format will produce very little change in the estimates beyond a certain "stability point" reached quite early in the data collection.

In this section, we study the dynamic behavior of survey estimates as the data collection proceeds. We measure the progression of the population total estimates for three variables as a function of the number of the call attempt, defined more precisely as the attempt at which an interviewer made successful contact with a sampled person and data delivery occurred. The three variables are register variables, used here as study variables. Their values are therefore known for all sampled units, not only for responding ones. To let three register variables play the role of study variables restricts to some extent the pool of available auxiliary variables, but it is a price worth paying in order to realize the methodological objectives.

Some notation is needed. The finite population $U = \{1, \ldots, k, \ldots, N\}$ consists of N units indexed $k = 1, 2, \ldots, N$. A probability sample s is drawn from U; in this sampling, unit k has the known inclusion probability $\pi_k = \Pr(k \in s) > 0$ and the known design weight $d_k = 1/\pi_k$. We denote the value of the study variable y as y_k. The target parameter for estimation is the population total $Y = \sum_U y_k$. (A sum $\sum_{k \in A}$ over a set of units $A \subseteq U$ will be written as $\sum_A$.) Normally, the survey involves many study variables and many totals to be estimated.

The response set is the set of units for which the value y_k has been recorded. Since we follow the data collection as a function of the call attempt number, there is a series of successively larger response sets. In a completely rigorous notation, we would denote these increasingly large response sets as $r^{(a)}$, where a refers to "call attempt number", $a = 1, 2, \ldots$, and

$$r^{(1)} \subset r^{(2)} \subset \ldots \subset r^{(a)} \subset \ldots \tag{4.1}$$

But in order to not burden the notation, it is sufficiently clear to let the notation r refer to any one of the increasingly larger response sets. Data collection stops before the expanding r has reached the full probability sample s. The value y_k recorded for $k \in r$ provide, together with auxiliary variable values, the material for estimating the parameter $Y = \sum_U y_k$.

The (design-weighted) survey response rate is

$$P = \sum_r d_k / \sum_s d_k. \tag{4.2}$$

In the context of the LCS, all d_k are equal because of simple random sampling from the Swedish TPR, so P is simply number of individuals responding divided by number in sample, but for more generality, the formulas that follow are expressed in arbitrary design weights d_k. The response probability of unit k, denoted $\theta_k = \Pr(k \in r | k \in s)$, is a conceptually defined, nonrandom, unknown number. The response rate P is an estimate of the (unknown) mean response probability in the population, $\bar{\theta}_U = \sum_U \theta_k / N$.

Auxiliary information is crucial. We denote as $\mathbf{x}_k$ the auxiliary vector value for unit k, assumed available at least for all units $k \in s$, possibly for all $k \in U$. If $J \geq 1$ auxiliary variables are used, then $\mathbf{x}_k = (x_{1k}, \ldots, x_{jk}, \ldots, x_{Jk})'$, where x_{jk} is the value for unit k of the jth auxiliary variable, x_j. We consider auxiliary vectors $\mathbf{x}_k$ of a form such that for some constant vector $\boldsymbol{\mu}$ we have $\boldsymbol{\mu}'\mathbf{x}_k = 1$ for all k. This is not a major restriction. Vectors of importance in practice are usually of this kind, such as when $\mathbf{x}_k = (1, x_k)'$ and $\boldsymbol{\mu} = (1, 0)'$.

The LCS 2009 data file analyzed here contains the observed values y_k of a number of study variables ("the y-variables") and the values x_{jk} of a number of auxiliary variables

("the *x*-variables") which may be either continuous or categorical equal to 1 or 0 to code the presence or the absence of a given trait of unit k. Some of these auxiliary values are obtained from the Swedish Total Population Register, while others are derived by matching from other reliable Swedish administrative registers, using the personal identification number.

We compute estimation weights calibrated on auxiliary information about $\mathbf{x}_k$ for $k \in s$. The weight given to the value y_k observed for $k \in r$ is $d_k m_k$, the product of the sampling weight $d_k = 1/\pi_k$ and the adjustment factor

$$m_k = \left(\sum_s d_k \mathbf{x}_k\right)' \left(\sum_r d_k \mathbf{x}_k \mathbf{x}_k'\right)^{-1} \mathbf{x}_k.$$

Hence the resulting calibration estimator is

$$\hat{Y}_{CAL} = \sum_r d_k m_k y_k. \tag{4.3}$$

The weights $d_k m_k$ are constructed to deliver unbiased estimates for the variables in the auxiliary vector, as expressed by the calibration equation $\sum_r d_k m_k \mathbf{x}_k = \sum_s d_k \mathbf{x}_k$. For any $\mathbf{x}$-vector and any response set r, the mean adjustment factor is $\sum_r d_k m_k / \sum_r d_k = 1/P$. Consequently, when the data collection progresses and r gets increasingly larger, the increasing proportion P of observed sample units in the Estimator (4.3) is correspondingly matched by a decreasing average mean adjustment factor $1/P$. But it is the composition of the response set r, the particular units that are in r at any given point, that determines the more or less pronounced bias of $\hat{Y}_{CAL}$. We want "the right kind of units" to be in r in the end.

For the theory behind calibration for nonresponse, see, for example Särndal and Lundström (2005). Calibration will generally reduce the nonresponse bias, and quite considerably if the auxiliary vector is powerful, but without eliminating it entirely. At Statistics Sweden, many potential auxiliary variables are typically available for the estimation. The question then arises about the best choice among those. Indicators for this purpose are given in Särndal and Lundström (2008, 2010).

Remark: The calibrated weights in (4.3) use an auxiliary vector $\mathbf{x}_k$ known for the sample units. In practice at Statistics Sweden the calibration estimates ordinarily draw on auxiliary information at two levels: at the population level, transmitted by a vector $\mathbf{x}_k^*$, and at the sample level, transmitted by a vector $\mathbf{x}_k^\circ$. The population total $\sum_U \mathbf{x}_k^*$ is known, while $\sum_U \mathbf{x}_k^\circ$ is unknown but estimated without bias by $\sum_s d_k \mathbf{x}_k^\circ$, which helps the reduction of nonresponse bias. The auxiliary vector is $\mathbf{x}_k = \begin{pmatrix} \mathbf{x}_k^* \\ \mathbf{x}_k^\circ \end{pmatrix}$, and to benefit from the potential for reduced variance when $\sum_U \mathbf{x}_k^*$ is known, the weights are calibrated to satisfy $\sum_r w_k \mathbf{x}_k = \mathbf{X} = \begin{pmatrix} \sum_U \mathbf{x}_k^* \\ \sum_s d_k \mathbf{x}_k^\circ \end{pmatrix}$. The published survey estimate is $\hat{Y}_{CAL}^* = \sum_r w_k y_k$ with weights $w_k = d_k \left\{ \mathbf{X}' \left(\sum_r d_k \mathbf{x}_k \mathbf{x}_k'\right)^{-1} \mathbf{x}_k \right\}$. But for the purposes of this article, it is deemed appropriate to use $\hat{Y}_{CAL} = \sum_r d_k m_k y_k$ in (4.3) with weights calibrated as $\sum_r d_k m_k \mathbf{x}_k = \sum_s d_k \mathbf{x}_k$.

In our analysis of the LCS, the adjustment factors m_k in (4.3) are computed on an auxiliary vector $\mathbf{x}_k$ of dimension eight considered suitable for monitoring the estimates over the course of the data collection and composed of the following categorical auxiliary variables: *Phone access* (equaling 1 for a person with accessible phone number; 0 otherwise), *Education level* (equaling 1 if high; 0 otherwise), *Age group* (four zero/one coded groups; age brackets -24, $25–64$, $65–74$, $75+$ years); *Property ownership* (equaling 1 for a property owner; 0 otherwise); *Country of origin* (equaling 1 if born in Sweden; 0 otherwise). We refer to this vector as the *standard* x-*vector* (to distinguish it from the *experimental* x-*vector* needed in Section 6). These variables are a subset of those used to produce the published calibration estimates in the LCS 2009.

The variable *Property ownership* equals one for a person identified in the property tax register as having paid taxes on real estate property owned. The variable *Phone* equals one for a person whose phone number is available and ready to be used at the very beginning of the data collection period. All persons in Sweden have access to a phone, whether a landline or cell phone. When the sample of persons has been drawn from the Swedish Total Population Register, it is matched to the phone register and, if found, the number is noted. "Found" or "not found" defines the dichotomous variable *Phone*. For different reasons, not all phone numbers are in the phone register. About 90% of the needed phone numbers are found. Before the start and during the field work, the interviewers try to trace the persons with phone numbers as yet missing using various sources, for example the internet. In this manner, telephone numbers are found and can be used for about one third of those with initial value zero on the *Phone* variable.

In Tables 1 and 2 (as in later tables), the entries for "Attempt number" a (where $a = 1$, $2, 3 \ldots$) are computed on the union of the sets of persons having responded at attempts 1, 2, $\ldots$, a, as expressed by (4.1). Not all call attempts are shown in the tables, but changes for deleted rows are minor. The entries for "End ordinary field work" are computed on the respondents at the end of the five week ordinary data collection period; "Final" is based on the total response recorded at the end of the follow-up period.

The three register variables used here as study variables are: *Sickness insurance benefits* (for simplicity called *Benefits*, a categorical variable equaling 1 for a recipient of such benefits; 0 otherwise), *Income* (a continuous variable including employment as well as retirement income), and *Employed* (a categorical variable equaling 1 for an employed person; 0 otherwise). We chose these three register variables because they are central aspects of living conditions as studied in the LCS. The use of these register variables as study variables meets a methodological objective: We can follow the progression of estimators of interest, and study the benefits of calibrated weighting.

For the three register variables, y_k is available for $k \in s$, and we can for comparison compute the unbiased full sample (Horvitz-Thompson) estimate

$$\hat{Y}_{FUL} = \sum_s d_k y_k. \tag{4.4}$$

The computable percentage relative difference between $\hat{Y}_{FUL}$ (unbiased) and $\hat{Y}_{CAL}$ (biased to some extent) is

$$RDF_{CAL} = 100 \cdot (\hat{Y}_{CAL} - \hat{Y}_{FUL})/\hat{Y}_{FUL}. \tag{4.5}$$

The calibration estimator generated by the primitive auxiliary vector, $\mathbf{x}_k = 1$ for all units k serves as a benchmark; it is the <u>expansion</u> estimator given by

$$\hat{Y}_{EXP} = \left(\sum_s d_k\right)\left(\sum_r d_k y_k\right) / \left(\sum_r d_k\right). \tag{4.6}$$

Its often large relative deviation from the unbiased $\hat{Y}_{FUL}$ is

$$RDF_{EXP} = 100 \cdot (\hat{Y}_{EXP} - \hat{Y}_{FUL})/\hat{Y}_{FUL}. \tag{4.7}$$

Table 1 shows RDF_{CAL} (computed with the standard $\mathbf{x}$-vector) and RDF_{EXP} for the three variables and for a number of steps in the LCS 2009 data collection. We note that:

- The numerically important changes in RDF_{CAL} and RDF_{EXP} occur early in the series of attempts because of important data inflows. From around attempt five onwards, both follow quite a stable pattern; later changes are small, moving in smooth continuous fashion. The changes are necessarily minute when the data collection has gone on for some time, because small amounts of new data are added to a substantial

Table 1. The LCS 2009 data collection: Progression of the response rate P (in per cent) and of RDF for three selected register variables. The calibration estimator is based on the standard x-vector explained in this section

Attempt number	$100 \times P$	Benefits		Income		Employed	
		RDF_{CAL}	RDF_{EXP}	RDF_{CAL}	RDF_{EXP}	RDF_{CAL}	RDF_{EXP}
1	12.8	10.5	− 10.0	− 0.05	0.3	− 1.3	− 9.0
2	24.6	3.3	− 13.9	− 1.1	0.4	− 2.0	− 8.1
3	32.8	1.6	− 12.1	− 0.4	1.6	0.2	− 4.7
4	39.6	2.7	− 10.1	0.2	2.9	0.4	− 2.4
5	44.3	3.7	− 7.2	0.7	3.6	1.1	− 1.1
6	47.8	2.7	− 7.0	1.2	4.5	1.7	0.4
7	50.9	1.6	− 7.3	2.1	5.5	2.5	1.6
8	53.0	1.0	− 7.4	2.4	6.2	2.4	2.3
9	54.6	0.2	− 8.0	2.8	6.4	2.6	2.5
10	55.7	0.2	− 8.0	2.8	6.6	2.6	2.8
11	56.8	− 0.5	− 8.5	2.7	6.5	2.6	3.0
12	57.7	0.1	− 7.9	3.0	6.8	2.5	3.1
13	58.3	− 0.3	− 8.0	3.0	6.9	2.7	3.4
14	58.7	− 0.1	− 7.7	3.0	6.9	2.7	3.6
15	59.1	− 0.5	− 8.0	3.1	7.1	2.8	3.8
⋮							
20	60.1	− 0.5	− 7.7	3.4	7.5	3.0	4.1
End ord. fieldwork	60.4	− 0.9	− 7.9	3.3	7.4	2.9	4.2
		Follow-up					
1	61.4	− 1.0	− 8.0	3.3	7.1	2.9	4.1
2	62.6	− 1.6	− 8.2	3.1	6.7	3.0	3.9
3	63.8	− 2.5	− 9.2	3.0	6.7	3.2	4.2
4	64.6	− 2.8	− 9.3	3.1	6.7	3.3	4.3
5	65.3	− 2.7	− 9.0	3.1	6.8	3.1	4.3
⋮							
10	66.8	− 2.9	− 8.9	2.9	6.7	3.0	4.5
Final	67.4	− 3.6	− 9.4	2.9	6.7	3.1	4.8

body of existing data. Every new contact attempt brings progressively smaller amounts of new information. For this we use the word "stabilization", as evidenced in Table 1 for *RDF* and in later tables for other statistics.

- For all three study variables, RDF_{CAL} is small, in fact near zero, early in the data collection. For example, for the *Benefits* variable, RDF_{CAL} hovers/ around zero in the range from 9 to 14 call attempts. The other two variables have near-zero RDF_{CAL} even earlier in the data collection. Nevertheless, at the very end (the row "Final"), the value of RDF_{CAL} is large, -3.6%, 2.9%, and 3.1% respectively. The LCS with its unchanging data collection plan does not result in small estimation error.

- The large departures from the unbiased estimate $\hat{Y}_{FUL}$, signaled by high values of RDF_{EXP}, indicate that the LCS 2009 data collection ends up with a markedly skewed response. In most of the steps in Table 1, RDF_{EXP} is greater than RDF_{CAL}. Hence the auxiliary information that $\hat{Y}_{CAL} = \sum_r d_k m_k y_k$ can draw on is valuable for reducing the departure from the unbiased estimate for all three variables, although the end result is short of satisfactory. The difference between RDF_{EXP} and RDF_{CAL} is less pronounced for *Employed*; for this variable, the auxiliary vector is less effective.

5.   Further Tools: Indicators of Balance, Distance and Representativity

In the theoretical first part of this section, we explain several indicators designed for monitoring the data collection. The indicators reflect well known statistical concepts. Later in the section we illustrate the indicators numerically by computing them on the LCS 2009 data. Furthermore, in Section 6, those indicators will be used in an experiment with the LCS 2009 data, whereby we intervene "in retrospect" in the data collection process, aiming to achieve a better balanced, or more representative, ultimate response set than if no action were taken. The indicators can be computed from the auxiliary variable values, known for both respondents and nonrespondents.

We distinguish three types of concept from which to construct an indicator: (i) Balance (of the response set, for selected auxiliary variables), (ii) Distance (between respondents and nonrespondents), and (iii) Variability of (estimated) response probabilities. All depend on the idea of imbalance now to be defined. Desirable features are high balance, low distance and small variability of the response probabilities. We use the indicators to observe the dynamic pattern as the data collection unfolds and to allow interventions to be made at suitable points.

All the indicators rely on an auxiliary vector, denoted in general by $\mathbf{x}$ with value $\mathbf{x}_k = (x_{1k}, \ldots, x_{jk}, \ldots, x_{Jk})'$ known for the units $k \in s$ (or possibly for $k \in U$). The dimension J is arbitrary. For the j:th auxiliary variable x_j, with value x_{jk} for unit k, we compute the difference $D_j = \bar{x}_{jr} - \bar{x}_{js}$ between the respondent mean, $\bar{x}_{jr} = \sum_r d_k x_{jk}/\sum_r d_k$, and the full sample mean, $\bar{x}_{js} = \sum_s d_k x_{jk}/\sum_s d_k$. If $D_j = 0$ for all J auxiliary variables, then r is a *perfectly balanced* response set. In vector form, $\mathbf{D} = (D_1, \ldots, D_j, \ldots, D_J)' = \bar{\mathbf{x}}_r - \bar{\mathbf{x}}_s$ with vector mean $\bar{\mathbf{x}}_r = \sum_r d_k \mathbf{x}_k/\sum_r d_k$ for the respondents and $\bar{\mathbf{x}}_s = \sum_s d_k \mathbf{x}_k/\sum_s d_k$ for the full sample. Under perfect balance, $\mathbf{D} = \mathbf{0}$, the zero vector.

We must seek balance on the auxiliary variables, because unlike real study variables, they are individually known for the full sample (or for the whole population). What benefit can we expect from balancing on a chosen vector $\mathbf{x}_k$? Is there reason to expect that balance on the

x-vector will produce, if not perfect, at least good balance for the y-variables in the survey, in particular for the highly important y-variable? Let us consider these questions. The concept of balance refers to the equality of response set mean and full sample mean. We have x-vector balance if $\bar{\mathbf{x}}_r = \bar{\mathbf{x}}_s$. We can strive to come close to this during data collection. The desirable goal of y-variable balance is expressed as $\bar{y}_r = \bar{y}_s$. Whether or not this comes close to being satisfied for a real study variable y will never be known. But balancing on a chosen x-vector can bring us closer. Using the property $\boldsymbol{\mu}'\mathbf{x}_k = 1$ for all k, we express the difference $\bar{y}_r - \bar{y}_s$ (which we would like to be zero or close to zero) as a sum of two terms,

$$\bar{y}_r - \bar{y}_s = (\bar{\mathbf{x}}_r - \bar{\mathbf{x}}_s)'\mathbf{b}_r + (\mathbf{b}_r - \mathbf{b}_s)'\bar{\mathbf{x}}_s \tag{5.1}$$

where $\mathbf{b}_r = \left(\sum_r d_k \mathbf{x}_k \mathbf{x}_k'\right)^{-1} \sum_r d_k \mathbf{x}_k y_k$ and $\mathbf{b}_s = \left(\sum_s d_k \mathbf{x}_k \mathbf{x}_k'\right)^{-1} \sum_s d_k \mathbf{x}_k y_k$ are linear regression coefficients (regressing y on x), for the response set and for the full sample respectively. That $\mathbf{b}_r$ and $\mathbf{b}_s$ may differ is an expression of a dilemma well known in regression analysis: Non-random selection of cases causes biased regression. The computable first term of (5.1), $(\bar{\mathbf{x}}_r - \bar{\mathbf{x}}_s)'\mathbf{b}_r$, is zero if the x-balance $\bar{\mathbf{x}}_r = \bar{\mathbf{x}}_s$ is realized. This by itself does not imply that the second (not computable) term $(\mathbf{b}_r - \mathbf{b}_s)'\bar{\mathbf{x}}_s$ is zero or small. But often that term, which is what remains of the difference $\bar{y}_r - \bar{y}_s$ after complete balance on the x-vector, is smaller than what that difference would be in the absence of any balancing. One situation where the second term is small is when y is well explained by the x-vector, so that $y_k \approx \boldsymbol{\beta}'\mathbf{x}_k$ and therefore $\mathbf{b}_r \approx \mathbf{b}_s$. In other words, if the response is balanced for a vector $\mathbf{x}_k$ highly related to the study variable y, then we are close to y-variable balance. Another condition under which the second term $(\mathbf{b}_r - \mathbf{b}_s)'\bar{\mathbf{x}}_s$ is small occurs if the data collection can be directed to yield a response set r that is an essentially random subset of s. In many situations there is a strong incentive to seek balance on a suitable x-vector, because it will likely bring us closer to y-variable balance. Multiplying by $\hat{N} = \sum_s d_k$ shows Equation (5.1) in a different light:

$$\hat{Y}_{EXP} - \hat{Y}_{FUL} = (\hat{Y}_{EXP} - \hat{Y}_{CAL}) + (\hat{Y}_{CAL} - \hat{Y}_{FUL}).$$

Here the computable difference $\hat{Y}_{EXP} - \hat{Y}_{CAL}$ is the adjustment we apply to the primitive estimate $\hat{Y}_{EXP}$ to arrive at the improved estimate $\hat{Y}_{CAL}$. The term $\hat{Y}_{CAL} - \hat{Y}_{FUL}$, unknown for a real y-variable, is not zero, but may be small compared with the adjustment $\hat{Y}_{EXP} - \hat{Y}_{CAL}$. There is no choice of x-vector that will completely eliminate the nonresponse error $\hat{Y}_{CAL} - \hat{Y}_{FUL}$. Some bias always remains after calibration. Expressed differently, there exists no x-vector that realizes missing at random, given x.

Normally in practice, $\mathbf{D} \neq \mathbf{0}$, suggesting departure from balance. We transform the multivariate $\mathbf{D}$ into a suitable univariate measure of *imbalance*, for the given survey outcome (s, r) and the given composition of $\mathbf{x}_k$. The imbalance is a quadratic form in $\mathbf{D}$ defined as

$$\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D} = (\bar{\mathbf{x}}_r - \bar{\mathbf{x}}_s)'\boldsymbol{\Sigma}_s^{-1}(\bar{\mathbf{x}}_r - \bar{\mathbf{x}}_s) \tag{5.2}$$

with weighting matrix $\boldsymbol{\Sigma}_s = \sum_s d_k \mathbf{x}_k \mathbf{x}_k' / \sum_s d_k$. Increased mean differences D_j tend to increase the imbalance $\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}$. Interposing the inverse of the weighting matrix permits an upper bound to be stated on the imbalance: For any outcome (s, r) and any composition of $\mathbf{x}_k$ we have $0 \leq \mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D} \leq Q - 1$ with $Q = 1/P$ (see Särndal 2011a). For most data

encountered in practice, $\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}$ is not a large number, often 0.3 or less. As the data collection unfolds and the response rate P gets larger, one often finds that $\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}$ decreases, because $\bar{\mathbf{x}}_r$ moves closer to $\bar{\mathbf{x}}_s$ when the response r grows toward the full sample s, although the question depends also on what particular units happen to be in the set r at a given moment.

Balance is imbalance with a reversed sign. We use two indicators of balance, measured on the unit interval scale and such that the value "1" implies perfect balance. The first is

$$BI_1 = 1 - \sqrt{\frac{\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}}{Q-1}}. \tag{5.3}$$

It follows from $0 \leq \mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D} \leq Q-1$ that $0 \leq BI_1 \leq 1$. Because $P(1-P) \leq 1/4$, an alternative indicator also contained in the unit interval is

$$BI_2 = 1 - 2P\sqrt{\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}}. \tag{5.4}$$

For most data, $\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}$ does not come near $Q-1$. It is not a sharp upper bound. Consequently, both indicators transmit an inflated notion of balance, often greater than 0.8 for both BI_1 and BI_2. The lower portion of the unit interval is not effectively used. The notion of distance now to be discussed is less subject to this criticism.

Our concept of distance, which contrasts respondents with nonrespondents for the chosen $\mathbf{x}$-vector, is a transformation of mean difference vector, $\bar{\mathbf{x}}_r - \bar{\mathbf{x}}_{nr}$, where $nr = s - r$ is the nonresponse set with mean $\bar{\mathbf{x}}_{nr} = \sum_{s-r} d_k \mathbf{x}_k / \sum_{s-r} d_k$. This distance is

$$dist_{r|nr} = [(\bar{\mathbf{x}}_r - \bar{\mathbf{x}}_{nr})'\boldsymbol{\Sigma}_s^{-1}(\bar{\mathbf{x}}_r - \bar{\boldsymbol{x}}_{nr})]^{1/2}. \tag{5.5}$$

If respondents and nonrespondents agree on average for every variable in the $\mathbf{x}$-vector, then $\bar{\mathbf{x}}_r = \bar{\mathbf{x}}_{nr}$ and $dist_{r|nr} = 0$. From (5.2) and the equation $\bar{\mathbf{x}}_r - \bar{\mathbf{x}}_s = (1-P)(\bar{\mathbf{x}}_r - \bar{\mathbf{x}}_{nr})$ follows that $dist_{r|nr}$ is a simple transformation of the imbalance $\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}$:

$$dist_{r|nr} = \frac{1}{1-P}(\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D})^{1/2} \tag{5.6}$$

From $\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D} \leq Q-1$ follows that $dist_{r|nr} \leq 1/\sqrt{P(1-P)}$. Thus for nonresponse in the range 20% to 50%, $dist_{r|nr}$ can never exceed 2.5. But for data encountered in practice $dist_{r|nr}$ is normally much lower, rarely exceeding 0.6. One reason is that the upper bound covers any vector composition $\mathbf{x}_k$ and even the most extreme response outcome r that can occur for the given sample s. The measure $dist_{r|nr}$ reacts distinctly but smoothly to the steps in the data collection and is a more expressive indicator than BI_1 or BI_2, which tend to concentrate in the upper quarter of the unit interval. For example, in Table 3 (Subsection 6.2), $dist_{r|nr}$ roughly doubles from 0.23 at the beginning to 0.47 at the end of the data collection, while BI_1 only moves from 0.85 to 0.72.

Simple relationships between $dist_{r|nr}$ and the balance indicators follow from (5.3), (5.4) and (5.6):

$$BI_1 = 1 - \sqrt{P(1-P)} \times dist_{r|nr} \quad ; \quad BI_2 = 1 - 2P(1-P) \times dist_{r|nr}. \tag{5.7}$$

The principal tools for the empirical work reported later are the balance indicator BI_1 and the distance $dist_{r|nr}$. It is important to follow their progression as the response set r expands

and P increases. Increasing balance and decreasing distance are signs of a satisfactory data collection. But the undesirable opposite can happen, as in an empirical illustration that follows. The level of the indicators depend on the choice of $\mathbf{x}_k$, notably on the number of x-variables in $\mathbf{x}_k$ – it is harder to obtain balance on more variables – and on their relationship with response. The ultimate response set r should have high balance and low distance; however, it is hard to formulate definite ultimate target values for BI_1 and $dist_{r|nr}$, because of their strong dependence on $\mathbf{x}_k$.

If $dist_{r|nr}$ decreases when P increases, then the balance, as measured by BI_1 or by BI_2, may or may not improve, that is, get larger. If the distance $dist_{r|nr}$ increases when P increases towards 50%, then the balance, measured by BI_1 or by BI_2, will necessarily deteriorate.

The third concept behind an indicator for the data collection is the variability of (estimated) response probabilities. It was used in the RISQ project mentioned earlier. The resulting indicators are called R-indicators (with R for representativity); see, for example Schouten et al. (2009). Let $\hat{\theta}_k$ be the estimated response probability for unit $k \in s$. Their variance is

$$S_{\hat{\theta}s}^2 = \sum_s d_k(\hat{\theta}_k - \bar{\hat{\theta}}_s)^2 / \sum_s d_k \tag{5.8}$$

where $\bar{\hat{\theta}}_s = \sum_s d_k\hat{\theta}_k / \sum_s d_k$. The R-indicator is defined as

$$R = 1 - 2S_{\hat{\theta}s} \tag{5.9}$$

The rationale behind the construction (5.9) is that if the data collection can be directed to reduce variability in the estimated response probabilities, then the representativity of the response set, measured by R, is said to be improved.

For the chosen specification $\mathbf{x}_k$, the estimates $\hat{\theta}_k = f(\mathbf{x}_k'\hat{\boldsymbol{\beta}})$ can be obtained via different link functions. For the linear response function, using weighted least squares, we determine $\boldsymbol{\beta}$ to minimize $\sum_s d_k(I_k - \mathbf{x}_k'\boldsymbol{\beta})^2$, where I_k is the response indicator, $I_k = 1$ for $k \in r$ and $I_k = 0$ for $k \in s - r$. As a result, $\hat{\boldsymbol{\beta}} = \left(\sum_s d_k\mathbf{x}_k\mathbf{x}_k'\right)^{-1}\left(\sum_r d_k\mathbf{x}_k\right)$, and for $k \in s$, the response probability estimate is $\hat{\theta}_k = t_k$ with $t_k = \mathbf{x}_k'\hat{\boldsymbol{\beta}}$. If we denote by S_{ts}^2 the variance (5.8) computed with $\hat{\theta}_k = t_k$, the R-indicator (5.9) is $R = 1 - 2S_{ts}$. Because $S_{ts}^2 = P^2 \times \mathbf{D}'\Sigma_s^{-1}\mathbf{D}$, the R-indicator for the linear response function is equal to the balance measure (5.4): $1 - 2S_{ts} = BI_2$.

Schouten et al. (2009) closely examine the case where $\hat{\theta}_k$ is obtained through a logistic response function. By logistic regression fit, we obtain first $\hat{\boldsymbol{\beta}}$, then $\hat{\theta}_{k,\log} = \exp(\mathbf{x}_k'\hat{\boldsymbol{\beta}})/[1 + \exp(\mathbf{x}_k'\hat{\boldsymbol{\beta}})]$ for $k \in s$. Their variance, denoted $S_{\hat{\theta}\log,s}^2$, is computed in the manner of (5.8) with $\hat{\theta}_k = \hat{\theta}_{k,\log}$, and the resulting logistic R-indicator is

$$R = 1 - 2S_{\hat{\theta}\log,s}. \tag{5.10}$$

Bethlehem et al. (2011) consider a bias-adjusted R-indicator, reflecting a desire to reduce the bias that (5.10) may have when viewed as an estimate of a corresponding population quantity. With our data, (5.10) differed negligibly from its bias-corrected counterpart, not shown in Table 2.

A noteworthy property of the imbalance $\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}$ is its simple relation to the coefficient of variation of the estimates $\hat{\theta}_k = t_k$ for $k \in s$. Because $\bar{t}_s = \sum_s d_k t_k / \sum_s d_k = P$ we have

$$cv_{ts} = S_{ts}/\bar{t}_s = (\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D})^{1/2}. \tag{5.11}$$

Often close in value to cv_{ts} is the coefficient of variation of the adjustment factors m_k in the estimator $\hat{Y} = \sum_r d_k m_k y_k$. It can be shown that

$$cv_{mr} = S_{mr}/\bar{m}_r = (\mathbf{D}'\boldsymbol{\Sigma}_r^{-1}\mathbf{D})^{1/2}$$

where $\boldsymbol{\Sigma}_r = \sum_r d_k \mathbf{x}_k \mathbf{x}_k' / \sum_r d_k$; $S_{mr} = \left[\sum_r d_k (m_k - \bar{m}_r)^2 / \sum_r d_k\right]^{1/2}$ and $\bar{m}_r = \sum_r d_k m_k / \sum_r d_k = 1/P$. The two coefficients of variation differ only in the inverted weighting matrix: $\boldsymbol{\Sigma}_s^{-1}$ in the former, $\boldsymbol{\Sigma}_r^{-1}$ in the latter. The statistic cv_{mr} is used in

Table 2.   *The LCS 2009 data collection: Progression of the response rate P (in per cent), balance BI_1 and BI_2, logistic R−indicator (5.10), distance $dist_{r|nr}$, and square root of imbalance $(\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D})^{1/2} = cv_{ts}$. Computations based on the standard x-vector explained in Section 4*

Attempt number	$100 \times P$	Balance BI$_1$	Balance BI$_2$	R-indicator formula (5.10)	Distance $dist_{r|nr}$	Sqrt. imbalance cv_{ts}
1	12.8	0.855	0.904	0.902	0.433	0.378
2	24.6	0.802	0.829	0.829	0.460	0.347
3	32.8	0.779	0.793	0.794	0.470	0.316
4	39.6	0.770	0.775	0.780	0.471	0.285
5	44.3	0.767	0.769	0.775	0.469	0.261
6	47.8	0.763	0.763	0.770	0.475	0.248
7	50.9	0.756	0.756	0.763	0.488	0.240
8	53.0	0.751	0.752	0.758	0.499	0.234
9	54.6	0.750	0.752	0.757	0.501	0.227
10	55.7	0.748	0.749	0.756	0.508	0.225
11	56.8	0.746	0.749	0.754	0.512	0.221
12	57.7	0.747	0.750	0.756	0.513	0.217
13	58.3	0.744	0.748	0.754	0.519	0.217
14	58.7	0.742	0.746	0.753	0.523	0.216
15	59.1	0.741	0.745	0.752	0.527	0.215
⋮						
20	60.1	0.737	0.743	0.751	0.536	0.214
End ordinary	60.4	0.738	0.744	0.752	0.536	0.212
Follow-up						
1	61.4	0.736	0.743	0.751	0.542	0.210
2	62.6	0.734	0.742	0.750	0.550	0.206
3	63.8	0.730	0.741	0.748	0.561	0.203
4	64.6	0.728	0.740	0.747	0.569	0.201
5	65.3	0.727	0.740	0.748	0.573	0.199
⋮						
10	66.8	0.719	0.736	0.742	0.596	0.198
Final	67.4	0.717	0.735	0.742	0.603	0.197

selecting auxiliary variables at the estimation stage, as in Särndal and Lundström (2010) and Särndal (2011b).

Table 2 shows the progression during the LCS 2009 data collection of BI_1, BI_2 and the logistic R-indicator (5.10), all viewed as functions of the call attempt number. The three measures are numerically close, but all are subject to the criticism that they fail – here and in other applications – to effectively use the whole unit interval. Often around 0.8 or higher, they seldom fall below 0.7. Here, as for most other data, the imbalance $\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}$ does not come near its upper bound $Q - 1$, so (5.3) and the related (5.4) fall predominantly into the upper end of the unit interval. Because BI_1, BI_2 and the logistic R-indicator (5.10) tell essentially the same story, we focus in the following on BI_1. Table 2 also shows the progression of the distance $dist_{r|nr}$ between respondents and nonrespondents.

The desired pattern of reduced distance and increased balance does not happen for the LCS 2009 data collection. Instead Table 2 shows that the balance indicators and the distance $dist_{r|nr}$ go the wrong way: The balance decreases; the distance $dist_{r|nr}$ gets larger. Thus Table 2 reinforces the message already conveyed in Table 1 of a weakness in the LCS data collection. It raises the question of whether the ordinary field work should proceed as long as it currently does, instead of ending after say ten attempts. The follow-up does not bring improvement; the indicators continue in the wrong direction.

Also shown in Table 2 is $(\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D})^{1/2} = cv_{ts}$ defined in (5.11). Here the imbalance $\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}$ goes from an initial value 0.14 to a final value 0.04, a decrease explained in large part by the increasing proportion P, which makes $\bar{\mathbf{x}}_r$ move closer to $\bar{\mathbf{x}}_s$ and $\mathbf{D}$ closer to the zero vector.

6. Experimental Data Collection Strategies

6.1. Auxiliary Vector for the Experiments

There is strong evidence that realizing a predefined "respectable" overall response rate should no longer be accorded the same dominant importance in future renditions of the LCS. It is hard to justify a costly effort for a possible five per cent greater ultimate response rate unless accompanied by concrete measures of quality in the response set, such as progressively better balance and closeness of respondents to nonrespondents. As Tables 1 and 2 have shown, those features are lacking in the LCS 2009 data collection.

This section presents the results of three "experiments in retrospect" carried out with the existing LCS 2009 data file. We cannot add more data, but we can delete data from that file to show the effects of different interventions in the data collection, in particular the trend in the balance indicators BI_1 and BI_2, and in the distance $dist_{r|nr}$. Increasing balance and decreasing distance are features we hope to find. Our experiments consist in treating data collection as terminated, at suitably chosen points in the data inflow, for sample groups with relatively high response. For example, it might stop in some groups after a suitable number of call attempts because realistic expectations for the response have already been met, whereas for the other groups, data collection would continue for some time yet before stopping, and for remaining low-responding groups it would continue until the very end of the data collection period. We refer to the points where stopping occurs as *intervention points*.

In this manner we delete data in the existing LCS 2009 data file pretending that data collection has been terminated at given points for relatively high-responding sample subgroups. In other words, for those groups, we sacrifice some data y_k that were in reality available beyond the intervention points. The imbalance measured by $\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}$ plays a key role in the analysis. It determines the balance measures BI_1 and BI_2 given respectively by (5.3) and (5.4) and the distance $dist_{r|nr}$ given by (5.6). An aim for the data collection should be to reduce the differences D_j that make up the vector $\mathbf{D} = \bar{\mathbf{x}}_r - \bar{\mathbf{x}}_s = (D_1, \ldots, D_j, \ldots, D_J)'$.

The imbalance $\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}$ has a particularly transparent expression when the vector $\mathbf{x}_k$ is defined by J mutually exclusive and exhaustive traits or characteristics, for example when "Age" is defined by, say, $J = 3$ traits, Young, Middle-aged and Elderly. But usually in practice, several categorical variables are crossed to define a set of mutually exclusive and exhaustive groups. The trait of unit k is then uniquely coded by the J-vector $\mathbf{x}_k = (\gamma_{1k}, \ldots, \gamma_{jk}, \ldots, \gamma_{Jk})' = (0, \ldots, 1, \ldots, 0)'$ (with a single entry "1"), or equivalently by the J-vector $\mathbf{x}_k = (1, \gamma_{1k}, \ldots, \gamma_{J-1,k})'$, where $\gamma_{jk} = 1$ if k has the trait j and $\gamma_{jk} = 0$ otherwise. Denote by s_j the (non-empty) subset of the sample s consisting of the units k with the trait j, and let r_j be the corresponding responding subset of the whole response set r; $r_j \subseteq s_j$. For trait j, denote by $W_{js} = \sum_{s_j} d_k / \sum_s d_k$ that trait's share of the whole sample s. Then the imbalance is a sum of non-negative terms expressed as

$$\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D} = \sum_{j=1}^{J} C_j \qquad (6.1)$$

with

$$C_j = W_{js} \times \left(\frac{P_j}{P} - 1\right)^2 \qquad (6.2)$$

where $P_j = \sum_{r_j} d_k / \sum_{s_j} d_k$ is the response rate for the jth group and P is the overall response rate given by (4.2). We call $(P_j - P)/P$ the *response rate differential* for the jth group. Together, the J differentials $(P_j - P)/P$ describe the state of the response for the set of groups at any given point in the data collection. The differentials are positive, negative or zero. If all are zero, the imbalance is zero, and the balance is perfect for the chosen $\mathbf{x}$-vector: $BI_1 = BI_2 = 1$. The differentials change continuously during data collection and can be substantially different, although experience shows that they are seldom greater in absolute value than 0.3. At any given point in the data collection, their weighted average is zero: $\sum_{j=1}^{J} W_{js} \times ((P_j/P) - 1) = 0$. The imbalance (6.1) is therefore the variance over the J groups of the differentials $(P_j - P)/P$. If the maximum $|P_j - P|/P$ over the J groups equals, say, 0.5, it follows that $\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D} \leq 0.25$ and that $(\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D})^{1/2} = cv_{ts} \leq 0.5$. If all J response rates P_j are equal then $\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D} = 0$.

The analysis in Sections 4 and 5 was based on the *standard* $\mathbf{x}$-*vector*, close to the auxiliary vector used to produce the calibration estimates for LCS 2009. Here we choose a more appropriate vector that identifies a set of particularly important sample subgroups. Using this *experimental* $\mathbf{x}$-*vector* we carry out three "experiments in retrospect" on the LCS 2009 data, each based on an *experimental data collection strategy* defined by one or more intervention points and a stopping rule for each intervention point. This vector points

out membership in one of $J = 8$ mutually exclusive and exhaustive sample groups. Every intervention point marks a change in the data collection. The stopping rule is formulated in terms of a predefined target response rate for each group, so that data collection will be deemed terminated at a given intervention point for groups having at that point reached the specified response rate. This is a simple form of responsive design, made possible by the categorical nature of the experimental vector.

The experimental **x**-vector is defined by the crossing of three dichotomous auxiliary variables: *Education level* (high, not high), *Property ownership* (owner, non-owner), *Country of origin* (Sweden, other). There are $J = 2^3 = 8$ mutually exclusive and exhaustive groups coded by the experimental x-vector $\mathbf{x}_k = (\gamma_{1k}, \ldots, \gamma_{8k})'$, where $\gamma_{jk} = 1$ if k belongs to group j and $\gamma_{jk} = 0$ otherwise. Group membership, and hence the value $\mathbf{x}_k$, is known auxiliary information for all $k \in s$. Those eight groups are important to monitor because experience has shown their response rates to be considerably different and indeed strikingly low for some, as Table 5 confirms.

6.2. The Actual LCS 2009 Data Collection Analyzed with the Experimental x-vector

By the *actual LCS data collection* we mean the data collection as actually carried out, with all the contact attempts and realized responses, resulting in the actual LCS 2009 response set. We compare it in Subsection 6.3 and 6.4 with three experimental data collections where the LCS 2009 data set is censored by stopping rules for data collection in certain sample subgroups.

To put the experiments in their proper light, we analyze first the actual LCS 2009 data collection in the light of the experimental **x**-vector defined in Subsection 6.1. Summary results are shown in Tables 3 and 4.

As Table 1 showed for the standard **x**-vector, Table 3 shows that RDF_{CAL} (with weights now calibrated on the experimental **x**-vector) does not terminate at desirable near-zero levels. The balance, measured by BI_1 and BI_2, decreases as the data collection proceeds, and the distance $dist_{r|nr}$ increases. This again indicates an inefficiency in the 2009 data collection, with its predefined unchangeable format.

Table 4 shows the progression over the LCS 2009 data collection of the eight terms C_j defined by (6.2), whose total $\mathbf{D}'\mathbf{\Sigma}_s^{-1}\mathbf{D}$ is given in the bottom line. Both are multiplied by 100. A low variability in the C_j is a goal, because if all C_j are equal in the end, the imbalance is zero. The groups in lines 1 and 8 stand out, but for different reasons. In both cases, C_j remains high from attempt 5 (where the data collection has gained a certain stability) until the very end. High values of C_j also prevail throughout for lines 4, 5 and 6. The very low-responding line 1 group, *education not high, non-owner, foreign origin*, has a large negative response differential $(P_j - P)/P$, and although $100 \times C_j$ decreases somewhat from 1.44 at attempt 5 to a final value of 1.18, the decrease is much weaker than desired. A negative response differential and a large C_j also characterizes the line 5 group. By contrast, distinctly positive response differentials $(P_j - P)/P$ characterize lines 4, 6 and 8. Most prominent of these is the line 8 group, *high education, property owner, Swedish origin*, for which $100 \times C_j$ decreases somewhat, from 0.58 at attempt 5 to a final value of 0.44, but less than one would like to see.

Table 3.   The actual LCS 2009 data collection: Progression of RFD_{CAL} (for three study variables), BI_1 and $dist_{r|nr}$. Computations based on the experimental x-vector of dimension eight defined in Subsection 6.1

Attempt number	$100 \times P$	RFD_{CAL} Benefits	Income	Employed	BI_1	$dist_{r|nr}$
1	12.8	− 8.4	− 2.7	− 10.2	0.922	0.233
2	24.6	− 13.2	− 3.2	− 9.7	0.887	0.263
3	32.8	− 11.5	− 2.3	− 6.3	0.867	0.283
4	39.6	− 8.5	− 1.5	− 4.4	0.850	0.306
5	44.3	− 5.8	− 0.5	− 3.0	0.846	0.310
⋮						
8	53.0	− 5.7	1.7	0.2	0.812	0.377
⋮						
12	57.7	− 6.1	2.5	1.2	0.805	0.394
⋮						
20	60.1	− 6.0	3.1	2.2	0.795	0.418
⋮						
End ordinary field work	60.4	− 6.2	3.1	2.3	0.796	0.417
		Follow-up				
1	61.4	− 6.2	3.0	2.3	0.796	0.418
⋮						
4	64.6	− 7.9	2.8	2.6	0.792	0.435
⋮						
Final	67.4	− 7.9	2.9	3.1	0.779	0.471

In Subsections 6.3 and 6.4 we contrast these results on the group factors C_j with results from three experimental strategies obtained through interventions in the LCS 2009 data base. All three use the experimental **x**-vector of dimension $J = 2^3 = 8$ as defined in Subsection 6.1, but they differ in the points of intervention and in the stopping rules.

Table 4.   The actual LCS 2009 data collection: values of the eight terms C_j of $\mathbf{D}' \mathbf{\Sigma}_s^{-1} \mathbf{D}$ (both multiplied by 100). Computations based on the experimental x-vector defined in Subsection 6.1

| Group characteristic | | | $100 \times C_j$ | | | | | | |
| | | | Ord. field work attempt | | | | Follow-up attempt | | |
Education	Property ownership	Origin	1	5	12	End	1	4	Final
Not high	Non-owner	Abroad	1.49	1.44	1.26	1.23	1.25	1.16	1.18
Not high	Non-owner	Sweden	0.00	0.06	0.11	0.11	0.08	0.07	0.07
Not high	Owner	Abroad	0.06	0.01	0.00	0.00	0.00	0.00	0.00
Not high	Owner	Sweden	0.72	0.24	0.21	0.19	0.17	0.17	0.18
High	Non-owner	Abroad	1.28	0.39	0.29	0.26	0.25	0.23	0.22
High	Non-owner	Sweden	0.11	0.26	0.25	0.24	0.21	0.20	0.23
High	Owner	Abroad	0.18	0.01	0.03	0.03	0.03	0.02	0.04
High	Owner	Sweden	0.29	0.58	0.64	0.66	0.62	0.53	0.44
		$100 \times \mathbf{D}' \mathbf{\Sigma}_s^{-1} \mathbf{D}$	4.13	2.99	2.78	2.72	2.61	2.37	2.36

Table 5. *Response rate P (in per cent) at three points in the actual LCS 2009 data collection for the eight groups formed by the experimental x-vector*

Group characteristic			Response rate P (per cent)			
Education	Property ownership	Origin	Attempt 12 ordinary	Attempt 2 follow-up	Final	Individuals in sample
Not high	Non-owner	Abroad	37.5	41.8	44.6	847
Not high	Non-owner	Sweden	54.6	59.8	64.6	3210
Not high	Owner	Abroad	58.5	62.3	66.8	171
Not high	Owner	Sweden	63.0	67.6	73.2	2036
High	Non-owner	Abroad	39.4	44.9	48.7	236
High	Non-owner	Sweden	66.8	71.6	77.6	816
High	Owner	Abroad	68.1	73.6	81.9	72
High	Owner	Sweden	72.2	77.4	81.5	832
		Total	57.7	62.6	67.4	8220

6.3. *Experimental Strategy 1 and Its Results*

We define Experimental Strategy 1 to have two intervention points, Attempt 12 of the ordinary data collection (point 1), and Attempt 2 of the follow-up (point 2); the stopping rule is to declare data collection terminated (so that no further *y*-values are taken into account) in a group that has realized at least 65% response. Table 5 shows the response rates for the actual LCS 2009 data at the two intermediate points and at the very end

Table 6. *Experimental strategy 1: the eight terms C_j of $\mathbf{D}'\mathbf{\Sigma}_s^{-1}\mathbf{D}$ (multiplied by 100), the response rate P (in per cent), the balance BI_1, and the distance $dist_{r|nr}$, computed on the experimental x-vector at three points in the data collection*

Group characteristic			Value of $100 \times C_j$ at		
Education	Property ownership	Origin	Attempt 12 ordinary	Attempt 2 follow-up	Final
Not high	Non-owner	Abroad	1.26	1.06	0.94
Not high	Non-owner	Sweden	0.11	0.03	0.00
Not high	Owner	Abroad	0.00	0.00	0.00
Not high	Owner	Sweden	0.21	0.24	0.08
High	Non-owner	Abroad	0.29	0.21	0.16
High	Non-owner	Sweden	0.25	0.07	0.02
High	Owner	Abroad	0.03	0.01	0.00
High	Owner	Sweden	0.64	0.31	0.17
	$100 \times \mathbf{D}'\mathbf{\Sigma}_s^{-1}\mathbf{D}$		2.78	1.93	1.39
	$100 \times P$		57.7	61.5	63.9
	BI_1		0.805	0.824	0.843
	$dist_{r|nr}$		0.394	0.361	0.326

(Final). It follows that the Strategy 1 data collection is deemed terminated at point 1 for the groups in lines 6, 7 and 8, and at point 2 for the group in line 4, while remaining groups continue until the very end. For the low-responding line 1 and 5 groups, the final response rate is still far from 65%.

For the data collection of Strategy 1, Table 6 shows the progression of the terms C_j and their total $\mathbf{D}'\mathbf{\Sigma}_s^{-1}\mathbf{D}$ (both multiplied by 100). Data collection has ended at point 1 for the relatively high-responding groups in lines 6, 7 and 8. The ensuing marked decrease in C_j for lines 6 and 8 occurs because the response differential $(P_j - P)/P$ drops when the increasing P draws nearer the unchanging P_j. The low-responding group 1 accounts for the largest $100 \times C_j$. It drops from 1.26 at point 1 but still ends at a fairly high 0.94. Both P_1 and P increase; they are getting closer, and $|P_1 - P|/P$ is reduced, but not enough. Although only two interventions are used in Strategy 1, the imbalance $100 \times \mathbf{D}'\mathbf{\Sigma}_s^{-1}\mathbf{D}$ is greatly reduced, from 2.78 at first intervention to 1.39 at the end. As Table 6 also shows, both balance and distance now go in the desired directions: The balance BI_1 increases from 0.805 to 0.843 and the distance $dist_{r|nr}$ decreases from 0.394 to 0.326. The ultimate response rate for Strategy 1 is 63.9%, as compared with 67.4% in the actual LCS 2009 data collection.

6.4. Experimental Strategies 2 and 3

Experimental strategies 2 and 3 use sharpened stopping rules for the data collection in the eight groups defined by the experimental **x**-vector defined in Subsection 6.1. The objective is to attempt to confirm the expectation that still better balance can be achieved.

Strategy 2 is defined to declare data collection terminated (in the ordinary data collection or in the follow-up) for a group as soon as its response has reached 60%. The resulting five intervention points are shown in Table 7: Five groups terminate at four different points in the ordinary data collection, and one group terminates at follow-up attempt 3. The low-responding line 1 and line 5 groups continue to the end, but still do not come near 60% response.

Table 8 shows the terms $100 \times C_j$, which sum to the total imbalance $100 \times \mathbf{D}'\mathbf{\Sigma}_s^{-1}\mathbf{D}$. Compared with Strategy 1 in Table 6, we see that Strategy 2 brings improvement in that for all but the line 1 group, C_j is reduced to near-zero levels at the end (the column Final). The column total $100 \times \mathbf{D}'\mathbf{\Sigma}_s^{-1}\mathbf{D}$ is reduced markedly from 3.07 at first intervention to a final value of 0.82, considerably lower than the final value 1.39 for Strategy 1. As a result, the balance improves markedly, and the distance $dist_{r|nr}$ is reduced in the end to 0.220, as compared with 0.326 for Strategy 1.

The stopping rule for experimental strategy 3 is to declare data collection terminated for a group whose response rate has reached 50%. For this more stringent rule, data collection terminates with still fewer attempts than in Strategy 2. The improvement in the indicators becomes further pronounced, giving better balance and decreased distance compared with Strategies 1 and 2. The distance $dist_{r|nr}$ now ends at 0.089, as compared with final values of 0.220 in Strategy 2 and 0.326 in Strategy 1. Strategy 3 in the end leaves very little variation in the response differentials, which explains the low value 0.20 of $100 \times \mathbf{D}'\mathbf{\Sigma}_s^{-1}\mathbf{D}$.

Table 7. *Experimental strategies 2 and 3: termination points for data collection and response rate (in per cent) at termination for the eight groups*

Group			Strategy 2		Strategy 3	
Education	Property ownership	Origin	Termination point	Response rate at termination	Termination point	Response rate at termination
Not high	Non-owner	Abroad	Final	44.6	Final	44.6
Not high	Non-owner	Sweden	Att. 3 follow-up	61.0	Att. 8 ordinary	50.0
Not high	Owner	Abroad	Att. 15 ordinary	60.2	Att. 7 ordinary	51.5
Not high	Owner	Sweden	Att. 9 ordinary	60.0	Att. 6 ordinary	52.6
High	Non-owner	Abroad	Final	48.7	Final	48.7
High	Non-owner	Sweden	Att. 7 ordinary	60.2	Att. 5 ordinary	50.5
High	Owner	Abroad	Att. 8 ordinary	62.5	Att. 6 ordinary	52.8
High	Owner	Sweden	Att. 7 ordinary	63.5	Att. 4 ordinary	50.1
Entire data collection				58.9		50.3

Table 8.   Experimental strategy 2: the eight terms C_j of $\mathbf{D'}\mathbf{\Sigma}_s^{-1}\mathbf{D}$ (multiplied by 100), the balance BI_1, and the distance $dist_{r|n}$, computed on the experimental x-vector at six points in the data collection. Column "7 ord." refers to "Attempt 7 in the ordinary data collection"; analogous for other columns

| Group | | | Value of $100 \times C_j$ at data collection point | | | | | |
Education	Property ownership	Origin	7 ord.	8 ord.	9 ord.	15 ord.	3 fol.-up	Final
Not high	Non-owner	Abroad	1.39	1.40	1.29	0.99	0.78	0.60
Not high	Non-owner	Sweden	0.12	0.09	0.05	0.00	0.07	0.05
Not high	Owner	Abroad	0.00	0.00	0.00	0.01	0.00	0.00
Not high	Owner	Sweden	0.25	0.33	0.33	0.13	0.02	0.01
High	Non-owner	Abroad	0.35	0.31	0.31	0.22	0.15	0.09
High	Non-owner	Sweden	0.33	0.21	0.14	0.05	0.01	0.00
High	Owner	Abroad	0.02	0.03	0.02	0.01	0.00	0.00
High	Owner	Sweden	0.61	0.44	0.33	0.18	0.07	0.06
		$100 \times \mathbf{D'}\mathbf{\Sigma}_s^{-1}\mathbf{D}$	3.07	2.81	2.49	1.59	1.09	0.82
		$100 \times P$	50.9	52.5	53.8	56.0	58.6	58.9
		BI_1	0.822	0.824	0.830	0.858	0.876	0.892
		$dist_{r|nr}$	0.357	0.353	0.341	0.287	0.252	0.220

Table 9.   Experimental strategy 3: the eight terms C_j of $\mathbf{D'}\mathbf{\Sigma}_s^{-1}\mathbf{D}$ (multiplied by 100), the balance BI_1, and the distance $dist_{r|nr}$, computed on the experimental x-vector at six points in the data collection. Column headed "4 ord." refers to "Attempt 4 in the ordinary data collection"; analogous for other columns

| Group | | | Value of $100 \times C_j$ at data collection point | | | | | |
Education	Property ownership	Origin	4 ord.	5 ord.	6 ord.	7 ord.	8 ord.	Final
Not high	Non-owner	Abroad	1.51	1.39	1.23	1.10	1.05	0.13
Not high	Non-owner	Sweden	0.05	0.03	0.02	0.00	0.03	0.00
Not high	Owner	Abroad	0.01	0.00	0.00	0.01	0.01	0.00
Not high	Owner	Sweden	0.26	0.30	0.46	0.25	0.16	0.05
High	Non-owner	Abroad	0.59	0.38	0.35	0.27	0.22	0.00
High	Non-owner	Sweden	0.27	0.30	0.12	0.06	0.03	0.01
High	Owner	Abroad	0.00	0.01	0.02	0.01	0.01	0.00
High	Owner	Sweden	0.72	0.21	0.07	0.02	0.01	0.00
		$100 \times \mathbf{D'}\mathbf{\Sigma}_s^{-1}\mathbf{D}$	3.42	2.62	2.26	1.73	1.51	0.20
		$100 \times P$	39.6	43.8	46.4	47.8	48.7	50.3
		BI_1	0.850	0.857	0.860	0.874	0.880	0.955
		$dist_{r|nr}$	0.306	0.288	0.281	0.252	0.240	0.089

6.5.   A Comparison of the Data Collection Strategies

All three experimental strategies use interventions in the LCS 2009 data, with successively more stringent stopping rules, as described in Subsections 6.3 and 6.4. Table 10 summarizes the experiments and compares them with the actual LCS 2009 data collection (with no interventions). For comparability, at the end of the respective data collections the entries in Table 10 are computed on the standard auxiliary vector defined in Section 4, which resembles the one used to produce the LCS estimates in 2009. The entries for the actual LCS 2009 data collection (the first line) are taken from the bottom line, "Final", in Tables 1 and 2.

Table 10.   The three experimental strategies compared with the actual LCS 2009 data collection; response rate P (in per cent), RDF_{CAL}, BI_1, $dist_{r|nr}$ and reduction (in per cent) of the number of call attempts. Computations based on the standard x-vector explained in Section 4

End data collection	$100 \times P$	RDF_{CAL} Benefits	Income	Employment	BI_1	$dist_{r|nr}$	Reduction in %
Actual LCS 2009	67.4	− 3.6	2.9	3.1	0.717	0.603	0.0
Strategy 1	63.9	− 1.6	2.7	3.0	0.765	0.489	8.2
Strategy 2	58.9	− 1.2	2.6	3.2	0.787	0.433	20.2
Strategy 3	50.3	1.0	1.0	2.3	0.808	0.383	36.4

Table 10 shows that each experimental strategy improves on the preceding one. The relative deviation RDF_{CAL} is reduced in each step for all three register variables used as study variables (if we disregard a slightly higher value for the variable *Employment* in Strategy 2). For *Income* and *Employment*, the major reduction in RDF_{CAL} occurs in the step from Strategy 2 to Strategy 3.

Both the balance and the response-to-nonresponse distance improve in each step. The distance $dist_{r|nr}$ drops from 0.603 to 0.383. The balance BI_1 increases from 0.717 to 0.808, the greatest step occurring from the actual LCS 2009 to Experimental Strategy 1. The balance shown in Table 10 is lower than the balance for the corresponding experimental strategy in Tables 6, 8 and 9. This is because the **x**-vectors are different; it is harder to achieve high balance for a more extensive vector.

A striking benefit from the experimental strategies is an implicit reduction of data collection cost through significantly fewer call attempts. To reach the 67.4% response in the complete 2009 LCS data collection, 53,258 attempts were used, but to reach the 63.9% response in experimental Strategy 1, only 48,883 attempts are used, a reduction of 8.2%. The reduction in call attempts is even more striking for the other two experimental strategies: 20.2% for Strategy 2 and 36.4% for Strategy 3. In practice, such cost savings should be used to improve other aspects of the survey design; one could for example afford a larger size sample *s* to begin with.

7.   Discussion and Implications for the Future

The concepts proposed in this article, more specifically those presented in Sections 4 and 5, are general in scope and can be applied to a variety of sample surveys. We have chosen the 2009 Swedish Living Conditions Survey as an instrument to illustrate the use of these concepts, which are also being tested and evaluated in other surveys at Statistics Sweden.

In Section 5 we introduced the concept of imbalance, defined mathematically by the quadratic form $\mathbf{D}'\mathbf{\Sigma}_s^{-1}\mathbf{D}$, Formula (5.2). The imbalance, a function of the chosen auxiliary vector $\mathbf{x}_k$, determines important tools presented in Section 5: The *balance of the set of respondents* and the *distance between respondents and nonrespondents*. Signs of a good data collection are increasing balance and decreasing distance during the course of the data collection.

In this article we have used these tools and paradata from the Swedish CATI-system to examine the data collection in the 2009 Swedish Living Conditions Survey. Earlier studies at Statistics Sweden had cast doubt on the merits of conducting a follow-up in the LCS; our

results in Sections 4 and 5 confirm those earlier findings. In Table 1 we studied the changes in the estimates for three register variables as the data collection progresses. The follow-up does not produce the improvement one would hope for.

Table 2 shows that the balance indicators BI_1 and BI_2 have a decreasing trend over the course of the LCS 2009 data collection. Contrary to reasonable expectations, the set of respondents is thus less balanced after the follow-up than at the end of the ordinary fieldwork. Furthermore, in Table 2 the distance $dist_{r|nr}$ shows an increasing rather than a decreasing trend as the data collection unfolds. This adds to earlier doubts about the efficiency of the current LCS data collection.

When the auxiliary vector $\mathbf{x}_k$ codifies membership in one of J mutually exclusive and exhaustive sample subgroups, the imbalance $\mathbf{D}'\mathbf{\Sigma}_s^{-1}\mathbf{D}$ is particularly transparent: It is a sum of non-negative terms, $\mathbf{D}'\mathbf{\Sigma}_s^{-1}\mathbf{D} = \sum_{j=1}^{J} C_j$, where C_j is the contribution to imbalance of the j:th group. This representation allows us to focus on each specific group in the data collection. Problematic groups are those for which C_j remains high throughout the data collection. This happens in the LCS 2009 data collection, as illustrated in Table 4. To reduce imbalance, one should direct the data collection so that all group contributions C_j are small in the end.

Section 6 described three experiments carried out by interventions in the LCS 2009 data file. A set of eight important sample subgroups was defined, and data collection was deemed terminated when subgroup response meets specified levels. These experiments, summarized in Table 10, showed that appropriate interventions in the data collection can bring considerable improvement – increased balance, reduced distance – compared with the actual LCS data collection. The cost savings realized by fewer call attempts might instead be used to improve other aspects of survey quality.

To use the conclusions from these experiments in practice, we must anticipate a "reasonable expectations" response rate to be used as a stopping rule for data collection in a group. In a regularly repeated survey such prior information is usually available, but this may not be the case in a survey carried out for the first time. But an assessment would be necessary.

In practice, whether or not responsive design has been used, nonresponse weighting adjustment for nonresponse will necessarily take place at the estimation stage. Balancing does not guarantee that the nonresponse bias is eliminated. The question then arises whether one could just as well delay the use of some of the auxiliary variables – those chosen to inform the responsive design – until the estimation stage, where they, usually together with other auxiliary variables, will determine the calibrated adjustment weights. In our opinion, although the responsive data collection can be an advantage, it does not eliminate the need for efficient calibrated weighting at the estimation stage. The question needs to be addressed in future research.

An issue in the LCS survey is the frame over-coverage; certain sample subgroups contain highly mobile people, some of whom may no longer reside in the country. It is clear that groups with chronically low response rate deserve particular attention in the data collection. In particular, improvements are needed to reach immigrants and younger persons whose style of living and interest in the survey may differ substantially compared with a majority of the population. Future savings could be realized by transferring interviewer effort from "easy-to-reach" respondents to the more problematic groups.

A central question is the choice of auxiliary variables to enter into the vector $\mathbf{x}_k$ that determines the imbalance $\mathbf{D}'\mathbf{\Sigma}_s^{-1}\mathbf{D}$ used to direct the data collection. This question needs to be addressed further in the future. Auxiliary variables are used first during the data collection and then with a somewhat different perspective at the estimation stage. In the data collection, the selected auxiliary variables serve to monitor the balance of the response set and the distance $dist_{r|nr}$ between respondents and nonrespondents. At the data collection stage, the auxiliary vector should thus be one that lends itself well to contrasting respondents with nonrespondents. At the estimation stage, on the other hand, the auxiliary vector serves to yield the most accurate estimates, particularly for the most important survey variables, and this vector is likely to contain more variables than the one used in monitoring the data collection.

8. References

Bethlehem, J., Cobben, F., and Schouten, B. (2011). Handbook of Nonresponse in Household Surveys. New York: Wiley.

Groves, R. (2006). Research Synthesis: Nonresponse Rates and Nonresponse Bias in Household Surveys. Public Opinion Quarterly, 70, 646–675. DOI: http://www.dx.doi.org/10.1093/poq/nfl033

Groves, R.M. and Heeringa, S.G. (2006). Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs. Journal of the Royal Statistical Society: Series A, 169, 439–457. DOI: http://www.dx.doi.org/10.1111/j.1467-985X.2006.00423.x

Hörngren, J., Lundquist, P., and Westling, S. (2008). Effects of Number of Call Attempts on Nonresponse Rates and Nonresponse Bias – Result from Some Case Studies at Statistics Sweden. Proceedings 24th International Methodology Symposium, Statistics Canada, Session 17, catalogue no. 11-522-XIE. Available at: http://www5.statcan.gc.ca. (accessed September 19, 2013).

Laflamme, F. (2009). Experiences in Assessing, Monitoring and Controlling Survey Productivity and Costs at Statistics Canada. Proceedings of the 57th Session of the International Statistical Institute, South Africa. (August 16–22). Available at: http://www.statssa.gov.za/isi2009/ScientificProgramme/IPMS/0049.pdf (accessed October 11, 2013).

Lundquist, P. and Särndal, C.-E. (2012). Aspects of Responsive Design for the Swedish Living Conditions Survey. R&D report 2012:1, Statistics Sweden. Available at: www.scb.se. (accessed September 19, 2013).

Mohl, C. and Laflamme, F. (2007). Research and Responsive Design Options for Survey Data Collection at Statistics Canada. Proceedings of the American Statistical Association, Section on Survey Research Methods. (July 29–August 2) Available at: https://www.amstat.org/sections/SRMS/Proceedings/y2007/Files/JSM2007-000421.pdf (accessed October 11, 2013).

Peytchev A, Baxter, R.K., and Carley-Baxter, L.R. (2009). Not All Survey Effort is Equal. Reduction of Nonresponse Bias and Nonresponse Error. Public Opinion Quarterly, 73, 785–806. DOI: http://www.dx.doi.org/10.1093/poq/nfp037

Peytchev, A., Riley, S., Rosen, J., Murphy, J., and Lindblad, M. (2010). Reduction of Nonresponse Bias in Surveys Through Case Prioritization. Survey Research Methods, 4, 21–29.

Peytcheva, E. and Groves, R.M. (2009). Using Variation in Response Rates of Demographic Subgroups as Evidence on Nonresponse Bias in Survey Estimates. Journal of Official Statistics, 25, 193–201.

Rao, R.S., Glickman, M.E., and Glynn, R.J. (2008). Stopping Rules for Surveys with Multiple Waves of Nonrespondent Follow-Up. Statistics in Medicine, 27, 2196–2213. DOI: http://www.dx.doi.org/10.1002/sim.3063

Schouten, B. and Bethlehem, J. (2009). Representativeness Indicator for Measuring and Enhancing the Composition of Survey Response. RISQ work package 8, deliverable 9. Available at: http://www.risq-project.eu/. (accessed September 19, 2013).

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.

Särndal, C.-E. (2011a). Dealing with Survey Nonresponse in Data Collection, in Estimation. Journal of Official Statistics, 27, 1–21.

Särndal, C.-E. (2011b). Three Factors to Signal Nonresponse Bias, with Applications to Categorical Auxiliary Variables. International Statistical Review, 79, 233–254. DOI: http://www.dx.doi.org/10.1111/j.1751-5823.2011.00142.x.

Särndal, C.-E. and Lundström, S. (2005). Estimations in Surveys with Nonresponse. New York: Wiley.

Särndal, C.-E. and Lundström, S. (2008). Assessing Auxiliary Vectors for Control of Nonresponse Bias in the Calibration Estimator. Journal of Official Statistics, 24, 251–260.

Särndal, C.-E. and Lundström, S. (2010). Design for Estimation: Identifying Auxiliary Vectors to Reduce Nonresponse Bias. Survey Methodology, 36, 131–144.

Wagner, J. (2008). Adaptive Survey Design to Reduce Nonresponse Bias. Ph.D. thesis, University of Michigan, Ann Arbor. Available at: http://www.google.se/-books?hl=sv&lr=&id=ilVbF2qPlTgC&oi=fnd&pg=PR3&dq=Wagner+(2008)+Adaptive+survey+design+to+reduce+nonresponse+bias&ots=iqKD3_BnEQ&sig=cRtrqb-zLogBQERLCHdBiuhyNu2k&redir_esc=y#v=onepage&q=Wagner%20(2008)%20Adaptive%20survey%20design%20to%20reduce%20nonresponse%20bias&f=false (accessed October 11, 2013).

Wagner, J. (2012). Research Synthesis: A Comparison of Alternative Indicators for the Risk of Nonresponse Bias. Public Opinion Quarterly, 76, 555–575. DOI: http://www.dx.doi.org/10.1093/poq/nfs032

Wagner, J. and Raghunathan, T.E. (2010). A New Stopping Rule for Surveys. Statistics in Medicine, 29, 1014–1024, DOI: http://www.dx.doi.org/10.1002/sim.3834