JOS 2287—4/5/2011—SANKARS—391109

Journal of Official Statistics, Vol. 27, No. 2, 2011, pp. 1–24

Indicators for Monitoring and Improving Representativeness of Response

Barry Schouten[1], *Natalie Shlomo*[2], *and Chris Skinner*[3]

The increasing efforts and costs required to achieve survey response have led to a stronger focus on survey data collection monitoring by means of paradata and to the rise of adaptive and responsive survey designs. Indicators that support data collection monitoring, targeting and prioritising in such designs are not yet available. Subgroup response rates come closest but do not account for subgroup size, are univariate and are not available at the variable level.

We present and investigate indicators that support data collection monitoring and effective decisions in adaptive and responsive survey designs. As they are natural extensions of R-indicators, they are termed partial R-indicators. We make a distinction between unconditional and conditional partial R-indicators. Unconditional partial R-indicators provide a univariate assessment of the impact of register data and paradata variables on representativeness of response. Conditional partial R-indicators offer a multivariate assessment.

We propose methods for estimating partial indicators and investigate their sampling properties in a simulation study. The use of partial indicators for monitoring and targeting nonresponse is illustrated for both a household and a business survey. Guidelines for the use of the indicators are given.

Key words: Auxiliary variable; business survey; nonresponse; response propensity.

1. Introduction

In the recent literature on survey nonresponse reduction and adjustment, much attention is paid to data collection monitoring and targeting of subpopulations in adaptive and responsive survey designs. Availability of register data and frame data is either very limited or provides little explanation of nonresponse behaviour. For this reason the focus has shifted partially towards data about the data collection process, so-called paradata (e.g., Kreuter et al. 2010). Paradata may consist of the outcomes of the various substeps in obtaining a response, like making contact, screening for eligibility or gaining participation, may represent the actual realisations of survey design features like the interviewer or the incentive used, or may include observations on the households and addresses themselves. Adaptive and responsive survey designs (see Groves and Heeringa

[1] Statistics Netherlands, Division of Methodology and Quality, PO Box 4000, 2270 JM Voorburg, The Netherlands. Email: bstn@cbs.nl
[2] University of Southampton, Highfield, Southampton SO17 1BJ, UK. Email: n.shlomo@soton.ac.uk
[3] University of Southampton, Department of Social Statistics, Highfield, Southampton SO17 1BJ, UK. Email: cjs@soton.ac.uk
Acknowledgments: This research was supported under the Representativity Indicators for Survey Quality (RISQ) project, funded by the European Commission. We thank colleagues on the RISQ project for valuable input of ideas to this research.

2006; Wagner 2008; Peytchev et al. 2010) employ the combined set of register data, frame data and paradata, in order to target and tailor the data collection strategy to the sample. For instance, households in urban areas may receive increased effort because their response is lower and addresses where interviewers observe physical impediments may be assigned to a different interview mode.

Both data collection monitoring and data collection targeting need quality and cost indicators to support decisions. To date, effective and easy-to-use indicators for targeting and prioritising sample cases are lacking. In this article we present indicators that can be used for data collection monitoring and the identification of relevant subgroups for adaptive and responsive designs. The indicators decompose the variation in response propensities and are directly linked to so-called R-indicators (see Schouten et al. 2009). For this reason we term them partial R-indicators.

Indicators for data collection monitoring and targeting require four properties. They should be easy to interpret, they should be based on available auxiliary data and survey data only, they should be relevant or in other words lead to effective survey designs, and they should allow analysis at different levels of detail. The last property is especially important when many auxiliary variables are available and the number of indicators increases very rapidly. In surveys with large samples, the ideal measure of nonresponse error might be taken to be nonresponse bias. However, this is rarely measurable directly and, moreover, most surveys are designed to produce a large number of survey estimates and the corresponding number of nonresponse biases might be too great to serve many needs of quality indicators, e.g., between-survey comparisons. The indicators that come closest to quality indicators are subgroup response rates, e.g., the response rates for rural versus urban areas. Response rates have the advantage of simplicity and ease of calculation (e.g., Biemer and Lyberg 2003, Section 3.5), but they also suffer from often having only a limited relation to nonresponse bias (e.g., Groves 2006; Groves and Peytcheva 2008). There are three main drawbacks to using subgroup response rates in monitoring and targeting nonresponse. First, subgroup response rates do not depend on the size of the subgroup, i.e., small subgroups may appear equally important as large subgroups. Second, subgroup response rates cannot be given at the variable level. As a consequence different variables cannot be evaluated and compared in their impact on response. Third, subgroup response rates are univariate and do not allow conditioning on other variables in an easy way. There is therefore a need for other quality indicators to supplement their use (see also Groves et al. 2008).

Schouten et al. (2009) proposed one alternative indicator, which they called an R-indicator, with "R" standing for representativeness. This indicator is designed to measure the degree to which the respondents to a survey resemble the complete sample. The contrast between the respondents and the sample is defined with respect to specified auxiliary variables. The R-indicator is motivated by the potential for systematic differences on auxiliary variables between respondents and nonrespondents to be predictive of nonresponse bias. The indicator will be most effective in capturing nonresponse bias in a survey estimate when the auxiliary variables are, in combination, strong predictors of the survey item(s) upon which the estimate is based. This will not always be the case (e.g., Kreuter et al. 2010), but these survey items are deliberately excluded from the definition of the R-indicator, since a key purpose is to support comparisons of surveys, which may have different items. When different surveys are compared, the same auxiliary variables need to be selected. However, when the representativeness of a single survey is evaluated, the selection of auxiliary variables may be based on their relation to the main survey items and may also include paradata. See Särndal and Lundström (2008) and Andridge and Little (2010) for some alternative possible approaches.

R-indicators themselves, like response rates, do not provide means to identify subgroups for targeting and prioritising. Partial R-indicators are designed to evaluate the contribution of a single specified auxiliary variable to a lack of representative response. They will be defined in relation to this variable or in terms of the categories of the variable when it is categorical. We shall make a distinction between unconditional and conditional partial R-indicators. The definitions we shall present are designed to supplement and be used in conjunction with R-indicators.

In this article we present indicators but do not give a detailed account of how to go from monitoring data collection to interventions in data collection. Loosveldt and Beullens (2009) discuss how to use partial R-indicators for the identification of effective treatments. Partial R-indicators have also been used on an experimental basis in data collection at Statistics Netherlands (Luiten and Wetzels 2010) and Statistics Norway (Kleven et al. 2010), as part of the RISQ project (http://www.risq-project.eu). In particular, Luiten and Wetzels (2010) found that they could be used to help design interventions in a household survey which significantly increased representativeness, while maintaining the response rate and substantially reducing costs.

In Section 2 we define the partial indicators and discuss their estimation. The sampling properties of the estimators are assessed in a simulation study in Section 3. Section 4 provides guidelines for the use of partial R-indicators. Applications to a household and a business survey are presented in Section 5, followed by some concluding discussion in Section 6.

2. Partial R-indicators

In this section we present definitions of partial indicators, designed to evaluate the contribution of a single specified auxiliary variable Z to a lack of representative response. Our primary interest is when Z is a component of the vector X used to define the response propensities, but we are also interested in the case when this does not hold. We shall only consider the case when Z is categorical and leave the case of continuous Z to further work.

We introduce two types of partial indicators. We define unconditional partial indicators in Section 2.3 to measure the contribution of single variables to a lack of representative response. Conditional partial indicators are defined in Section 2.4 to measure the contribution of single variables to a lack of representative response *given* other variables, i.e., with respect to conditional representative response.

Both types of indicators are based on definitions of R-indicators which are reviewed in Section 2.1 together with basic notation. Some further preliminaries are set out in Section 2.2. The definitions of partial indicators are set out in Sections 2.3 and 2.4 and estimation is considered in Section 2.5.

Table 2.1 shows an example taken from the Dutch Labour Force Survey (LFS). The response rate and R-indicator are stable over the two years investigated. The question is

JOS 2287—4/5/2011—SANKARS—391109

what subgroups may be identified to further improve response to the LFS. We use this example throughout Section 2 for illustration and return to the example in Section 5 where we detail the analyses.

2.1. Response Propensities and R-indicators

Let U denote the set of population units and s the set of sample units. Define a response indicator variable R_i which takes the value 1 if unit i in the population responds and the value 0 otherwise. The *response propensity* is defined as the conditional expectation of R_i given the value x_i of the vector X of auxiliary variables:

$$\rho_X(x_i) = E(R_i = 1|X = x_i) = P(R_i = 1|X = x_i) \tag{1}$$

We assume that the values x_i are known for all sample units, i.e., for both respondents and nonrespondents, and can include both specified variables and survey fieldwork conditions. Thus, X may include variables such as mode of data collection, whether there has been an advance contact, the number of callbacks, reissuance constraints etc. The response propensity is thus defined conditional on design choices which have been previously made at a particular point in time and the propensity might change over time for a given unit if new design choices are introduced.

Schouten et al. (2009) define the R-indicator, $R(\rho_X)$, as:

$$R(\rho_X) = 1 - 2S(\rho_X) \tag{2}$$

where $\bar{\rho}_X = N^{-1}\sum_U \rho_X(x_i)$ and $S^2(\rho_X) = 1/(N-1)\sum_U[\rho_X(x_i) - \bar{\rho}_X]^2$ are the population mean and variance, respectively, of the response propensities ρ_X. It can be shown that $S(\rho_X)$ lies between 0 and 0.5 and the transformation from $S(\rho_X)$ to $R(\rho_X)$ in (2) is designed to ensure that the R-indicator lies between 0 and 1, with 1 denoting fully representative response and 0 indicating the least possible representativity. Schouten et al. (2009) discuss some associated measures, in particular:

$$B(X) = \frac{1 - R(\rho_X)}{2\bar{\rho}_X} \tag{3}$$

which is shown to be the maximal absolute relative bias when estimating a population mean of a survey variable, under the scenario where nonresponse correlates maximally to this variable.

Example 2.1 Consider two simple, arbitrary auxiliary variables, job (yes/no) and nonnative (yes/no). The following population distributions and estimated response

Table 2.1. Response rate and R-indicator for the 2006 and 2008 LFS

	LFS	
	2006	*2008*
Response rate	63.2%	63.4%
R-indicator	0.889	0.884

propensities are taken from the LFS 2008. In Section 2.5 we provide details about the Q2 estimation of the propensities (Tables 2.2 and 2.3).

The overall response rate is 63.4%. The standard deviation of the estimated response propensities given the two variables equals 0.046 and the R-indicator is 0.908. The R-indicator in Table 1.1 is slightly lower as it is based on a larger set of auxiliary variables.

2.2. Preliminaries for Defining Partial Indicators

Let Z denote the auxiliary variable for which we should like to define the partial indicator. We first assume Z is categorical with categories $k = 1, 2, \ldots, K$. Partial indicators are denoted by $P(Z,\rho_X)$ for the overall contribution of variable Z and $P(Z = k, \rho_X)$ for the contribution of a single category k of Z. In both cases indicators are computed given response propensities defined with respect to X.

In Sections 2.3 and 2.4 we employ the ANOVA decomposition with respect to Z to the variance, $S^2(\rho_X)$, underlying the R-indicator

$$S^2(\rho_X) = S_b^2(\rho_X|Z) + S_w^2(\rho_X|Z) \tag{4}$$

where

$$S_b^2(\rho_X|Z) = \frac{1}{N-1}\sum_{k=1}^{K} N_k(\bar{\rho}_{X,k} - \bar{\rho}_X)^2 \cong \sum_{k=1}^{K} \frac{N_k}{N}(\bar{\rho}_{X,k} - \bar{\rho}_X)^2 \text{ and} \tag{5}$$

$$S_w^2(\rho_X|Z) = \frac{1}{N-1}\sum_{k=1}^{K} \sum_{i \in U_k} (\rho_X(x_i) - \bar{\rho}_{X,k})^2 \tag{6}$$

are the within and between variances, U_k is the set of units in category k, N_k is the size of U_k, and $\bar{\rho}_{X,k}$ is the average response propensity in U_k.

2.3. Unconditional Partial Indicators

The unconditional partial R-indicator for Z is taken as the Euclidean distance to representative response as defined by Schouten et al. (2009), i.e., as equal response propensities. The unconditional partial indicator for the variable Z then equals

$$P_u(Z, \rho_X) = S_b(\rho_X|Z) \tag{7}$$

where $S_b(\rho_X|Z)$ is the square root of (5). This indicator is necessarily nonnegative. From (4), it is bounded above by $S(\rho_X)$, which itself is bounded above by 0.5. The larger the value of $P_u(Z,\rho_X)$, the greater the contribution of the variable Z to the lack of representativeness. When the indicator is zero, Z does not contribute to selective

Table 2.2. Population distribution

	No job	Job	
Native	24.5%	55.7%	80.2%
Nonnative	8.0%	11.8%	19.8%
	32.5%	67.5%	

Table 2.3. Response propensities

	No job	Job	
Native	63.4%	66.5%	65.6%
Nonnative	51.4%	57.2%	54.9%
	60.4%	64.9%	63.4%

nonresponse. At the upper bound with $P_u(Z, \rho_X) = S(\rho_X)$, the variable Z accounts entirely for the lack of representativeness arising from X.

The unconditional partial indicator for category k of Z is defined as:

$$P_u(Z = k, \rho_X) = \sqrt{\frac{N_k}{N}}(\bar{\rho}_{X,k} - \bar{\rho}_X) \tag{8}$$

It follows from (5) that the variable-level indicator $P_u(Z,\rho_X)$ in (7) is the squared root of the sum of squared values of the category-level indicators $P_u(Z = k, \rho_X)$ across k. Hence the $P_u(Z = k, \rho_X)$ may be used to elaborate the lack of representativeness arising from the variable Z. The measure $P_u(Z = k, \rho_X)$ may be positive or negative, indicating either over-representation or under-representation of the category, respectively. It may take values between -0.5 and $+\ 0.5$, where again a value of zero indicates no contribution. Used in conjunction with the R-indicator, these partial indicators assist in the individual analysis of representativity and can be especially useful for field work monitoring in localising subgroups for targeted data collection.

Example 2.2 Consider the setting of example 2.1. We compute the unconditional variable-level and category-level partial R-indicators for job status and ethnicity
Q2 separately and for the combined four-category variable (Tables 2.4 and 2.5).

The variable-level partial R-indicator for Ethnicity is 0.043 and is, therefore, close to the overall standard deviation of 0.046. The strongest positive impact comes from natives with a job and the strongest negative impact from nonnatives without a job.

2.4. *Conditional Partial Indicators*

For conditional partial indicators, we assume that Z is included in the vector of variables X used to define the response propensities. We write X^- as that part of X excluding Z so that we may write: $X = (X^-, Z)$. In this article, we assume that X^- is made up of categorical variables, defining a set of strata U_l, $l = 1, \ldots, L$.

We first introduce the definition of *conditional representative response*. The response to a survey is called conditionally representative for Z given X^- when the conditional response propensities are equal for all choices of X^-. Hence, when response is conditionally representative, the propensities for X equal the propensities for X^-.

Table 2.4. Variable-level

	$Pu(Z)$
Job status	0.021
Ethnicity	0.043

Table 2.5. Category-level

$Pu(Z = k)$	No job	Job	
Native	0.000	0.023	0.019
Nonnative	−0.034	−0.021	−0.038
	−0.017	0.012	0.046

This definition allows us to analyse the contribution of variables to nonrepresentative response adjusted for the impact of other variables.

Analogous to the R-indicator and unconditional partial R-indicator, the conditional partial R-indicator is taken as the Euclidean distance to conditional representative response, i.e., the Euclidean distance between ρ_X and ρ_{X^-}. Consequently, the conditional variable-level partial R-indicator again amounts to a standard deviation, the within standard deviation given X^-

$$P_c(Z, \rho_X) = \sqrt{S_w^2(\rho_X | X^-)} = S_w(\rho_X | X^-) \tag{9}$$

where $S_w^2(\rho_X | X^-)$ is defined as in (6), with the strata U_l replacing the subpopulations U_k defined by the categories k of Z. The larger the value of $P_c(Z,\rho_X)$, the greater must be the variability of the response propensities within the strata. Since this variation can only be attributable to Z (given the definition of X^-), we may interpret $P_c(Z,\rho_X)$ as measuring the contribution of Z to the R-indicator after first controlling for the contribution of all remaining variables, denoted by X^-. Again (9) takes values between 0 and 0.5, where a value of zero means no conditional contribution of Z.

Assuming again that Z is categorical, let δ_k be the 0–1 dummy variable that is equal to 1 if $Z = k$ and 0 otherwise. The conditional partial indicator for category $Z = k$ is defined as the within standard deviation of $\rho_X(x_i)$ restricted to units in this category:

$$P_c(Z = k, \rho_X) = \sqrt{\frac{1}{N-1} \sum\nolimits_{l=1}^{L} \sum_{U_l} \delta_{k,i} [\rho_X(x_i) - \bar{\rho}_{X,l}]^2} \tag{10}$$

where $\bar{\rho}_{X,l}$ is the average of the response propensities $\rho_X(x_i)$ in Stratum l of X^-. It follows from (6) that the variable-level indicator $P_c(Z,\rho_X)$ in (9) is the squared root of the sum of squares of the category-level indicators $P_c(Z = k, \rho_X)$ across categories k. Hence the $P_c(Z = k, \rho_X)$ enable explanation of the lack of representativeness reflected by $P_c(Z,\rho_X)$. The category-level indicator ranges from 0 to 0.5, where a value of zero implies no conditional contribution of the category.

Example 2.3 Consider again the setting of Example 2.1. We now compute the conditional partial R-indicators for both variables (Tables 2.6 and 2.7).

Q2

The variable-level partial R-indicator for job status and ethnicity dropped from 0.021 to 0.018 and from 0.043 to 0.041, respectively, when conditioning on the other variable. Hence, both variables do not show strong collinear response behaviour and both variables can be viewed as having a separate impact on representativeness. From the category-level indicators we conclude that the strongest conditional contribution comes from nonnatives without a job.

Table 2.6. Variable-level

	$Pc(Z)$
Job status	0.018
Ethnicity	0.041

2.5. Estimation

We base the estimation of the propensities on a logistic regression model, where β denotes the vector of regression coefficients and x_i the corresponding vector of explanatory variables, which may involve transformation of the original auxiliary variables (e.g., by including interaction terms). The estimator of the response propensity is $\hat{\rho}_X(x_i) = \exp(x_i{}'\hat{\beta})/[\exp(x_i{}'\hat{\beta}) + 1]$, where $\hat{\beta}$ is an estimator of β. The estimator of the variance of the response propensities is $\hat{S}^2(\hat{\rho}_X) = 1/(N-1)\sum_s d_i(\hat{\rho}_X(x_i) - \hat{\bar{\rho}}_X)^2$, where $d_i = \pi_i^{-1}$ is the design weight and $\hat{\bar{\rho}}_X = 1/N \sum_s d_i \hat{\rho}_X(x_i)$. We estimate the population-level R-indicator in (2) by $\hat{R}(\hat{\rho}_X) = 1 - 2\hat{S}(\hat{\rho}_X)$. We use design weights so that this indicator is estimated approximately unbiasedly.

We estimate the partial indicators in a similar way, plugging in the estimated propensities. For example, we estimate the within and between variances in Expressions (5) and (6) by:

$$\hat{S}_w^2(\hat{\rho}_X|Z) = \frac{1}{N-1}\sum_{l=1}^{L}\sum_{i \in s_l} d_i(\hat{\rho}_X(x_i) - \hat{\bar{\rho}}_{X,l})^2 \tag{11}$$

$$\hat{S}_b^2(\hat{\rho}_X|Z) = \sum_{l=1}^{L}\frac{\hat{N}_l}{N}(\hat{\bar{\rho}}_{X,l} - \hat{\bar{\rho}}_X)^2 \tag{12}$$

where s_i is the set of sample units in Stratum l, and $\hat{N}_l = \sum_{s_l} d_i$ is the estimated population size of that stratum.

3. Simulation Study

The partial indicators defined in Section 2 enable the R-indicators to be analysed according to different subsets of the population. The benefits of increasing analytic detail need to be balanced, however, against the potential for greater estimation error as the subsets and their associated sample sizes become smaller. In this section we conduct an empirical investigation of this estimation error via a simulation study based upon a population obtained from the 1995 Israel Census. The estimators of the response propensities, and hence the indicators, are based upon samples (combining respondents and nonrespondents) and hence the magnitude of the estimation error (measured by both

Table 2.7. Category-level

$Pc(Z = k)$	No job	Job	
Native	0.000	0.023	0.019
Nonnative	0.034	0.021	0.037
	0.014	0.011	

bias and variance) may be expected to depend upon the sample size. This is therefore varied in the simulation study.

The 1995 Israel Census was based on two types of questionnaires: a short form for every household and a long form that was distributed to every fifth household in addition to the short form. Census questionnaires were delivered and collected by Census enumerators who visited every household. The simulation study is based on a population defined by all individuals aged 15 and over at the time of the Census who responded to the long form questionnaire ($N = 753{,}711$).

For this simulation, population response propensities $\rho_X(x_i)$ were calculated using a 2-step process:

1. Response rates were specified according to the following auxiliary X variables based on those achieved from a recent Labour Force Survey with an income component in Israel: child indicator, income from earnings groups, age group, gender, number of persons and locality type. Based on these response rates, initial population values of the response indicator R_i were generated.
2. Using the initial values of the response indicator as the dependent variable, we fit a logistic regression model on the population using the above explanatory variables including an interaction between the number of persons and locality type. The predictions from this model serve as the "true" response propensities $\rho_X(x_i)$.

Table 3.1 presents the "true" response rates generated in the population for the different variables implied by the population response propensities calculated from the logistic regression model in Step 2. The overall response rate in the population is 78.5% and the true R-indicator is 86.8%.

Response propensities and partial R-indicators were estimated from 1,000 samples drawn from the population. We drew 1,000 samples under three sample fractions: 1:50 (sample size is 15,074), 1:100 (sample size is 7,537) and 1:200 (sample size is 3,768), using simple random sampling. For each of the 1,000 samples, a new set of respondents was generated using the response propensities. The study therefore captures the full variability in estimation error arising from both sampling and nonresponse. We present results through a series of box plots in Figures 3.1 to 3.8. Box plots show the mean, the median and the spread of the distribution for the estimated partial R-indicators across the 1,000 simulations. In each figure, the variables are labelled according to the name of the variable (or category). Each variable has three box plots associated with it according to the sampling fraction, which we denote by "L" for the large sample (1:50), "M" for the medium sample (1:100) and "S" for the small sample (1:200). To save space, we present results only for the Z variables age group, number of persons in household and type of locality, where these were selected since they had the largest true values of the variable-level conditional partial indicator and include the values of Z with the two largest values of the unconditional partial indicator.

3.1. The Unconditional Partial Indicators

We first present estimates of the unconditional indicator $P_u(Z,\rho_X)$, defined in (7), in Figure 3.1. The estimated values of $P_u(Z,\rho_X)$ are seen to be roughly unbiased, although

Table 3.1. Summary of response rates in simulated population according to auxiliary variables

Variable	*Category*	*Response rate*
Total		78.5%
Gender	Male	77.4%
	Female	79.5%
Children	None	77.3%
	1 +	82.2%
Type of locality	3 large cities	74.5%
	Urban	79.8%
	Rural	78.1%
Age group	15–17	84.0%
	18–34	74.3%
	35–44	74.7%
	45–54	78.0%
	55–69	79.9%
	70 +	84.3%
Persons in household	1	74.3%
	2	75.7%
	3	82.3%
	4	85.9%
	5	76.6%
	6 +	72.5%
Income groups	No income	79.5%
	Low	77.3%
	Medium	76.9%
	High	76.7%

there is a slight tendency for increasing upward bias as the sample sizes gets smaller. As expected, larger sample sizes result in smaller variation in the estimated values. The figure demonstrates that, for the kinds of sample sizes and true values considered here, the variability of the estimation error for the partial indicators (as measured by the interquartile range, say) tends to be less than the difference between the average values of the estimators. It can be seen in Figure 3.1 that the type of locality has a lower unconditional partial indicator than the other variables, which means less variability of response propensities between the categories.

In Figures 3.2 through 3.4, we present estimates of the category-level partial indicator $P_u(Z = k, \rho_X)$, defined in (8), for different categories k of the Z variables, age group, type of locality and number of persons. Values of $P_u(Z = k, \rho_X)$ indicate categories of variables that are underrepresented (below zero) and overrepresented (above zero). Examples of underrepresented groups in this simulation are: persons aged 18–44, 3 large cities and small household sizes of 1 or 2 and large household sizes 5 and over. The results of the underrepresented groups also coincide with lower response rates as seen in Table 3.1. The estimates show little evidence of bias in the figures. As in Figure 3.1 the sampling errors of these estimates seem small enough, at least for these sample sizes and true values, for differences between the categories to be estimated with reasonable precision.

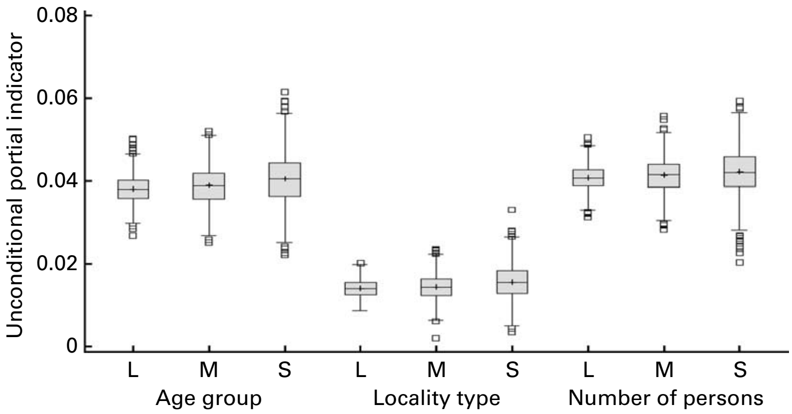

Fig. 3.1. Unconditional partial indicator $P_u(Z,\rho_X)$ for Z = age, type of locality and number of persons in household for three sampling fractions (1:50 (L), 1:100 (M) and 1:200 (S)). Population values are: age = 0.0369, type of locality = 0.0135, number of persons = 0.0401

3.2. The Conditional Partial Indicators

We next consider estimates of the partial indicator $P_c(Z,\rho_{X,Z})$, defined in (9). Figure 3.5 shows the performance of the estimated value $\hat{P}_c$ for the three choices of Z and for three sample sizes. There is evidence of upward bias, which increases as the sampling fraction decreases. The smallest sample size (1:200 sampling fraction) results in over-estimation of the contribution to the lack of representativity compared to the other sample sizes. The 1:200 sample fraction overestimates by approximately 4% compared to the 1:50 sample fraction for age group and number of persons and by 13% for locality type. The dispersion in the values of $\hat{P}_c$ is similar to that for $\hat{P}_u$ in Figure 3.1, although the true values for the three variables are now more similar and the sampling variation tends to dominate the differences between the variables.

The variable-level conditional partial indicators for age group and number of persons are about the same as their corresponding unconditional partial indicators (compare

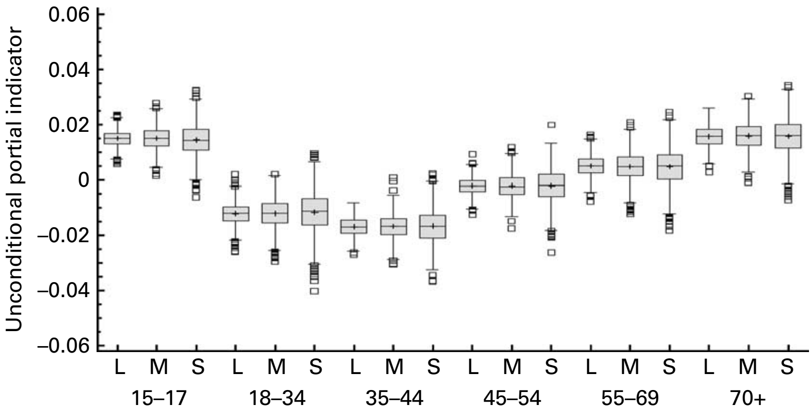

Fig. 3.2. Unconditional partial indicator $P_u(Z = k, \rho_X)$ for categories of age group for three sampling fractions (1:50 (L), 1:100 (M) and 1:200(S)). Population values are: 15–17 = 0.0148, 18–34 = −0.0122, 35–44 = −0.0169, 45–54 = −0.0022, 55–69 = 0.0049, 70+ = 0.0156

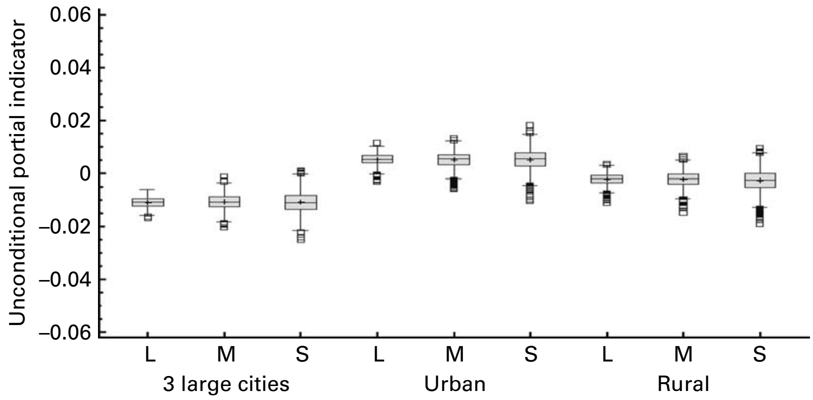

Fig. 3.3. Unconditional partial indicator $P_u(Z = k, \rho_X)$ for categories of type of locality for three sampling fractions (1:50 (L), 1:100 (M) and 1:200 (S)). Population values are: 3 large cities $= -0.0109$, urban $= 0.0051$, rural $= -0.0022$

Figure 3.1 and Figure 3.5), suggesting that these variables have a separate impact on representativity. Type of locality, however, has a slightly smaller conditional partial indicator than the unconditional partial indicator and therefore some part of the contribution of type of locality to response behaviour is accounted for by the other variables.

In Figures 3.6 through 3.8, we present estimates of the category-level partial conditional indicator $P_c(Z = k, \rho_X)$ defined in (10), for different categories k of the Z variables, age group, type of locality and number of persons. Lower values of $P_c(Z = k, \rho_X)$ indicate categories of variables that have high collinear response behaviour. Examples of this property are persons aged 18–34 and 70 and over, 3 large cities and household sizes of 4 persons. The estimates show evidence of upward bias as the sample sizes get smaller. As in Figure 3.5, the sampling errors of these estimates seem small for differences between the categories to be estimated with reasonable precision.

In this simulation study, we assessed the estimation error with respect to bias and variance of the partial indicators. As seen in the variation of the partial indicators

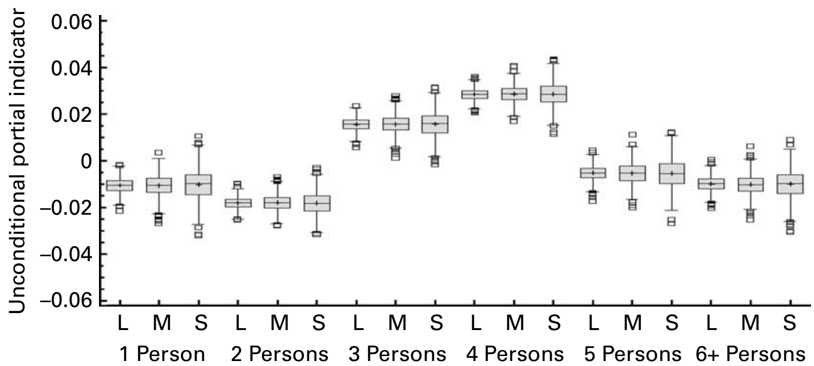

Fig. 3.4. Unconditional partial indicator $P_u(Z = k, \rho_X)$ for categories of number of persons in the household for three sampling fractions (1:50 (L), 1:100 (M) and 1:200 (S)). Population values are: $1 = -0.0107$, $2 = -0.0181$, $3 = 0.0156$, $4 = 0.0283$, $5 = -0.0052$, $6+ = -0.0099$

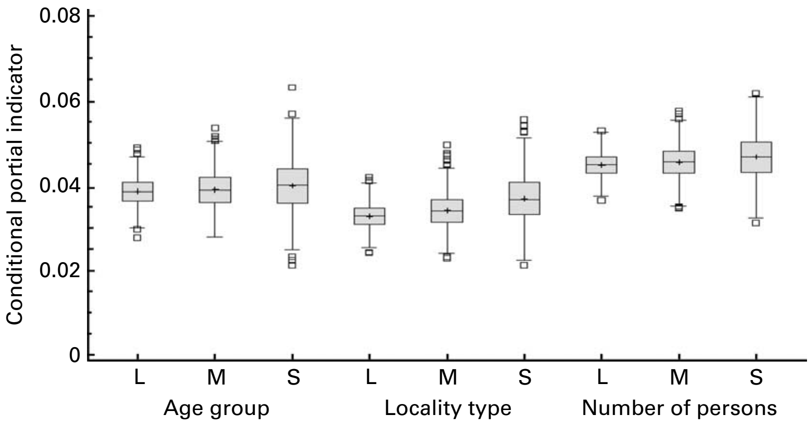

Fig. 3.5. Conditional partial indicator $P_c(Z, \rho_{X,Z})$, for Z = age group, type of locality or number of persons in household and for three sampling fractions (1:50 (L), 1:100 (M) and 1:200 (S)). Population values are: age = 0.0379, type of locality = 0.0311, number of persons = 0.0446

presented in the boxplots in Figures 3.1 to 3.8, the variances should not be neglected as they can be relatively large, especially for small sample sizes. In Shlomo et al. (2008), a bias adjustment was developed for the R-indicator. As seen in the results of this simulation study, we obtain some bias in the estimates of the partial R-indicators as the sample sizes get smaller. The bias is bigger for the conditional partial R-indicators, which is not surprising as these indicators arise from a more detailed stratification than the unconditional indicators.

4. How to Use Partial R-indicators in Monitoring and Targeting Nonresponse?

R-indicators and partial R-indicators describe multivariate breakdowns of nonresponse behaviour on a selected set of variables from register data, frame data and paradata into simple measures of representativeness. But how to use these measures? And equally important, given the dependence of the indicators on the set of auxiliary variables, how to

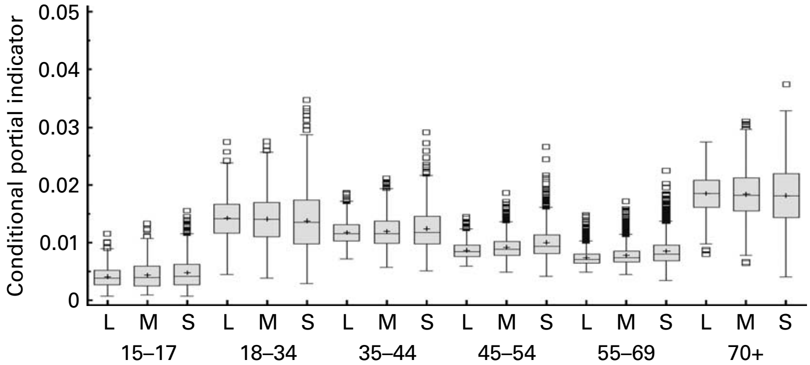

Fig. 3.6. Conditional partial indicator $P_c(Z = k, \rho_X)$ for categories of age group for three sampling fractions (1:50 (L), 1:100 (M) and 1:200 (S)). Population values are: 15–17 = 0.0040, 18–34 = 0.145, 35–44 = 0.0114, 45–54 = 0.0080, 55–69 = 0.0068, 70 + = 0.0187

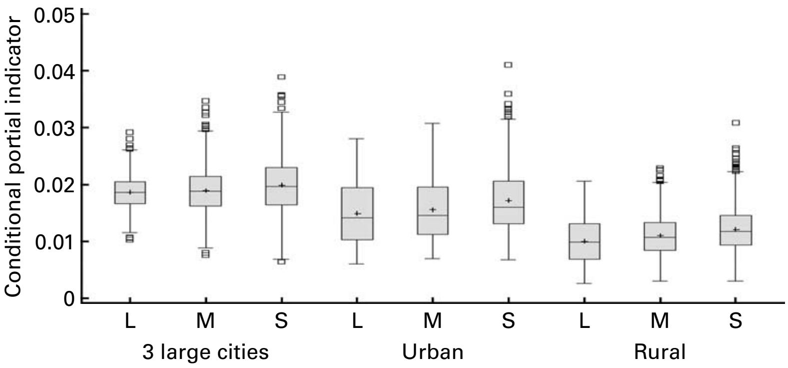

Fig. 3.7. Conditional partial indicator $P_c(Z = k, \rho_X)$ for categories of type of locality for three sampling fractions (1:50 (L), 1:100 (M) and 1:200 (S)). Population values are: 3 large cities = 0.0181 urban = 0.0140 rural = 0.0084

select the variables and their categories? In this section we provide some guidelines with regard to both questions.

As indicated in the introduction, ideally indicators allow analysis of nonresponse on different levels of detail. The R-indicators, the unconditional and conditional variable-level partial indicators and the unconditional and conditional category-level partial indicators allow such an analysis. Monitoring and possibly intervening may be done using a number of steps that can be repeated during data collection:

1. Compute the R-indicator and compare to previous waves of the same survey.
2. Assess the unconditional variable-level partial R-indicators for all selected auxiliary variables; the variables that have the highest values are the strongest candidates for being involved in design changes and increased follow-up efforts.
3. Assess the conditional variable-level partial R-indicators for all selected auxiliary variables; the conditional values are needed in order to check whether some of the

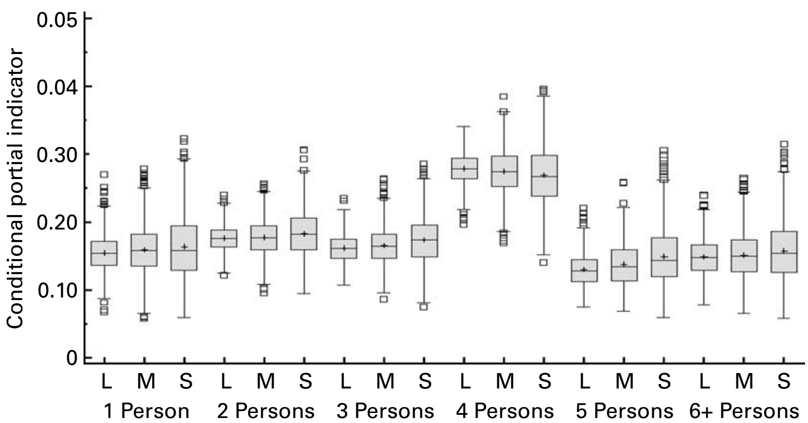

Fig. 3.8. Conditional partial indicator $P_c(Z = k, \rho_X)$ for categories of number of persons in the household for three sampling fractions (1:50 (L), 1:100 (M) and 1:200 (S)). Population values are: 1 = 0.0155, 2 = 0.0174, 3 = 0.0158, 4 = 0.0289, 5 = 0.0121, 6+ = 0.0147

variables are strongly collinear. If indicator values remain high, then the strongest variables are selected. If indicator values vanish by conditioning, then it is sufficient to focus only on a subset of the variables.

4. Repeat Steps 1 and 2 but now for the category-level partial R-indicators and for the auxiliary variables selected in Step 3 only; the subgroups that need to be targeted in design changes are those categories that have large negative unconditional values and large conditional values.

The subgroups that are selected in Step 4 may form the input to responsive and adaptive survey designs. A few remarks are in place. First of all, it is crucial to realise that any attempt to improve the representativeness of response must be viewed jointly with the associated costs and with the design features that can be changed. A survey that has a low budget may accept different levels for the indicators than surveys with a high budget. Also, for example, the options to increase efforts are different in web and face-to-face surveys. Second, the values of the indicators must be confronted with their standard errors before it can be concluded that contributions to nonrepresentative response are signifcant. Hence, analytic approximations of standard errors are needed. Third, one may choose to intervene during data collection or to change the design for future waves of the survey. The first option is usually referred to as a responsive survey design and requires careful monitoring of both response representativeness and costs. The second option is termed an adaptive survey design. Such a design assumes that historical response propensities apply to future waves and hence can be used as input parameters to a mathematical optimization of representativeness given constraints on costs. Responsive designs need thresholds for prioritising sample cases. Adaptive designs require robustness of the estimated response propensities.

The selection of auxiliary variables is important when using the indicators. When indicators are used to compare multiple surveys, and partial R-indicators could be part of such a comparison, then generally available auxiliary variables should be selected for which literature has shown that they relate to nonresponse in most if not all surveys. For example, Statistics Netherlands uses age, type of household, urbanicity of address, ethnicity, average house value at postal code and job status to make general assessments of representativeness of its surveys. In Section 5 we focus on monitoring data collection and on identifying subgroups that are candidates for targeting and increased follow-up efforts in a single survey. When monitoring and imptoving response, it is imperative to select variables that 1) represent the main publication domains, 2) relate to the key survey items, and/or 3) relate to the survey-specific motives for and causes of nonresponse. The last two types of auxiliary variables should include paradata observations that are specifically designed for the survey under investigation. Since the number of variables and their numbers of categories affect the sampling variation of the partial R-indicators, it is important to use parsimonious selections of variables and categories.

5. Applications

In this section we present two applications: a household survey, the Dutch LFS, and a business survey, the Dutch STS. In both applications the main questions are: what variables have the strongest impact on representativeness of response and what subgroups should be monitored and targeted in adaptive survey designs.

5.1. The 2006 and 2008 LFS

The Dutch Labour Force Survey (LFS) is a monthly household survey conducted by face-to-face interviews. The key statistics of the LFS are the percentage of persons employed, unemployed and not in the labour force in the Netherlands and in various regional and socio-demographic subpopulations. The target population consists of persons of 15 years and older: the potential labour force population. Persons 65 years and older are subsampled as most persons in this group have retired and belong to the not in the labour force population. In the analysis we omit the persons 65 years and older. The contact strategy for the LFS consists of a maximum of six visits to the address. If no contact was made at the sixth visit, then the address is processed as a noncontact.

Table 1.1 in the introduction presents the response rates and the R-indicator for the 2006 and 2008 LFS. Both the response rate and R-indicator were stable for the two years. In this section we compare the partial R-indicators for both years. We compute partial R-indicators for contact and for overall response. We employ job status according to tax authorities, age and average value of houses at postal code area as auxiliary variables. All three variables relate strongly to the employment status. Partial R-indicators are also computed for response given contact using, in addition to the three register variables, the number of visits to contact. The contactability of a person is also known to relate strongly to employment status. Persons that are harder to reach more often are employed. Table 5.1.1 presents contact and response rates, R-indicators, maximal biases and variable-level partial R-indicators for age, house value and job status.

Table 5.1.1 shows that the contact representativeness and contact rates hardly changed from 2006 to 2008. As a consequence the maximal bias is comparable for these two years. The partial R-indicators are also similar in size and show that age and average house value have the largest impact on representativeness. The impact of job status is very small. The picture for the representativeness of response is similar: the R-indicator is almost the same for 2006 and 2008. Consequently, the overall impact on representativeness from

Table 5.1.1. Contact and response rates, R-indicators, maximal bias and variable-level partial R-indicators (Pu = unconditional, Pc = conditional) for the LFS 2006 and 2008

		Contact		*Response*	
		2006	*2008*	*2006*	*2008*
Rate		94.1%	94.9%	63.2%	63.4%
R-indicator		0.943	0.940	0.889	0.884
Maximal bias		0.030	0.032	0.088	0.091
Pu	Age	0.022	0.021	0.033	0.013
	House value	0.021	0.021	0.043	0.052
	Job	0.002	0.002	0.019	0.021
Pc	Age	0.019	0.019	0.031	0.017
	House value	0.018	0.021	0.036	0.050
	Job	0.001	0.002	0.024	0.023

participation given contact must have been the same too for both years. However, the contribution of the single variables has changed. In 2008 the response is more representative with respect to age but less representative with respect to average house value. The value for job status did not change.

Hence, between 2006 and 2008 the nature of the LFS response changed for participation.

For this reason we investigate the representativeness of response given contact. We first add the number of visits needed to contact the household. Table 5.1.2 shows the variable-level partial R-indicators for the LFS 2008 after two, four and all visits. Note that the number of visits has an increasing number of categories as data collection evolves.

From Table 5.1.2 we conclude that the number of visits is the strongest variable in all cases. Its variable-level partial R-indicators are considerably larger than for age, house value and job status. The R-indicator, participation rate given contact and maximal bias are relatively stable; the cases that require more visits show similar response and refusal behaviour. The unconditional and conditional indicator values are very similar. Hence, the four variables have a close to orthogonal impact on the representativeness and can be viewed as separate components of selective response.

Table 5.1.3 presents category-level partial R-indicator values for participation in increasing order for the unconditional partial R-indicators. The categories with large negative unconditional values and large conditional values are candidates for targeting and prioritising in adaptive survey designs. Of the 29 subpopulations formed by the categories of the auxiliary variables, 15 have a negative unconditional value. By far the most negative value is for persons that called Statistics Netherlands before visits to the address had been started. These persons call the phone number on the advance letter and mostly refuse further participation. The other subpopulations that have large negative scores are persons that required six visits, persons living in postal codes with an average house value between 100 and 150 thousand Euros, persons that do not have a job according to the tax authorities and persons that required five visits. In almost all cases the conditional and unconditional partial R-indicators have a similar size in an absolute sense, i.e., the corresponding

Table 5.1.2. Participation rates (given contact was established), R-indicators, maximal bias and variable-level partial R-indicators (Pu = unconditional, Pc = conditional) for the LFS 2008

		Participation		
		Two visits	*Four visits*	*Six visits*
Rate		67.9%	67.6%	66.9%
R-indicator		0.792	0.811	0.807
Maximal bias		0.153	0.140	0.144
Pu	Age	0.016	0.015	0.014
	House value	0.037	0.040	0.041
	Job	0.021	0.022	0.023
	Number of visits	0.093	0.082	0.083
Pc	Age	0.020	0.018	0.018
	House value	0.037	0.039	0.040
	Job	0.020	0.022	0.023
	Number of visits	0.092	0.081	0.082

Table 5.1.3. Category-level unconditional (Pu) and conditional partial R-indicators (Pc) for age, average house value, job status and number of visits in increasing order

Category	*Pu*	*Pc*	*Category*	*Pu*	*Pc*
Person called before 1st visit	−0.076	0.075	15–19 years	0.000	0.001
Six visits	−0.019	0.019	House value 200–250	0.000	0.002
No job	−0.019	0.019	House value >500	0.004	0.005
House value 100–150	−0.018	0.016	60–64 years	0.004	0.009
Five visits	−0.016	0.016	20–24 years	0.005	0.003
House value 75–100	−0.012	0.011	25–29 years	0.006	0.007
House value 150–200	−0.010	0.011	House value 250–300	0.006	0.005
40–44 years	−0.007	0.009	30–34 years	0.006	0.006
45–49 years	−0.006	0.006	House value 400–500	0.010	0.010
House value 0–75	−0.006	0.006	Two visits	0.011	0.011
50–54 years	−0.003	0.004	House value 300–400	0.013	0.013
35–39 years	−0.002	0.002	Job	0.013	0.013
55–59 years	−0.002	0.002	One visit	0.020	0.021
Three visits	−0.001	0.002	No house value available	0.028	0.027
Four visits	−0.001	0.001			

subpopulations have a separate impact on representativeness. There are some patterns in the partial R-indicator values. The unconditional values are increasing in line with the number of visits (apart from the group that called), the average house values above 200 thousand Euros perform better and persons between 35 and 59 years do worse.

From Table 5.1.3 we identify various subpopulations that need more effort during data collection. The call centre staff may receive special training to convert persons that call them. Interviewers may get additional instructions to deal with persons persons without a job, persons that require more than three visits and persons living in areas with a low average house value. Alternatively, the best-performing interviewers may be assigned to these cases.

5.2. The 2007 STS Survey

The monthly variant of Short Term Statistics (STS) was conducted by Statistics Netherlands in 2007. Sampling follows a fairly standard business survey design using stratification by size class and business type with businesses selected from the Statistical Business Register. Data collection takes place via three possible modes: paper questionnaires; web questionnaires; or response through Electronic Data Reporter software. Data collected using the last option has been removed from the data considered here, since this mode was not supported after 2007. Businesses may choose to report every month or use a four-week period (thus reporting 13 times a year). For simplicity we focus on the monthly reporters in the example as the four-week period reporters require an intermediate step in which their data is distributed over monthly periods.

Data will be considered on sampled businesses in two major categories of economic activity of interest: retail (sample size = 93,799) and industry (sample size = 64,413).

Despite being a mandatory survey, nonresponse occurs, with possible reasons including lack of awareness of the mandatory nature of the survey and forgetting or refusing to respond. More importantly, response to the STS may be too slow to include in STS statistics. Estimates from the STS survey are needed 30 days after the end of the reference month, and between three and five days is needed to process, edit, impute and aggregate survey data. For the accuracy of STS statistics it is imperative to assess the impact of nonresponse after different periods of data collection, especially between 25 and 30 days of data collection.

The questions that we would like to answer with the partial R-indicators are 1) Is response sufficiently representative after 25 days?; 2) If not, what types of businesses need more attention?; and 3) Does the additional response between 25 and 30 days have a strong impact, in other words is it worth delaying data processing?

A maximum period of 90 days was employed for fieldwork in the survey. A summary of response rates after varying periods from the start of data collection is presented in Table 5.2.1, from which we can see that between 25 and 30 days the response rates go up by 6.6% and 5.6% for Retail and Industry, respectively.

In order to investigate the impact of the length of fieldwork, the R-indicators and partial R-indicators were calculated after different time periods. Auxiliary variables used to define the indicators were: business type, business size and VAT reported to Tax Office in previous year. VAT and business size relate strongly to the STS reported turnover. Since the two variables are collinear they are combined into one single variable.

Tables 5.2.2 and 5.2.3 contain estimated R-indicators for the retail and industry parts of the survey for all available auxiliary variables. The two sectors show different patterns of the R-indicators over time. While the R-indicator for Industry grows steadily over time from 0.878 to 0.931, the R-indicator for Retail is very stable. Surprisingly, the R-indicator for Retail decreases between 25 and 30 days of data collection, suggesting that the additional response accentuates the difference between respondents and nonrespondents. The maximal nonresponse bias for Industry decreases with time since both response rate and the R-indicator go up. For Retail the maximal nonresponse bias after 60 days is considerably smaller than after 15 days, but between 25 and 30 days there is hardly any change because of the drop in the R-indicator.

Table 5.2.1. Summary of response rates in Short Term Statistics business survey

Time	*Retail*	*Industry*
15 days	49.5%	48.8%
25 days	71.4%	73.1%
26 days	72.9%	74.4%
27 days	74.5%	75.8%
28 days	75.7%	76.9%
29 days	76.9%	77.9%
30 days	78.0%	78.7%
45 days	85.8%	85.7%
60 days	88.2%	88.3%

Table 5.2.2. R-indicators and maximal bias for Retail after different data collection periods

	15d	*25d*	*26d*	*27d*	*28d*	*29d*	*30d*	*45d*	*60d*
R-indicator	0.890	0.887	0.886	0.884	0.883	0.882	0.881	0.887	0.893
Max bias	0.111	0.079	0.078	0.078	0.077	0.077	0.076	0.066	0.060

Table 5.2.4 contains the variable-level unconditional and conditional partial R-indicators for type of business after different periods of time. Three observations can be made about these indicators. First, the difference between the unconditional and conditional indicators is small. Thus, the impact of business type is not removed by controlling for business size and VAT, and has an almost orthogonal impact on the representativeness of response. Second, the values of the indicators for Industry are considerably larger. Given that the R-indicators are similar in size, and hence that the variation in response propensities is also similar, this means that business type has a stronger impact on representativeness for Industry than for Retail. This impact gradually diminishes with time. After 45 days of data collection the partial indicators for Retail and Industry are comparable in size. Implicitly this also means that business size and VAT have a much stronger impact for Retail. Third, the impact of business type is stable over time for Retail. When extra response comes in, there is no change in representativeness with respect to business type.

From these observations we conclude that there is the potential to improve representativeness for Industry by speeding up response for some business types. Furthermore, we conclude that for Retail it seems to pay off to focus on VAT and business size rather than on business type. Since conditional partial R-indicators are approximately similar to unconditional partial R-indicators in all cases, the impact is "independent" of the other business characteristics.

Figure 5.2.1 presents unconditional category-level partial indicators by type of business (NACE Categories 15 to 37) for the Industry sector, given the number of days of data collection. The business type category-level indicators become smoother as data collection proceeds. After 30 days of data collection the type of business that shows the biggest negative value is NACE 29 (chemical industry). Second and third come NACE 28 (petrochemical industry) and NACE 35 (machine manufacturing industry). Between 25 and 30 days the partial R-indicators gradually become less negative. It thus pays to wait for these businesses.

For Retail the unconditional category-level partial R-indicators for VAT $\times$ business size (not shown) show hardly any change during data collection. Hence, it does not pay to wait longer than 25 days to start producing STS statistics. The two categories that stand out very clearly are new businesses with a single employee and new businesses that have between two and four employees. A new business means that no VAT was available in the

Table 5.2.3. R-indicators and maximal bias for Industry after different data collection periods

	15d	*25d*	*26d*	*27d*	*28d*	*29d*	*30d*	*45d*	*60d*
R-indicator	0.878	0.891	0.894	0.891	0.897	0.901	0.903	0.928	0.931
Max bias	0.125	0.075	0.071	0.068	0.064	0.062	0.060	0.042	0.039

JOS 2287—4/5/2011—SANKARS—391109

Table 5.2.4. Unconditional and conditional partial R-indicators at the variable level for type of business. The conditional partial R-indicators are computed with respect to VAT and business size

	Retail		*Industry*	
Days	*Unconditional*	*Conditional*	*Unconditional*	*Conditional*
15	0.017	0.016	0.047	0.043
25	0.013	0.014	0.037	0.033
26	0.013	0.013	0.035	0.031
27	0.014	0.012	0.033	0.029
28	0.014	0.012	0.032	0.028
29	0.013	0.012	0.031	0.027
30	0.013	0.012	0.029	0.025
45	0.014	0.011	0.017	0.015
60	0.013	0.011	0.015	0.013

previous year. Hence, small, starting-up businesses in retail do not respond to the STS and may be targeted in adaptive survey designs. Although individually they contribute little to the total national turnover in retail, their large number leads to a considerable impact.

6. Discussion and Future Work

In this article we have defined partial indicators for representative response, described how they may be used to monitor survey data collection, carried out a simulation study of their

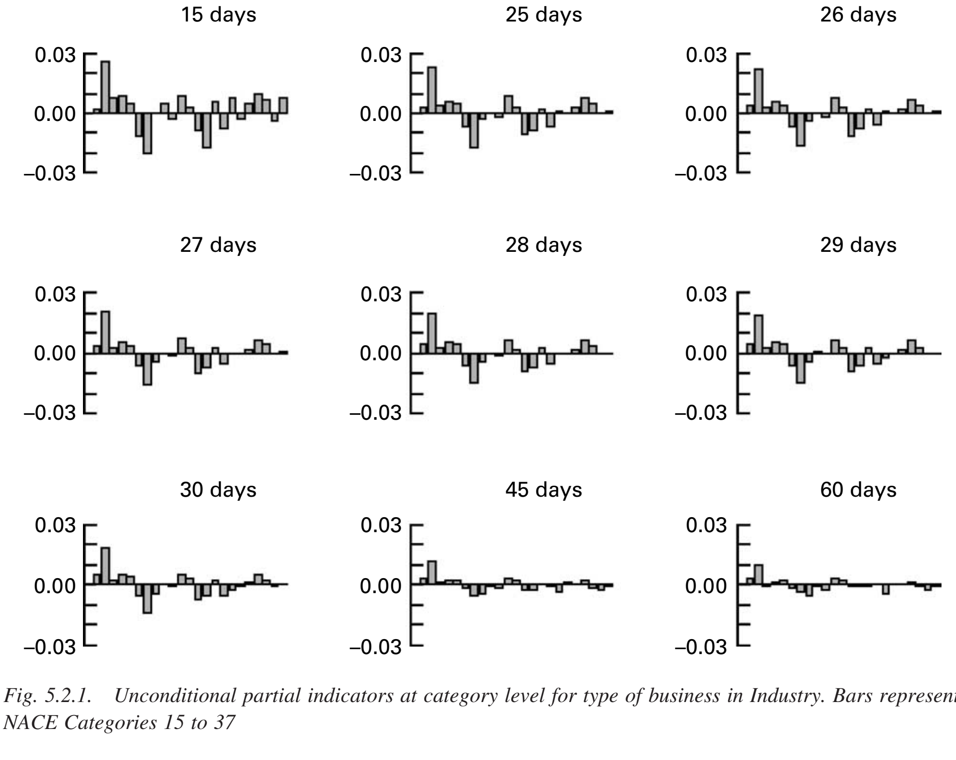

Fig. 5.2.1. Unconditional partial indicators at category level for type of business in Industry. Bars represent NACE Categories 15 to 37

sampling properties and presented two applications to show how they can provide insights into the influence of different auxiliary variables and categories of variables on lack of representativity. When used together with R-indicators and response rates, survey managers can target data collection resources to specific subgroups contributing to the lack of representativity, identify variables that might be used in survey estimation procedures to reduce non-response bias, assess future strategies for data collection modes and methods for a particular survey and compare different surveys with respect to their representativity.

The subgroups identified by R-indicators and partial R-indicators may form the input to responsive and adaptive survey designs. However, when is action required, i.e., what levels of the indicators are not acceptable, and how to set up such designs. These are important questions that ask for more experience and for benchmark studies. We see these as topics of future research.

There is one side-remark which it is important to make. Any indicator for representativeness can be artificially ameliorated by subsampling those subgroups with higher response rates. One simple way to do this for R-indicators is to subsample all subgroups using the ratio between its subgroup response rate and the lowest response rate over all subgroups. This results in constant subgroup response rates, equal to the lowest subgroup response rate identified. Hence, assessment of representativeness requires bona fide research.

This article can be viewed as a first exploration of partial indicators. We have provided basic guidelines for the use of the various indicators and for the selection of auxiliary variables. When monitoring the representativeness of a single survey, it is paramount that the selected auxiliary variables relate to the main publication domains, to key survey items or to survey-specific motives for nonresponse. Auxiliary variables may include paradata observations. From the simulation study we conclude that the estimated indicators behave broadly as expected with respect to their statistical properties. From the household and business survey applications we conclude that partial indicators can provide valuable insights to inform data collection strategies. Much is still to be learned, however, and more empirical evidence to support the fitness of the presented indicators for monitoring is key. More applications are also needed in order to assess acceptable values of indicators.

Further research into the use of partial indicators in practical settings is underway. Expressions for the linearisation standard errors of the different indicators are being developed. Two pilots were undertaken at Statistics Netherlands (Luiten and Wetzels 2010) and Statistics Norway (Kleven et al. 2010) under the RISQ project (http://www.risq-project.eu/) where R-indicators and partial indicators were used to monitor response representativeness during field work. In addition, we will employ more advanced models that distinguish different causes of nonresponse and include more fieldwork paradata.

Code in SAS and R for the computation of (partial) R-indicators can be downloaded from the RISQ website as well as a manual and test data set.

7. References

Andridge, R. and Little, R. (2010). Proxy Pattern-mixture Analysis for Survey Nonresponse. Paper presented at 21st International workshop on Household Survey Nonresponse, August 30–September 2, Nürnberg, Germany.

Biemer, P.P. and Lyberg, L.E. (2003). Introduction to Survey Quality. Hoboken: Wiley.

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

Groves, R.M. and Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias. Public Opinion Quarterly, 72, 1–23.

Groves, R.M., Brick, J., Couper, M., Kalsbeek, W., Harris-Kojetin, B., Kreuter, F., Pennell, B., Raghunathan, T., Schouten, B., Smith, T., Tourangeau, R., Bowers, A., Jans, M., Kennedy, C., Levenstein, R., Olson, K., Peytcheva, E., Ziniel, S., and Wagner, J. (2008). Issues Facing the Field: Alternative Practical Measures of Representativeness of Survey Respondent Pools. Survey Practice, October. Available at http://surveypractice.org/2008/10/30/issues-facing-the-field/

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.

Kleven, O., Fosen, J., Lagerstrøm, B., and Zhang, L.C. (2010) The Use of R-indicators in Responsive Survey Design – Some Norwegian Experiences. Paper presented at Q2010 Conference, Helsinki. Available at http://www.risq-project.eu/papers/

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

Loosveldt, G. and Beullens, K. (2009). Fieldwork Monitoring, Deliverable 5, Version 2 of Work Package 6, RISQ project. Available at http://www.risq-project.eu/papers/RISQ-Deliverable-5-V2.pdf

Luiten, A. and Wetzels, W. (2010). Differential Survey Strategies based on R-indicators. Paper presented at Q2010, Helsinki. Available at http://www.risq-project.eu/papers/

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.

Särndal, C.-E. and Lundström, S. (2008). Assessing Auxiliary Vectors for Control of Nonresponse Bias in the Calibration Estimator. Journal of Official Statistics, 24, 167–191.

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

Shlomo, N., Skinner, C.J., Schouten, B., Bethlehem, J., and Zhang, L.C. (2008). Statistical Properties of R-indicators, Work Package 3, Deliverable 2.1, RISQ Project http://www.risq-project.eu

Wagner, J. (2008). Adaptive Survey Design to Reduce Nonresponse Bias. PhD Thesis, University of Michigan, U.S.A.

Received May 2010
Revised March 2011

JOS 2287—4/5/2011—SANKARS—391109

Author Queries

JOB NUMBER: 2287

JOURNAL: JOS

Q1 We have inserted a running head. Please approve or provide an alternative.

Q2 We have inserted citations for Tables (2.2 to 2.7). Please approve or provide alternatives.