Is there a 'safe area' where the nonresponse rate has only a modest effect on nonresponse bias despite non-ignorable nonresponse?

Dan Hedlin

Is there a 'safe area' where the nonresponse rate has only a modest effect on nonresponse bias despite non-ignorable nonresponse?

Dan Hedlin, Department of Statistics, Stockholm University

Abstract

Rising nonresponse rates in social surveys makes the issue of nonresponse bias highly contentious. Nonresponse may induce bias and increase data collection costs. We study the relationship between response rate and bias, assuming non-ignorable nonresponse and focusing on estimates of totals or means.

We show that there is a 'safe area' enclosed by the response rate on the one hand and the correlation between the response propensity and the study variable on the other hand; in this area,
1) the response rate does not greatly affect the nonresponse bias
and
2) the nonresponse bias is small.

1. Introduction

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 the UK, 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).

In parallel, and mainly within the statistical community, many survey methodologists have re-evaluated the importance of high response rates, for example Kreuter (2013), Davern (2013) and Moore et al. (2016). See also other references in Brick (2013). Groves (2006, p. 670) put it succinctly: 'there is little empirical support for the notion that low response rate surveys de facto produce estimates with high nonresponse bias'. However, Bethlehem (2009, p. 212) writes: 'Due to the negative impact nonresponse may have on the quality of survey results, the response rate is considered to be an important indicator of survey quality'.

Another development is the broad emergence of surveys with response rates lower than 0.15. They arise because the sample selection in some current surveys is done in numerous steps. First, a database of email addresses of individuals are collected (either with random or nonrandom selection). Then, persons with the email addresses take part in a profile survey, and only after responding to the profile survey the owner of the email address may be invited to surveys that ask about study variables. The cumulation of nonresponse in the steps of the selection process makes it hard to reach response rates larger than roughly 0.15. Very low response rates can be observed also in some other types of surveys outside the area of official statistics.

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). These statements are not necessarily incorrect.

Theoretical work indicates that the nonresponse rate does play a direct and vital role for the nonresponse bias, see for example Bethlehem (2009, Ch. 9). Still, a number of recent empirical studies suggest rather the opposite. Groves (2006) and Groves and Peytcheva (2008) are often referenced. They compile nonresponse bias estimates from a large number of studies and conclude that the nonresponse rate is a poor predictor of nonresponse bias. Brick and Tourangeau (2017, p. 738) re-analysed the data of Groves and Peytcheva (2008) and concluded that 'response rates may not be very good predictors of nonresponse bias, but they are far from irrelevant' and that unit response rates do provide useful indicators of nonresponse bias.

The purpose of this paper is to clarify in what situations the nonresponse rate affects the nonresponse bias. However, we do not attempt to provide a tool for the survey practitioner with which she or he can assess the nonresponse bias of a survey. The bias is the result of the interplay between correlation between the response propensity and the study variable, which unknown due to nonresponse, and the response rate.

First we review some theoretical expressions for nonresponse bias. We focus on one from Bethlehem (1988). For our purposes Bethlehem's expression suffices, although Särndal and Lundström (2005) have presented a more general expression. Brick (2013) provides an excellent review of the literature on nonresponse models end bias expressions. In Section 3, we discuss Bethlehem's bias expression in practical terms. In Section 4 we attempt to transform the issue of what size of bias is acceptable into a line of thinking that can be used in practice.

2. Expressions for nonresponse 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 $r \subset s$. The inclusion probability of a unit k will be denoted by π_k. The inference framework is design-based.

We now recapitulate two expressions for nonresponse bias. 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

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

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.

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. I believe that this textbook and other papers by Cochran have been a major factor behind the view that high nonresponse rates are crucial. 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.

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

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

where $Cov(y, \theta)$ is the finite population covariance of $\mathbf{y}$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_N)'$, $\bar{\theta}_U$ 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.

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

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 $1 + \kappa/\bar{\theta}_U$. For example, if a survey organisation manages with a successful change of protocol to increase the mean response rate from 0.5 to 0.6, everything else equal, the reduction of bias is 17%. However, in practice this reduction may not be important, if the bias was small in the first place.

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

The expression (2) emanates from Bethlehem (1988; estimator (3.2)), $\hat{t}_{JB:r} = N \sum_r w_k y_k / \sum_r w_k$, which estimates the population total. The notation Σ_r stands for summation over the response set r and $w_k = \pi_k^{-1}$.

The bias of an estimator of a population total will be denoted by $Bias_{pq}(\hat{t}) = E_{pq}(\hat{t}) - t$. The relative bias is

$$Bias_{pq}(\hat{t}_{JB:r})/t_{y:U} \approx \rho(y,\theta)cv_y cv_\theta = \rho(y,\theta)cv_y \sigma_\theta/\bar{\theta}_U, \tag{3}$$

where $\rho(y,\theta)$ is the finite population Pearson correlation coefficient of $\mathbf{y}$ and $\boldsymbol{\theta}$, σ_θ is the population standard deviation of $\boldsymbol{\theta}$ and cv_y is the population coefficient of variation of $\mathbf{y}$ (see Bethlehem, 1988). If $\hat{t}_{JB:r}$ is replaced with the poststratification estimator, (2) becomes (as noted by Bethlehem, 1988)

$$Bias_{pq}(\hat{t}_{post:r}) = \sum_{h=1}^{H} N_h \frac{\sum_{U_h} \theta_k y_k}{t_{\theta:U_h}} - t_{y:U}, \tag{4}$$

where N_h is the number of units in poststratum h and $t_{\theta:U_h} = \sum_{U_h} \theta_k$, the sum taken over units in poststratum h. Note that poststratum refers here to a subset of the population, not a subset of the sample.

For a binary study variable,

$$y_k = \begin{cases} 0 \\ 1 \end{cases}$$

and an estimate of the population proportion rather than the total, you can show with a little algebra that (4) is

$$Bias_{pq}(\hat{p}_{post:r}) = -\sum_{h=1}^{H} P_h (1 - \bar{\theta}_{1h}/\bar{\theta}_{U_h}) \tag{5}$$

where P_h is the proportion of 'ones' in poststratum h, $\bar{\theta}_{1h} = \sum_{U_h} \theta_k y_k / M_h$, M_h being the number 'ones' and $\bar{\theta}_{1h}$ the average propensity of those who have $y_k = 1$ in poststratum h. For a given P_h the main driver of bias is the difference between the average propensity of everybody in poststratum h, and that of those who possess the characteristic of interest (i.e. having value $y_k = 1$).

One may at this stage ask how (2) and following formulae are connected to the concept of MAR. A missing at random (MAR) nonresponse mechanism is defined in Little and Rubin (2002, p. 12) in terms of the likelihood of missingness as

$$f(\mathbf{M}|\mathbf{Y},\phi) = f(\mathbf{M}|\mathbf{Y}_{obs},\phi) \text{ for all } \mathbf{Y}_{mis}, \phi,$$

where $\mathbf{Y}$ is an n x p dataset, consisting of the observed data $\mathbf{Y}_{obs}$ and the missing data $\mathbf{Y}_{mis}$, including auxiliary variables, and $\mathbf{M}$ is the n x p matrix of missingness indicators; ϕ are the unknown parameters. A design-based version of the MAR condition is

$$q(r|s, \boldsymbol{x}_s, \mathbf{y}_s) = q(r|s, \boldsymbol{x}_s, \mathbf{y}_{obs}), \tag{6}$$

where $\boldsymbol{x}_s$ is the auxiliary variables in the sample. The condition (6) is essentially the same as the condition for the response mechanism to be 'unconfounded', defined in Lee et al. (1994). In this paper, we shall say that the response mechanism is ignorable if (6) is satisfied.

If $\rho(y, \theta) = 0$ in (3), then the response propensities are constant within poststrata in (4) and the bias vanishes. If $\rho(y, \theta) = 0$ the response mechanism is ignorable. Also, if $\bar{\theta}_{1h} = \bar{\theta}_{U_h}$ for all h in (5), the response mechanism is ignorable. Note that ignorability hinges on the availability and choice of auxiliary variables in $\boldsymbol{x}_s$, which, for example, in poststratification is used to define poststrata. If one obtains auxiliary variables for poststratification that makes the response propensities constant within poststrata, one can say that 'the auxiliary variables completely explains the response mechanism'. Then there will be no nonresponse bias.

We adopt the view of Särndal and Lundström (2005, p. 105) when they say that "to hope that this $\boldsymbol{x}_k$ will achieve a 'complete explanation' of θ_k is utopian", that is, for the response mechanism to be completely ignorable. In other words, in this paper the missing data mechanism is always viewed as non-ignorable, a far more realistic assumption than MAR.

3. What do the expressions for nonresponse bias tell us?

Now we shall look at some numerical examples of relative bias under non-ignorable missing data mechanisms. Särndal et al. (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 response propensities and correlations with study variables are scarcer in the literature (Brick 2013). 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. The standard deviation of the response propensity should in most cases be at most $\sigma_\theta = 0.29$, which is the standard deviation of a uniform distribution. It should at least not be much larger than 0.29.

Figure 1 depicts $Bias_{pq}(\hat{t}_{JB:r})/t_y$ for $cv_y = 1$, $\sigma_\theta = 0.29$ and varying values of $\rho(y, \theta)$. For example, for $\rho(y, \theta) = 0.025$ the relative bias is smaller than 2.5% when $\bar{\theta}_U \geq 0.3$. The open rectangle encloses mean population propensities larger than 0.3, and relative biases smaller than 0.05 in absolute terms. Four important conclusions can be drawn from Figure 1. First, for smaller values of $\bar{\theta}_U$ than about 0.2 the relative bias will be high even for small values of $\rho(y, \theta)$. Second, for $\rho(y, \theta)$ smaller than about 0.05 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. Fourth, there seems to be a 'safe area' which is enclosed in roughly $\bar{\theta}_U >$

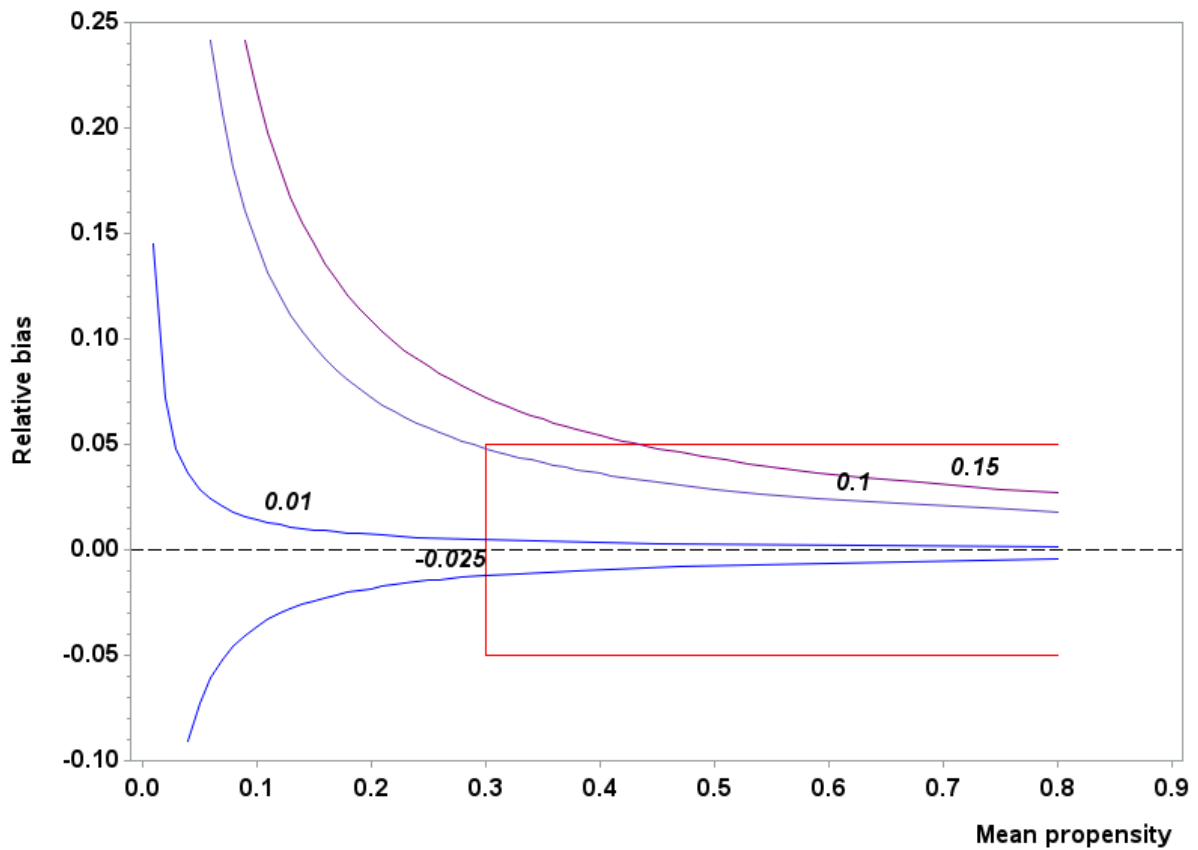0.30 and $|\rho(y, \theta)| < 0.05$, assuming that $cv_y = 1$.
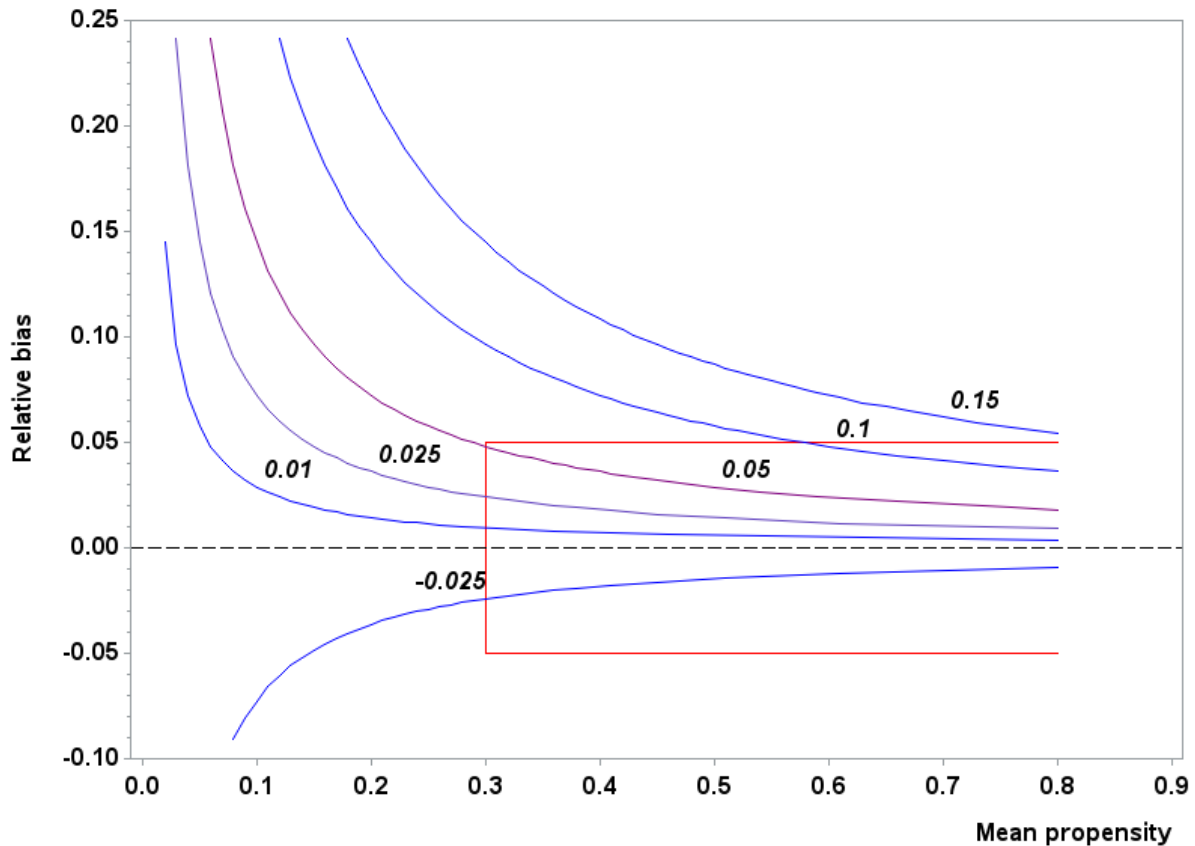
Figure 1. Size of relative bias, that is, the left-hand-side of (3), against mean propensity $\bar{\theta}_U$ for various values of $\rho(y, \theta)$, which are in italics. For all curves, $cv_y = 1$ and $\sigma_\theta = 0.29$.

Expressed in terms of non-ignorability and response rates, even for mildly non-ignorable missing data mechanisms, a response rate lower than about 0.2 will produce considerable bias. However, if the response rate is higher than about 0.3, mildly non-ignorable missing data mechanisms produce small biases, and the bias is in practice nearly independent of the level of the nonresponse rate (provided that it is larger than about 0.3). Note that if the missing data mechanism were MAR, the curves would coincide with the horizontal line $y = 0$, which would be the case if $\rho(y, \theta) = 0$.

Figure 2 is similar to Figure 1. In Figure 2, $cv_y = 0.5$, other than that it is essentially the same as Figure 1. Here the 'safe area' seems enclosed in roughly $\bar{\theta}_U > 0.30$ and $|\rho(y, \theta)| < 0.10$.

It should be noted that for smaller values of σ_θ than 0.29, the curves would be flatter and closer to zero bias for $\bar{\theta}_U > 0.30$, and have sharper bends and steeper ascents for $\bar{\theta}_U < 0.20$. In this sense,

6

Figures 1 and 2 are conservative.

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, $cv_y = 0.5$ and $\sigma_\theta = 0.29$.

Figure 3 shows bias in (5) for one poststratum with $P_h = 0.5$. For example, the dot in the top left corner represents $\bar{\theta}_{U_h} = 0.95$ and $\bar{\theta}_{1h} = 0.05$, for which the bias is 0.47. The straight line and the diagonal just above the straight line represent $\bar{\theta}_{U_h} = \bar{\theta}_{1h}$ and $\bar{\theta}_{U_h} - \bar{\theta}_{1h} = 0.025$, respectively. The bias is smaller than 0.05 for differences between $\bar{\theta}_{U_h}$ and $\bar{\theta}_{1h}$ of at most 0.025 in absolute terms for $\bar{\theta}_{U_h} \geq 0.25$. It is smaller than 0.03 for differences of at most 0.025 in absolute terms for $\bar{\theta}_{U_h} \geq 0.425$. Large biases in the lower right triangle have been supressed.

4. 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, $Bias(\hat{t})/\sqrt{V(\hat{t})}$, 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 $F = V(\hat{t}_{JB:r})/V(\hat{t}_{HT:s})$ be the 'estimator and nonresponse effect', where $\hat{t}_{HT:s}$ is the

Horvitz-Thompson estimator based on full response. Then, ignoring the finite population correction, the bias ratio is
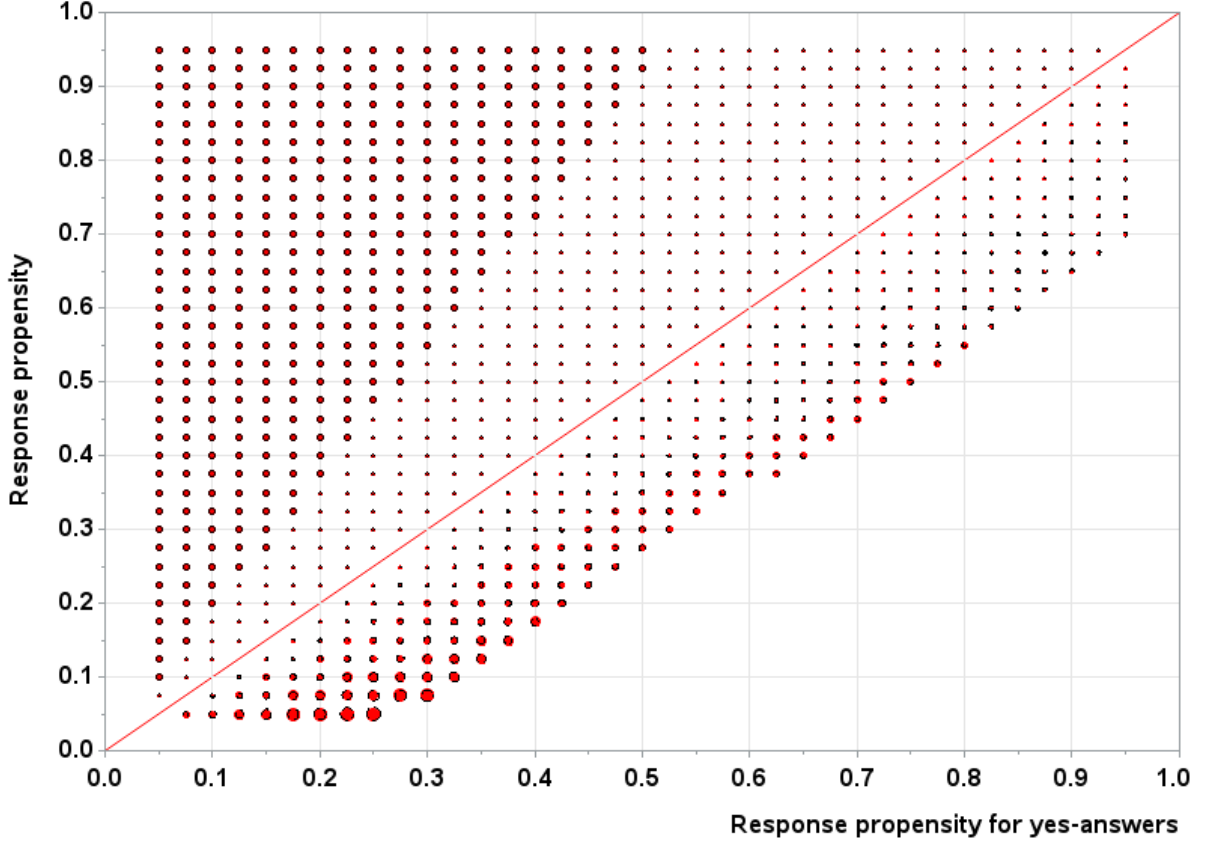
Figure 3. A bubble plot of the size of bias as in equation (5) in one poststratum, with $\bar{\theta}_{U_h}$ along the y-axis and $\bar{\theta}_{1h}$ along the x-axis. The straight line represents $\bar{\theta}_{U_h} = \bar{\theta}_{1h}$. The areas of the dots are proportional to the bias. The proportion to be estimated is 0.5.

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

where in the last approximation we have conservatively assumed that $F \approx n_s/n_r$, which makes cv_y to cancel out. To obtain a bias ratio of size b, then the mean response propensity should be

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

The benefit of (8) is that it turns our thinking from what bias we can accept, which is generally difficult to decide on, to the quantity bias ratio which is easier to discuss in terms of what coverage probability we may accept.

Consider a numerical example. For $\sigma_\theta = 0.29$, $\rho(y,\theta) = 0.05$, $b = 0.50$ and $n_r = 100$ in a stratum, $\bar{\theta}_U \geq 0.29$, which is a very modest response rate. For these numbers, the relative bias is 5% (see Figure 1). So if we accept a bias ratio of a half and the missing data mechanisms is mildly non-ignorable, we need to collect data from at least 100 out of 344 persons in a stratum. In this sense, we

can corroborate the term 'safe area' for the rectangle in Figure 1; the relative bias is not only rather flat within the rectangle even for greatly varying response rates, it is also small in terms of what coverage probability we may be willing to accept.

5. Discussion

We have focused on non-ignorable missing data mechanisms, where the correlation between the study variable and the response propensity, $\rho(y, \theta)$, is greater than zero in absolute terms. If it is zero within groups (e.g. poststrata), then there is no nonresponse bias. We believe that the assumption that $\rho(y, \theta) = 0$ is unrealistic.

The conflicting messages of, on the one hand, Groves (2006) and Groves and Peytcheva (2008) and on the other hand, the theoretical expression (3) seem resolved in Figures 1-3; for mild non-ignorability (small values of $\rho(y, \theta)$), the response rate is only weakly associated with the size of nonresponse bias as long as the response rate is not very low. The fact that Groves (2006) and Groves and Peytcheva (2008) do not detect a strong association suggests that in many surveys the correlation between the study variable and the response propensity is in general indeed low. However, there are exceptions, and both papers emphasise that there may be a large variation of bias over study variables in the same survey. Also note the importance of the availability and choice of auxiliary variables in sampling design and estimation; for one choice of poststrata the non-ignorability may be modest, for another, less successful, choice the nonresponse can turn out to be deleterious.

Another side of the coin is the fact that if $\rho(y, \theta)$ is larger than about 0.15 within groups (e.g. poststrata), the nonresponse bias may be considerable even for response rates as large as 0.70.

Brick and Tourangeau (2017) provide a useful typology of response propensity models. In their first three models, the random, the design-driven and the demographic-driven propensities model, most of the variation in response propensities are due to 'transient influences' (e.g. the sampled person is putting her or his baby in the bed when the call comes), design features or demographic characteristics that are only weakly associated with the characteristics of the sampled persons (e.g. incentives or age in some surveys), respectively. Of course, age is strongly associated with many common survey variables, but often you can poststratify by age and other demographic variables, and within poststrata the association may be weak. The fourth response propensities model is referred to as correlated propensities by Brick and Tourangeau (2017). This model is similar to the not missing at random response mechanism, NMAR (Little and Rubin, 2002), or a non-ignorable response mechanism. Brick and Tourangeau (2017) mentions 'a sense of civic obligation' as a cluster of variables (e.g. whether you vote) often related to response propensity. If such a variable is also a study variable, the correlated propensity model may be the model in the typology that fits best. We have in this paper focused on correlated propensity situations, where $|\rho(y, \theta)| > 0$. However, that does not mean that $\rho(y, \theta)$ must be large.

Let us now change to focus from $\rho(y, \theta)$ to the response rate. Going back to the statement by Särndal et al. (1992, p. 559) mentioned in the Introduction, that 'the greater the nonresponse, the more one has reason to worry about its harmful effects on the survey estimates', we have shown that this is to some extent true, there is a reason to be *worried*. In a situation when a survey practitioner has no knowledge about plausible sizes of $\rho(y, \theta)$, the *risk* for bias is higher if the nonresponse rate is 0.30 than if it is 0.70.

Lastly, a final word about surveys with very low response rates. If the response rate is very low, say 20% or lower, then we probably should not trust that survey, because then there may be considerable nonresponse bias even if $\rho(y, \theta)$ is miniscule.

References

Bethlehem, J. (1988). Reduction of nonresponse bias through regression estimation. Journal of Official Statistics, 4(3), 51-60.

Bethlehem, J. (2009). Applied survey methods: a statistical perspective. Hoboken: Wiley.

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

Brick, J.M. (2013). Unit nonresponse and weighting adjustments: a critical review. Journal of Official Statistics, 29(3), 329–353.

Brick, J.M. and Tourangeau, R. (2017). Responsive survey design for reducing nonresponse bias. Journal of Official Statistics, 33(3), 735-752.

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. 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. (2013). Facing the nonresponse challenge. The Annals of the American Academy of Political and Social Science, 645, 23-35.

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.

Lee, H., Rancourt, E. and Särndal, C.-E. (1994). Experiments with variance estimation from survey data with imputed values. Journal of Official Statistics, 10(3), 231-243.

Little, R.J.A. and Rubin, D.B. (2002). Statistical analysis with missing data, 2nd ed. Hoboken: Wiley.

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.

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. 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.

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

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.