Does nonresponse in business surveys matter?

Dan Hedlin, Department of Statistics, Stockholm University

Abstract

Nonresponse rates are rising in social surveys in many countries, but not necessarily in business surveys. Bethlehem (1988) and Särndal and Lundström (2005) have provided expressions for nonresponse bias. I shall discuss their interpretation in the context of business surveys. I focus on stratified simple random sampling. One underlying question that I'll attempt to answer is this: can we live with nonresponse levels larger than say 50%? This is not only an issue of nonresponse bias; official statistics is a matter of trust and high nonresponse rates are often viewed with grave misgivings. So if we believe that the nonresponse bias is low in one survey despite a high nonresponse rate, how can we communicate the (lack of) uncertainty?

Introduction

Social surveys, as opposed to business surveys, have been in the forefront of the debate and research about nonresponse issues. Rising nonresponse levels in important social surveys in many countries in Europe, Australia and Northern America (de Leeuw and de Heer 2002) have caused concern also outside the community of statisticians. For example, Örstadius (2015) wrote in the respected Swedish daily Dagens Nyheter that the trend with rising nonresponse levels distorts official statistics, an article that spurred a debate on nonresponse and public trust. In England, when the PISA survey fell short of the desired response rate in 2003, the reports from the OECD excluded the UK which led to a debate on trust in official statistics (Micklewright et al. 2012).

To go back in history, Cochran (1951, p. 652) stated that 'unfortunately, any sizeable percentage of nonresponse makes the results open to question by anyone who cares to do so'. His argument was that we know little about the nonresponding part of the sample and hence a wide spread of estimates are feasible within the realm of the knowledge gained from the response set. Särndal, Swensson and Wretman (1992, p. 559) claimed that 'the greater the nonresponse, the more one has reason to worry about its harmful effects on the survey estimates', a statement that – in the view of the significant development of methods of estimation in the presence of nonresponse that we have seen since 1992 – has been qualified, not least by Carl-Erik Särndal himself. Other quotations that indicate the thinking of the time before turn of the millennium include: the "'Best Practices' guide of the American Association for Public Opinion Research (AAPOR 1997, p. 5) states that 'a low cooperation or response rate does more damage in rendering a survey's results questionable than a small sample'" (quotation from Curtin et al. 2000, pp. 213-214).

Attractive features of the nonresponse rate were in the past (Groves et al. 2008):

1. 'its simplicity (a single number to characterize quality), and
2. its transparency (it was not a complicated statistical function of sample observations)'.

Laypeople perceive often a response rate as a clear indicator of quality or lack of quality. However, many survey methodologists have since re-evaluated the importance of high response rates (Moore et al., 2016, Groves, 2006), although in social surveys there may still be more focus on the nonresponse rate than nonresponse bias (Davern 2013). One rather neglected aspect of the information content of the nonresponse rate is its inaccuracy; one has to estimate the extent of overcoverage, which in the presence of nonresponse often requires some nontestable assumptions (Skalland 2011).

Theoretical work indicates that the nonresponse rate does play a direct and vital role, see for example Bethlehem and Bakker (2014). Still, a number of recent empirical studies focusing on social surveys suggest that the nonresponse rate is a poor predictor of nonresponse bias. Groves (2006) and Groves and Peytcheva (2008) are often referenced. However, we should take care not to conclude from these studies that nonresponse rate and nonresponse bias are not often associated. As the thinking used to be that a low response rate – meaning lower than about 0.75 only some decades ago – is deleterious, it is easy to imagine publication bias. If everyone agrees that a response level lower than 0.75 makes a survey untrustworthy, why would you publish results that only support this view?

What studies such as Merkle and Edelman (2002, 2009) and many other studies bring out is the importance of *the composition* of the response set, rather than its relative size. It may be worth stating the obvious fact that if we towards the end of the data collection manage to obtain responses from those who actually make the difference between respondents and nonrespondents smaller, then the bias will be smaller. However, efforts to increase the size of the response set may make the nonresponse bias worse if the composition is adversely affected. Several indicators that measure the 'goodness' of composition have been proposed, see Nishimura et al. (2016) and Tourangeau et al. (2017) for overviews, and also Lundquist and Särndal (2013) and Särndal et al. (2016) for a balance measure.

Whether the nonresponse rates have been rising in *business surveys* is not clear. Furthermore, there has been less debate about the importance of the nonresponse rate in business surveys than in social surveys. In business surveys it is not uncommon to focus on weighted nonresponse rates rather than the raw ones, and the weighted rates are in business surveys considerably smaller than the raw rates because after the end of data collection there is less nonresponse among large and medium-sized businesses than small ones.

Link between nonresponse rate and bias

We assume that the aim is to estimate the population total $t_{y:U} = \sum_U y_k$ or the population mean $\bar{Y}$ of a study variable $\mathbf{y} = (y_1, y_2, \ldots, y_N)'$ on a population *U* with unit labels $\{1,2, \ldots, N\}$. We assume further that there is an auxiliary variable $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)'$, with $\mathbf{x}_k$ known for each element in *U*. A sample *s* of size *n* is taken and y_k is observed for all units *k* in a response set *r*. The inclusion probability of a unit *k* will be denoted by π_k. The inference framework is design-based.

I now introduce three expressions briefly. A well-known expression for nonresponse bias found in many textbooks, for example Cochran (1977, p. 361) and Biemer and Lyberg (2003, p. 83) is

$$E(\bar{y}_r) - \bar{Y} = \frac{N_{nr}}{N}(\bar{Y}_r - \bar{Y}_{nr}), \quad (1)$$

where $\bar{y}_r$ is the average of the study variable among the respondents, $\bar{Y}$ is the population mean, $\bar{Y}_r$ is the population mean of those who respond with probability one, and $\bar{Y}_{nr}$ is the population mean of those who have zero probability to respond, and N_{nr}/N is the population proportion of the nonrespondents. Expression (1) assumes deterministic nonresponse and uses the expansion estimator, which does not take advantage of auxiliary information. Thus, (1) is very restrictive and, frankly, not very interesting.

Another well-known expression is due to Bethlehem (1988). A population unit is assumed to have a propensity (probability) θ_k to respond to a particular survey item at a particular point in time, using the survey protocol. Then the bias is

$$E_{pq}(\bar{y}_r) - \bar{Y} \approx Cov(y,\theta)/\bar{\theta}_U, \quad (2)$$

where $Cov(y,\theta)$ is the finite population covariance of $\mathbf{y}$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_N)'$, $\bar{\theta}_U$ is the population mean of the propensities and the expectation is taken over the sampling design $p(s)$, which is the probability that sample s is drawn, and the conditional response probability $q(r|s)$, which is the probability that the response set is r. A shift in response propensities will affect the denominator in (2) but not the location invariant numerator. Thus, as with expression (1), a higher average propensity of responding will lead to smaller bias. The ratio of the biases from two surveys, the first of which with response propensities $\boldsymbol{\theta}$, the other one with $\boldsymbol{\theta} + \kappa$, where κ is a constant, everything else equal, is $1 + \kappa/\bar{\theta}_U$.

The main approximation in (2) arises from a first-order Taylor series approximation, which is ubiquitous in the survey sampling literature.

The expression (2) emanates from Bethlehem's (1988) estimator (3.2), $\hat{t}_{JB:r} = N \sum_r w_k y_k / \sum_r w_k$, which estimates the population total. Here $w_k = \pi_k^{-1}$. This estimator recovers the Hájek estimator for full response.

The bias will be denoted by $Bias_{pq}(\hat{t}) = E_{pq}(\hat{t}) - t$.

We see in (2) that for constant response propensities the covariance vanishes and the bias is zero no matter the average response propensity, and hence the nonresponse rate.

The relative bias is

$$Bias_{pq}\left(\hat{t}_{JB:r}\right)/t_y \approx \rho(y,\theta) cv_y \sigma_\theta / \bar{\theta}_U, \quad (3)$$

where $\rho(y,\theta)$ is the finite population Pearson correlation coefficient of $\mathbf{y}$ and $\boldsymbol{\theta}$, σ_θ is the standard deviation of $\boldsymbol{\theta}$ and cv_y is the coefficient of variation of $\mathbf{y}$ (see Bethlehem, 1988).

In keeping with the more recent literature, I will change focus from estimation of the mean to the total. An interesting third expression (Lundström and Särndal, 1999, and Särndal and Lundström, 2005, p. 100) is

$$E_{pq}(\hat{t}_{cal:r}) - t_y \approx -\sum_U e_{\theta k}, \quad (4)$$

where $e_{\theta k} = y_k - \mathbf{x}_k' \mathbf{B}_{U;\theta}$ (5)

and $\mathbf{B}_{U;\theta} = (\sum_U \theta_k \mathbf{x}_k \mathbf{x}_k')^{-1} \sum_U \theta_k \mathbf{x}_k y_k$. (6)

The estimator can be any calibrated estimator. The quantities $e_{\theta k}$ are finite population model residuals. The stunning differences between (1) and (2), on the one hand, and (4) on the other hand are that (4) does not include the nonresponse *rate* in any simple way, although the response propensities do appear in $\mathbf{B}_{U;\theta}$, but on the other hand, (4) does include the auxiliary variables whereas (1) and (2) do not.

You can show (a forthcoming paper of mine) that (4) and (2) are equivalent for stratified simple random sampling and poststratification, where (2) is computed within strata or poststrata and summed up to population level.

Särndal, Swensson and Wretman (1992, p. 55) write that $0.5 \leq cv_y \leq 1$ for 'many variables and many populations'. The cv_y for income from salary was 1.7 in 2005 for all residents of Sweden 17 years of age or older (Statistics Sweden, 2008, p. 114). Data about the level and spread of propensities and correlations with study variables are scarcer in the literature. Kreuter et al. (2010) report on correlation between response and auxiliary variables in five large social surveys and find the most correlations are smaller than 0.10 in absolute terms.

However, for variables in business populations the cv_y can be considerably higher. The quotation from Särndal et al. (1992) seems implicitly confined to populations of individuals.

In Figure 1, I still focus on social surveys in my choice of numbers in where cv_y is fixed to unity. Figure 1 depicts $Bias_{pq}\left(\hat{t}_{JB:r}\right)/t_y$ for various values of $\rho(y,\theta)$ with $\sigma_\theta = 0.29$, which is the standard deviation of a uniform distribution. For example, for $\rho(y,\theta) = 0.025$ the relative bias is smaller than 2% when $\bar{\theta}_U \geq 0.37$. Three important conclusions can be drawn from Figure 1. First, for smaller values of $\bar{\theta}_U$ than about 0.25 the relative bias will be high even for fairly small values of $\rho(y,\theta)$. Second, for $\rho(y,\theta)$ smaller than about 0.10 the relative bias is rather flat for all values of $\bar{\theta}_U$ greater than about 0.30. Third, for large correlations, say $|\rho(y,\theta)| > 0.15$, the relative bias will be at least fairly large no matter the response rate. So there seems to be a 'safe area' in social surveys which is enclosed in roughly $\bar{\theta}_U > 0.30$ and $|\rho(y,\theta)| < 0.10$, assuming that $cv_y = 1$.

In business surveys we have to expect higher values than $cv_y = 1$, even within strata. This is particularly true for the completely enumerated stratum in a business survey with stratified simple random sampling. In Figure 2, $cv_y = 2$. Now the 'safe area' requires roughly $|\rho(y,\theta)| < 0.05$. So, perhaps surprisingly, business surveys seem more sensitive to nonresponse than social surveys.

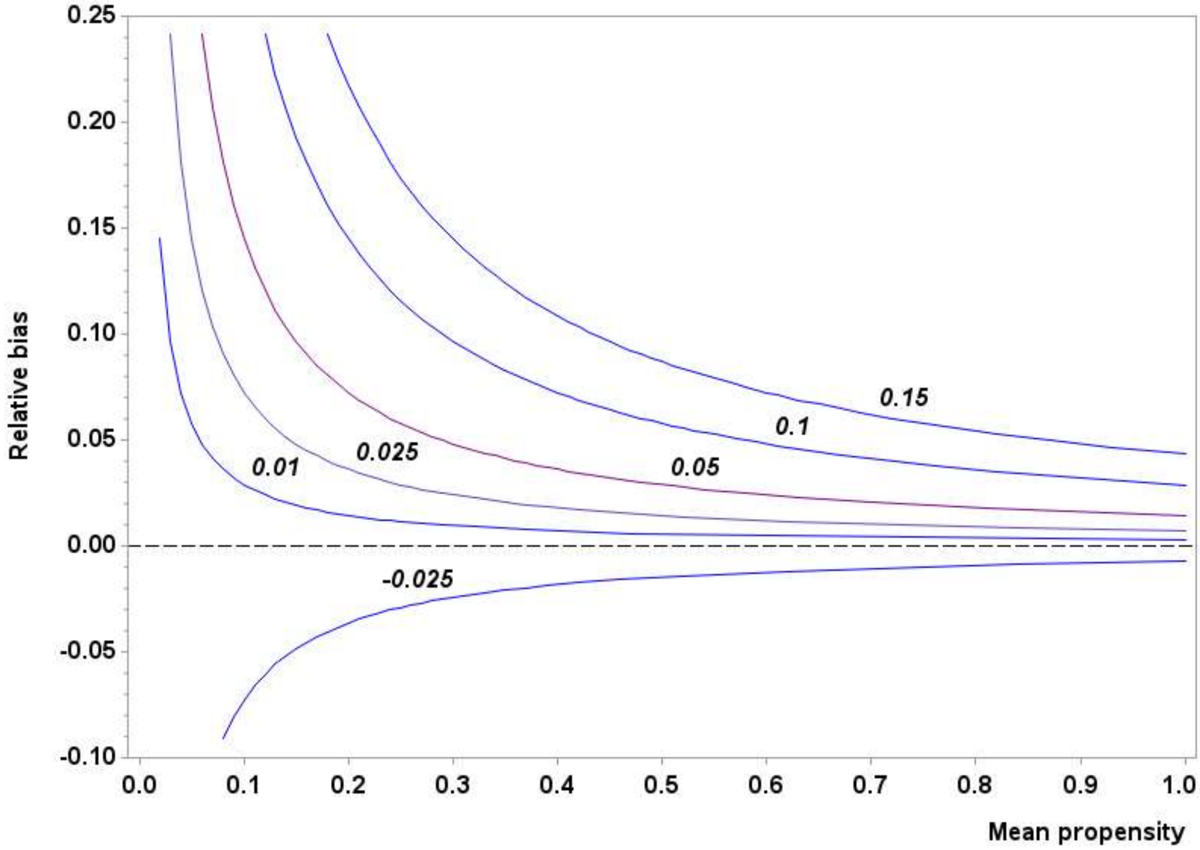

Figure 1. Size of relative bias, that is, the left-hand-side of (3 Error! Reference source not found.**), against mean propensity $\overline{\theta}_U$ for various values of $\rho(y, \theta)$, which are in italics. For all curves, $cv_y = 1$ and $\sigma_\theta = 0.29$.**

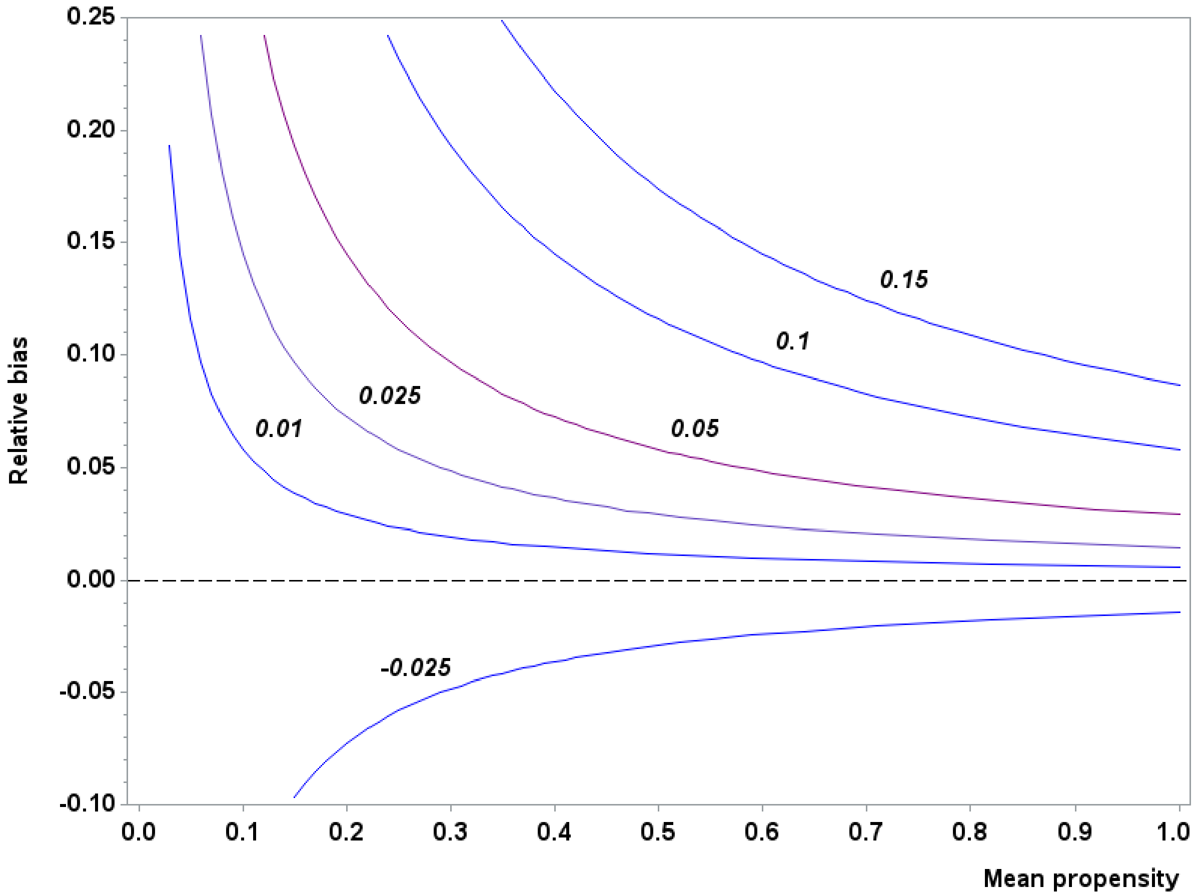

Figure 2. Size of relative bias, that is, the left-hand-side of (3), against mean propensity $\bar{\theta}_U$ for various values of $\rho(y,\theta)$, which are in italics. For all curves, $cv_y = 2$ and $\sigma_\theta = 0.29$.

Is there a bias level that is acceptable?

What can we say about the maximum *acceptable* bias? One view is to accept bias that does not distort the coverage probability of confidence intervals badly. A bias ratio, $Bias(\hat{t})/\left(V(\hat{t})\right)^{-0.5}$ less than 0.30 is in practice negligible for the coverage of a 95% confidence interval, see Särndal et al. (1992, pp. 164-165). In practice the coverage probability often falls short of 95% due to underestimation of the customary Taylor series variance of the generalised regression estimator. See Wu and Deng (1983) for coverage probabilities of the ratio estimator, and Hedlin (2002) for various estimators in two business surveys. In the light of the actually realised coverage probability, we may even accept a bias ratio equal to 0.5. Let $F = V\left(\hat{t}_{JB:r}\right)/V(\hat{t}_{HT:s})$ be the 'estimator and nonresponse effect', where $\hat{t}_{HT:s}$ is the HT estimator based on full response. Then, ignoring the finite population correction (admittedly very debatable in a business survey), the bias ratio is

$$\frac{Bias_{pq}\left(\hat{t}_{JB:r}\right)}{\sqrt{V(\hat{t}_{HT:s})\cdot F}} \approx \rho(y,\theta)cv_y cv_\theta t_y \left(V(\hat{t}_{HT:s})\cdot F\right)^{-0.5} = \rho(y,\theta)cv_\theta\sqrt{n/F} \approx \rho(y,\theta)cv_\theta\sqrt{n_r}\,, \qquad (7)$$

where in the last approximation we have assumed that $F \approx n/n_r$. If we accept a bias ratio of at most size b, then the mean response propensity should be

$$\bar{\theta}_U \geq \rho(y,\theta)\sigma_\theta b^{-1}\sqrt{n_r} \quad (8)$$

For $\sigma_\theta = 0.29$, $\rho(y,\theta) = 0.05$, $b = 0.50$ and $n_r = 100$ in a domain, $\bar{\theta}_U \geq 0.29$, which is a very modest response rate. Note that cv_y has cancelled out. For these numbers, the relative bias is 5%. So to go back to the question in the abstract: can we live with nonresponse levels larger than say 50%? The tentative answer is yes, provided that the $\rho(y,\theta)$ is small. However, whether it tends to be small in business surveys in general is not clear.

The maximum value of (3) is

$$\max\left|\frac{Bias_{pq}(\hat{t}_{JB})}{t_y}\right| \approx cv_y\sqrt{\frac{1-\bar{\theta}_U}{\bar{\theta}_U}} \quad (9)$$

(Bethlehem, 2016). To arrive at (9) it is assumed that $|\rho(y,\theta)| = 1$. While the conservatism makes (9) of little practical interest, we see that an upper bound of the absolute bias is driven by the coefficient of variation of $\mathbf{y}$ and the square root of the odds of the ‘average individual’ not responding. For $\bar{\theta}_U = 0.5$ the maximum relative bias equals the cv_y.

Time to replace the reporting of nonresponse rates with something else?

The highly influential textbook Cochran (1977, pp. 361-363 with the same text in the 1953 edition) paints a very bleak picture of the negative effects of even modest nonresponse rates. My guess is that this textbook and other papers by Cochran have been a major factor behind the view on nonresponse rates that I indicated in the beginning of this paper. For example, Hansen, Hurwitz and Madow (1953) do not convey as despondent a message as Cochran. Unfortunately, Cochran based his reasoning on the same deterministic response model as (1), which may have been the best researchers could do at the time, but, as we know today, it is not very useful.

I do not believe we can or even should stop reporting response rates in surveys. The response rate does convey some important information. The issue with the nonresponse rate is the difficulty in the interpretation of that information. Being difficult to interpret, however, is not a good reason to stop reporting it altogether. But there are modern measures that would be useful additions to response rates. Groves et al. (2008) discuss in a Survey Practice paper measures that could augment or even replace response rates. They distinguish between measures at the survey level and at the estimate level. I believe that measures at survey level are needed. Some of them that Groves et al. (2008) discuss measure ‘representativeness’, such as the R indicator. The indicators of the ‘goodness’ (representativeness) of the composition of the response set, which I mentioned at the beginning of this paper are interesting candidates, including the R indicator.

References

Bethlehem, J.G. (1988). Reduction of nonresponse bias through regression estimation. Journal of Official Statistics, 4, 51-60.

Bethlehem, J. (2016). Solving the nonresponse problem with sample matching? Social Science Computer Review, 34, 59-77.

Bethlehem, J. and Bakker, B. (2014). The impact of nonresponse on survey quality. Statistical Journal of the IAOS, 30, 243-248.

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

Cochran, W.G. (1951). General principles in the selection of a sample. American Journal of Public Health, 41(6), 647-653.

Cochran, W.G. (1953). Sampling techniques. New York: Wiley.

Cochran, W.G. (1977). Sampling techniques, 3rd ed. New York: Wiley.

Curtin, R., Presser, S. and Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment. Public Opinion Quarterly, 64, 413–428.

Davern, M. (2013). Nonresponse rates are a problematic indicator of nonresponse bias in survey research. Editorial. Health Services Research, 48(3), 905–912.

De Leeuw, E. and de Heer, W. (2002). Trends in household survey nonresponse: a longitudinal and international comparison. In Survey Nonresponse, eds R.M. Groves, D.A. Dillman, J.L. Eltinge and R.J.A. Little. New York: Wiley, 41-54.

Groves, R.M. (2006). Nonresponse rates and nonresponse bias in household surveys. Public Opinion Quarterly, 70, 646-675.

Groves, R.M. et al. (2008). Issues facing the field: alternative practical measures of representativeness of survey respondent pools. Survey Practice, 3(1), online http://surveypractice.org/index.php/SurveyPractice/article/view/221/html

Groves, R.M. and Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: a meta-analysis. Public Opinion Quarterly, 72, 167-189.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953). Sample survey methods and theory. Volumes 1 and 2. New York: Wiley.

Hedlin, D. (2002). Estimating totals in some UK business surveys. Statistics in Transition, 5, 943-968.

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.

Lundquist, P. and Särndal, C-E. (2013). Aspects of responsive design with applications to the Swedish Living conditions survey. Journal of Official Statistics, 29, 557-582.

Lundström, S. and Särndal, C-E. (2005). Calibration as a standard method for treatment of nonresponse. Journal of Official Statistics, 15, 305-327.

Merkle, D.M. and Edelman, M. (2002). Nonresponse in exit polls: a comprehensive analysis. In Survey Nonresponse, Eds R.M. Groves, D.A. Dillman, J.L. Eltinge and R.J.A. Little. New York: Wiley, 243-257.

Merkle, D.M. and Edelman, M. (2009). An experiment on improving response rates and its unintended impact on survey error. Survey Practice, 2(3), 1-5.

Micklewright, J., Schnepf, S.V. and Skinner, C. (2012). Non-response in surveys of schoolchildren: the case of the English Programme for International Student Assessment (PISA) samples. Journal of the Royal Statistical Society, series A, 175(4), 915-938.

Moore, J.C., Durrant, G.B. and Smith, P.W.F. (2016). Data set representativeness during data collection in three UK social surveys: generalizability and the effects of auxiliary covariate choice. Journal of the Royal Statistical Society, series A, Online DOI 10.1111/rssa.12256.

Nishimura, R., Wagner, J. and Elliot, M. (2016). Alternative indicators for the risk of non-response bias: A simulation study. International Statistical Review, 84 (1), 43–62.

Örstadius, K. (2015). The official statistics of Sweden may be misleading. Dagens Nyheter. (Swedish title: Sveriges officiella statistik hotar att bli missvisande). www.dn.se, published online 2015-01-18 (in Swedish).

Särndal, C-E., Lumiste, K. and Traat, I. (2016). Reducing the response imbalance: Is the accuracy of the survey estimates improved? Survey Methodology, 42(2), 219-238.

Särndal, C-E. and Lundström, S. (2005). Estimation in surveys with nonresponse. New York: Wiley.

Särndal, C.-E., Swensson, B. and Wretman J. (1992). Model assisted survey sampling. New York: Springer-Verlag.

Skalland, B. (2011). An alternative to the response rate for measuring a survey's realization of the target population. Public Opinion Quarterly, 75(1), 89–98.

Statistics Sweden (2008). Sample design – from theory to practice. Handbook 2008:1. Örebro: Statistiska centralbyrån (in Swedish; Swedish title 'Urval – från teori till praktik').

Tourangeau, R., Brick, J.M., Lohr, S. and Li, J. (2017). Adaptive and responsive survey designs: a review and assessment. Journal of the Royal Statistical Society, Series A, 180, 203-223.

Wu C.F. and Deng L.Y. (1983). Estimation of variance of the ratio estimator: an empirical study. in scientific inference, data analysis, and robustness, eds G.E.P. Box, T. Leonard, C.F. Wu. New York: Academic Press, 245-277.