

Responsive Survey Designs for Reducing Nonresponse Bias

J. Michael Brick¹ and Roger Tourangeau¹

Survey researchers have been investigating alternative approaches to reduce data collection costs while mitigating the risk of nonresponse bias or to produce more accurate estimates within the same budget. Responsive or adaptive design has been suggested as one means for doing this. Falling survey response rates and the need to find effective ways of implementing responsive design has focused attention on the relationship between response rates and nonresponse bias. In our article, we re-examine the data compiled by Groves and Peytcheva (2008) in their influential article and show there is an important between-study component of variance in addition to the within-study variance highlighted in the original analysis. We also show that theory implies that raising response rates can help reduce the nonresponse bias on average across the estimates within a study. We then propose a typology of response propensity models that help explain the empirical findings, including the relative weak relationship between nonresponse rates and nonresponse bias. Using these results, we explore when responsive design tools such as switching modes, giving monetary incentives, and increasing the level of effort are likely to be effective. We conclude with some comments on the use of responsive design and weighting to control nonresponse bias.

Key words: Adaptive design; response propensities; two-phase sampling.

1. Introduction

Responsive and adaptive survey designs are attempts to cope with a survey climate characterized by declining budgets, rising costs, and falling response rates (Tourangeau et al. 2017). The good news is that, to date, there is little evidence that falling response rates have had large effect on the quality of survey estimates. For example, Silver (2014) reports that polls done in the last three weeks of an election campaign were about as accurate in 2012 as in 2000, despite the steep fall in response rates over that period. Silver examined polls on presidential, gubernatorial, senatorial, and House races and in each case the accuracy was essentially flat over those twelve years.

In this article, we focus on the application of responsive or adaptive design to reduce nonresponse bias. While there are other reasons for using such designs, mitigating nonresponse bias has been the central issue in this since responsive design was first introduced by Groves and Heeringa (2006). To accomplish this, understanding the relationship between nonresponse rates and bias is critical so that responsive data collection approaches can specifically deal with nonresponse bias.

¹ Westat, 1600 Research Blvd, Rockville MD 20850 USA. Emails: mikebrick@westat.com and RogerTourangeau@westat.com

Acknowledgment: We would like to thank Drs. Groves and Peytcheva for providing the data set they used in their original work for our reanalysis.

Several studies have examined the issue of the relationship between nonresponse rates and nonresponse bias, including ones by Keeter and his colleagues (Keeter et al. 2000; Keeter et al. 2006), Curtin et al. (2000), Merkle and Edelman (2002), Groves (2006), and Chun (2009). All of these studies conclude that the relationship between the two is weak. For example, Merkle and Edelman (2002) examined the accuracy of exit polls, comparing the poll results for a precinct to the actual precinct vote shares. Over four elections, the authors found little relationship between response rates at a given precinct and the signed error in the estimates; these correlations ranged from $-.01$ to $.10$. But the most influential of these studies is the Groves and Peytcheva (2008) meta-analysis of 59 prior studies that had assessed the level of nonresponse bias in statistics. The studies drew on data from frames, administrative sources, screening interviews, and nonresponse follow-up surveys to characterize the nonrespondents and to estimate the bias in the survey statistics. Figure 2 in their article displays dramatic within-study variation in nonresponse biases. Within a single study, some estimates appear to be nearly unbiased, whereas others are off by as much as 70 to 100 percent. The Groves and Peytcheva (2008) analysis has been used to suggest that the survey response rate is irrelevant or at least problematic (Davern 2013), and that no one number is useful for characterizing the level of nonresponse bias in the estimates from a survey (Groves et al. 2008). If these interpretations of the Groves and Peytcheva findings were valid, then R-indicators (Schouten et al. 2009) and balance indicators (Särndal 2011) would be unable to characterize the nonresponse bias in a study – there is simply too much variation in the biases across the estimates from the same study.

We re-examine the Groves and Peytcheva data and look at both the within- and between-study components of variance in the nonresponse bias estimates in the next section. Our findings are more nuanced and give a different interpretation of the relationship between nonresponse bias and response rates than many researchers have drawn. We find that there is a substantial across-study variation in nonresponse bias along with the large within-study variation already observed.

In the third section of the article, we take a more theoretical perspective on the relationship between nonresponse rates and bias. The approach assumes the response propensities have a specified parametric distribution and then re-expresses the theoretical bias using the assumed model. Under these conditions, we show that higher nonresponse rates are likely to be associated with a higher level of nonresponse bias. We also propose a typology of models for response propensity mechanisms that can be used to describe both the survey process and patterns of nonresponse bias in the estimates. These models help explain the discrepancy between the strong relationship between nonresponse bias and nonresponse rates predicted by theory and the weaker relationship observed in practice.

The fourth section uses the relationships and proposes which responsive design strategies are likely to be effective in reducing nonresponse bias. These strategies include switching modes, providing incentives, and increasing levels of effort. Two-phase sampling also plays an important role in responsive designs because it helps control data collection costs when more costly or additional activities are implemented to reduce nonresponse bias. In this section, we clarify the role of two-phase sampling in responsive design.

The last section summarizes our findings and describes situations in which responsive designs may help reduce nonresponse bias. We also discuss the roles of responsive design and nonresponse weighting adjustments.

2. The Relationship between Nonresponse Rates and Nonresponse Bias: A Second Look

Groves and Peytcheva (2008) examined bias estimates for a total of 959 statistics from 59 surveys. They report “If a naïve OLS regression line were fit [to the 959 estimates] . . . , the R^2 would be 0.04” (Groves and Peytcheva 2008), implying a correlation between nonresponse bias and nonresponse rate of around 0.20. We re-examine these data here by looking at both the study-level and within-study contributions to the overall nonresponse bias in these data. By including the study-level contribution in the analysis, the importance of a study-level characteristic, including the response rate, can be better evaluated. Furthermore, this analysis helps to understand the potential effectiveness of adaptive designs that aim to increase response rates for a study.

The data set we received from the authors included all 959 bias estimates from 59 studies. Eight hundred and four of the bias estimates involved proportions; most of the rest involved means. We examine the absolute relative bias statistic (hereafter absolute relbias), or the absolute difference between the respondent estimate and the full sample estimate divided by the full sample estimate:

$$R_i = \frac{|\theta_{ri} - \theta_{ni}|}{\theta_{ni}}, \quad (1)$$

in which R_i is the absolute relbias for statistic i , θ_{ri} is the estimated value for that statistic based on the respondents, and θ_{ni} is the full sample estimate.

Looking at all 959 relbias estimates in our data set, we find a weak overall correlation between the absolute relbias statistics and the response rate – the correlation is $-.200$. This low overall correlation was one of the key findings from Groves and Peytcheva.

The high level of within-study variation in the relbias estimates found by Groves and Peytcheva (2008) does not mean that there is not substantial between-study variation as well. We therefore examined the study-level means and medians of the absolute relbias estimates. We refer to study-level means as average nonresponse bias estimates. For example, if bias estimates (or absolute relbias estimates) are computed for ten estimates for a study, then the mean of these estimates is the average nonresponse bias estimate. The correlation between the study-level mean absolute relative bias estimate and the response rate is $-.357$ and the correlation between the median absolute relbias difference and the response rate is $-.367$. In other words, the studies do reveal that the study-level nonresponse bias is correlated significantly with the response rate. Table 1 summarizes these findings.

Some of the studies contribute only one or two bias estimates so that the study-level means and medians are highly variable. If we weight each study by the number of bias estimates from the study, the correlations are $-.412$ for the mean absolute relbias of the estimates and $-.474$ for the median absolute relbias. Of course, the estimates within a study could be highly intercorrelated, so we also reweighted the study-level means using the average sample size on which the relbias estimates were based as the weight rather than the number of estimates from the study. (The sample sizes varied within study, since not every item applied to every case.) The bottom rows of Table 1 show these results. Again, we find correlations with the study response rates roughly twice those reported in

Table 1. Correlation between response rates and inverse of the response rate and relBias measures at the estimate and study level.

	Response rate	1/response rate
Unweighted estimate-level correlations		
Response rate and absolute relbias	– .200 (959)	.175 (959)
Unweighted study-level correlations		
Response rate and mean absolute relbias	– .357 (59)	.339 (59)
Response rate and median absolute relbias	– .367 (59)	.354 (59)
Study-level correlations – weighted by number of estimates		
Response rate and mean absolute relbias	– .412 (59)	.362 (59)
Response rate and median absolute relbias	– .474 (59)	.431 (59)
Study-level correlations – weighted by average sample size		
Response rate and mean absolute relbias	– .494 (59)	.556 (59)
Response rate and median absolute relbias	– .500 (59)	.575 (59)

Note: All the correlations in the table are significant at the .05 level or lower.

Groves and Peytcheva (2008)— – .494 for the mean relbias and – .500 for the median relbias. Table 1 also shows the pattern is quite similar if we examine the relationship between the relbias and the inverse of the response rate; it is the inverse of the response rate that, in theory, is related to the level of nonresponse bias (see, for example, Equation 2 below).

Figure 1 depicts these relationships graphically. The top panel of the figure is a bubble plot showing the relationship between the mean absolute relbias and the study response rate. The size of the bubbles in that panel is proportional to the number of relbias estimates from the study. The bottom panel in the figure shows the same data, this time weighting each study by the mean number of cases on which the relbias estimates are based.

One of the apparent conclusions from the Groves and Peytcheva study is that the within-study variation in nonresponse bias is so large that no one number can accurately characterize the overall risk of nonresponse bias in a study. Their analysis did not, however, directly examine the level of *between*-study variation (Because of sampling error, one would not expect the absolute relbiases or absolute differences to be zero. We redid some of our analyses incorporating a correction for sampling error. These analyses examined the absolute bias in an estimate less the expected absolute difference based on sampling error. This correction factor seemed only to attenuate all the relationships reported here; that is, the correction only muddled the waters.).

The findings in Table 1 support the existence of a substantial study-level component to nonresponse bias. If we do a simple one-way analysis of variance with the relbias estimates as the dependent variable and the studies as the independent variable, the resulting R^2 is .233 – that is, variation across studies is responsible for about a quarter of the variation in the absolute relbias estimates. This strongly suggests that nonresponse bias is partly a function of study-level characteristics; the correlational results indicate that one of those study-level characteristics is the study’s response rate. Response rates may not be very good predictors of nonresponse bias, but they are far from irrelevant.

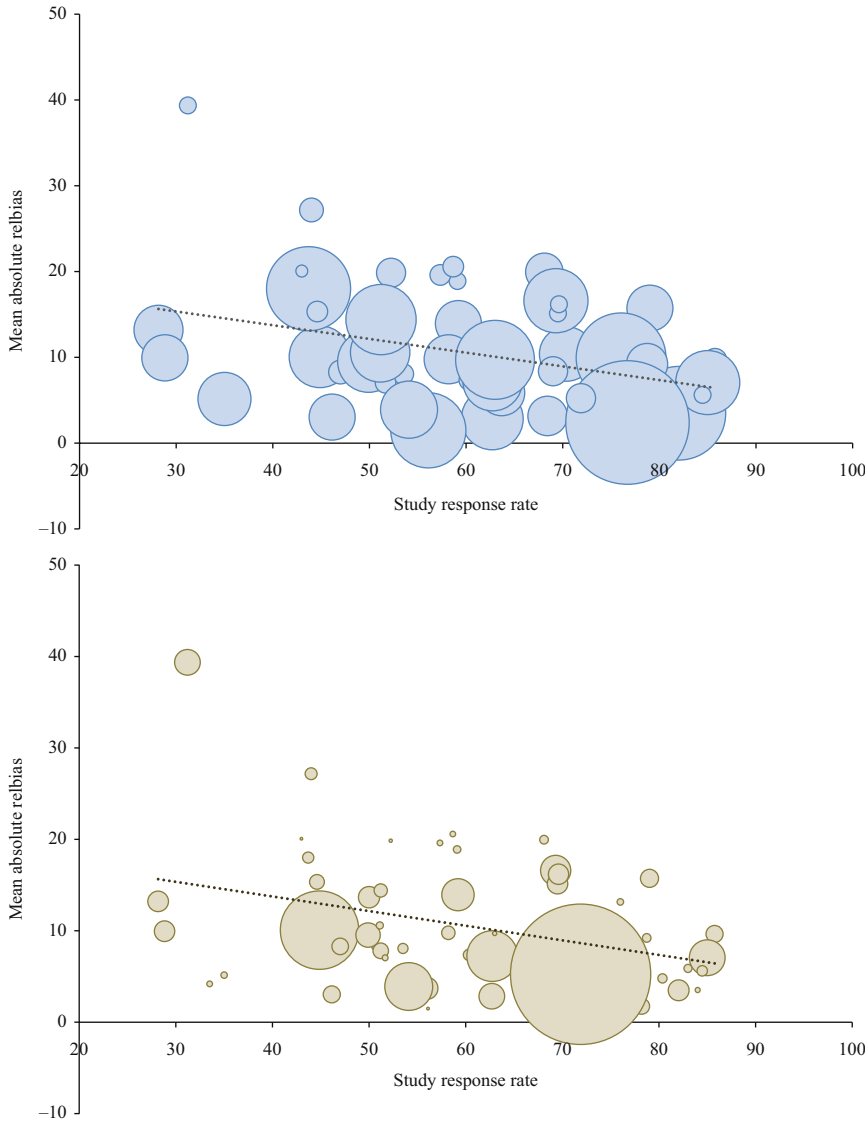


Fig. 1. Bubble plots of the mean absolute relbias of the estimates against the study's response rate. In the top figure, the bubble sizes are proportional to the number of bias estimates from the study. In the bottom figure, they are proportional to the average sample size for the estimates from the study. The dashed lines are the regression lines for the response rate-mean relbias relationship. The regressions were weighted by the number of estimates (top figure) and average sample size (bottom figure).

3. Theory for Response Propensity Distributions

Our reanalysis of the Groves and Peytcheva data shows that response rates empirically provide a useful study-level indicator of nonresponse bias. Thus, some reconsideration of the theoretical basis for a relationship between nonresponse rates and nonresponse bias is warranted. We take two approaches to the theory. First, we examine the statistical theory

by assuming the response propensity distributions have specific parametric forms (a continuous beta distribution and a discrete distribution). The value of exploring specific distributions is that it provides insights into the conditions under which increases in response rates will result in lower nonresponse biases. Previous nonparametric approaches have not been very instructive, and simulations may not be relevant to actual survey conditions.

Second, we propose a typology of response propensity models for sample surveys that identify conditions under which different correlations between the response propensities and outcome variables might arise. The goal is to provide models that link the statistical theory to the empirical evidence. These models can also be used to anticipate situations in which large nonresponse biases might be expected. If the models suggest the survey might produce large biases for key estimates without extra efforts, then adaptive survey designs could be implemented to reduce the potential biases.

3.1. Parametric Representation of Response Propensities

Using a stochastic perspective, [Bethlehem \(1988\)](#) expresses the bias of the simple expansion estimator of a mean or proportion (\hat{y}) as

$$\text{bias}(\hat{y}) \approx \frac{\sigma_\phi \sigma_y \rho_{\phi,y}}{\bar{\phi}}, \quad (2)$$

where $\bar{\phi}$ and σ_ϕ are the mean and standard deviation of the response propensities, σ_y is the standard deviation of the y 's, and $\rho_{\phi,y}$ is the correlation between the response propensities and the y 's; we assume that all units in the population have a nonzero chance of responding to the survey. As [Brick \(2013\)](#) noted, the response propensities are a function of the survey effort so it is very realistic that some units will not respond to a particular survey protocol.

One rationale for assuming lower response rates should be related to higher levels of nonresponse bias is that the mean of the response propensities is in the denominator of (2), which is why [Table 1](#) included the inverse of the study response rate as a possible predictor of the study-level nonresponse bias. However, in combination with growing empirical evidence of a more complex relationship, the other terms in (2), especially the correlation, have become the focus of theoretical explorations of nonresponse bias ([Little and Vartivarian 2005](#)). The R-indicator is a simple function of the standard deviation of the response propensities and (2) makes it clear why it should be related to the nonresponse bias.

Rewriting (2) as

$$\text{bias}(\hat{y}_{un}) \approx \bar{\phi}^{-1} \sigma_\phi \sigma_y \rho_{\phi,y} = cv(\phi) \sigma_y \rho_{\phi,y} \quad (3)$$

(where $cv(\phi) = \frac{\sigma_\phi}{\bar{\phi}}$ is the coefficient of variation of the response propensities) provides some insight into the role of the response rate in nonresponse bias. [Särndal and Lundström \(2005\)](#) used a similar expression but mainly for identifying auxiliary variables for weight adjustment to reduce nonresponse bias.

We examine the implications of (3) when the response propensities have two very different distributions. First, we assume the response propensities have a continuous, Beta distribution with shape parameters (α, β) , say $\phi \sim \text{Beta}(\alpha, \beta)$ and $\alpha > 1$, $\beta > 1$. The

mean of the Beta distribution is $\mu = \frac{\alpha}{\alpha+\beta}$ and its variance is $\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$. The Beta is often used in fitting densities because, by the appropriate choice of parameters, it can take very different shapes; Schouten et al. (2016) also use the Beta distribution to simulate response propensities. Figure 2 shows four Beta distributions with different shape parameters yet all have the same mean of 0.20. The bias of the mean can be very large when the response propensities have a Beta distribution if the correlation between the propensities and the survey variable ($\rho_{\phi,y}$) is large.

Since $\beta = \frac{\alpha(1-\mu)}{\mu}$, the coefficient of variation of a Beta distribution can be written as

$$cv(\phi; \alpha, \mu) = \sqrt{(1 - \mu)/(\alpha + \mu)}. \tag{4}$$

With this expression, the ratio of coefficients of variation of two Beta distributions with the same α is

$$RCV_{\beta}(\mu|\alpha, \mu^*) = \frac{cv(\phi; \alpha, \mu)}{cv(\phi; \alpha, \mu^*)} = \sqrt{\frac{1 - \mu}{1 - \mu^*} \frac{\alpha + \mu^*}{\alpha + \mu}}, \tag{5}$$

where the coefficient of variation in the denominator has the mean of the propensities fixed at a specified value, say μ^* . Figure 3 shows that $RCV_{\beta}(\mu; \alpha, \mu^* = 0.8)$ is almost linear and not very dependent on the value of α (values of $\alpha > 5$ are virtually indistinguishable from those with $\alpha = 5$). The figure also reveals that when the mean of the response propensities is smaller than μ^* , then the ratio of the coefficient of variation is greater than 1. For example, when the mean of the response propensities is 0.2 (a 20% response rate), the $cv(\phi)$ in (5) is roughly two times larger than the $cv(\phi)$ when the mean of the response propensities is 0.8.

This relationship between the mean response propensity and the cv of the propensities also holds when the response propensity distribution is discrete and bimodal – a

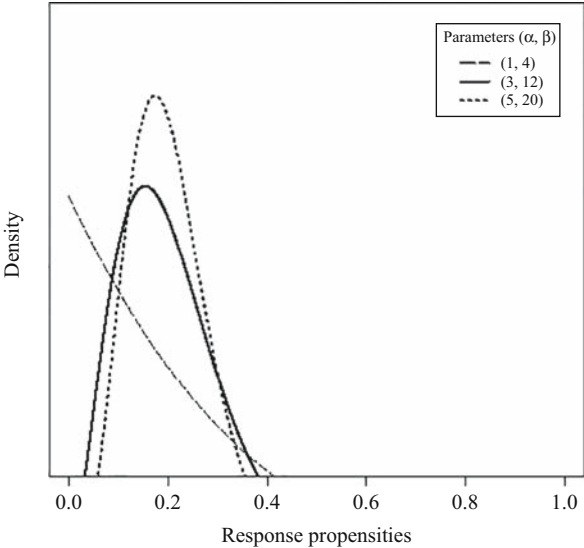


Fig. 2. Beta distributions with different shape parameters but a common mean of 0.2.

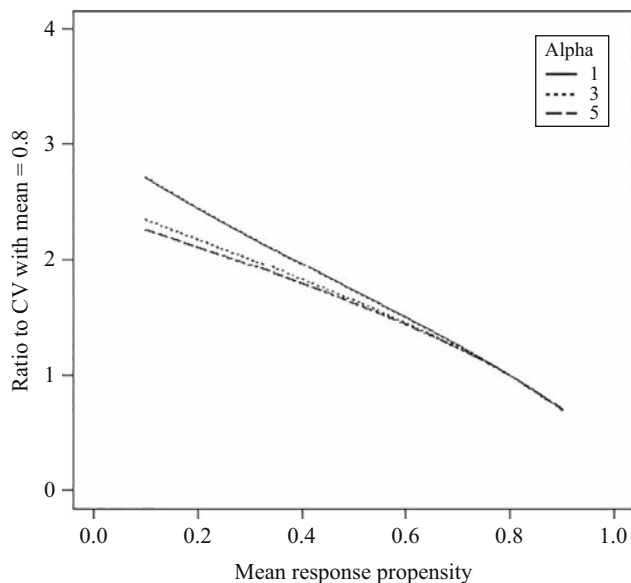


Fig. 3. Ratios of $cv(\phi)$'s ($RCV_{\beta}(\mu; \alpha, \mu^* = 0.8)$) for Betas with different α values, by the mean response propensity.

distribution very different from the continuous Beta distribution. Consider a distribution where some units have a low propensity to respond and the remaining units have a high propensity to respond. This distribution places all the mass at the two points corresponding to the values of low and high (we refer to this as a *point mass* distribution). In this case, the ratio of the coefficient of variation is easily shown to be a function of the mean of the distribution (the weighted average of the proportion with low and high propensities). We can express the equivalent of (4) for the point mass distribution as $RCV_p(\mu; \mu^*) = \sqrt{\frac{\mu^*}{\mu} \frac{(1-\mu)}{(1-\mu^*)}}$. The important point is that this ratio, like that with the Beta distribution, increases as the mean response propensity increases.

The two factors that influence the magnitude of the bias and vary with the data collection protocol are the $cv(\phi)$ and $\rho_{\phi,y}$. As shown above for both the Beta distribution and the point mass distribution, the $cv(\phi)$ increases substantially as the mean response rate decreases—lower response rates are associated with greater bias, even if $\rho_{\phi,y}$ increases somewhat.

A common reason for implementing a responsive design is to reduce nonresponse bias when the bias is substantial. (The bias is only substantial when $\rho_{\phi,y}$ is large, say greater than 0.5.) If the responsive design protocol succeeded in increasing the response rate from an initial 50 percent to 80 percent, then the final $\rho_{\phi,y}$ would have to be more than 1.6 (Figure 3) times larger than the value under the initial effort for the bias to increase (assuming the propensities were from a Beta distribution). With the point mass distribution the correlation would have to be more than twice as large. The point is that if a responsive design increases the overall response rate substantially, then it is likely to reduce the biases of estimated means, because it is very unlikely (but not impossible) that $\rho_{\phi,y}$ will increase at even a greater rate than the response rate increases.

Although the Beta and point mass distributions do not exhaust the possibilities, we regard them as plausible models that illustrate the importance of high response rates. Although it might be possible to find propensity distributions in which the response rates do not matter, we doubt they would be realistic candidates for most actual surveys.

What sort of responsive design protocol might cause an increase in the response rate *and* the nonresponse bias? Lundquist and Särndal (2013) show that following an initial protocol without major changes is a likely candidate – in other words, continuing more of the same effort that gave rise to the bias in the first place. For a more thorough examination of this issue, we examine several response propensity models.

3.2. Response Propensity Model Typology

Both our reanalysis of the Groves and Peytcheva (2008) data and the parametric representation of the response propensity distribution in the previous section indicate response rates should be correlated to nonresponse bias; a study with a higher response rates should have estimates that have lower *average* nonresponse bias. The empirical evidence finds that the strength of the relationship is not as large as we might expect based on the theory. Large differences in the correlations between response propensities and the different survey variables (differences in the $\rho_{\phi,y}$'s) would cause substantial within-study variation in the nonresponse bias of the estimates, but is there any reason to anticipate large differences in the correlations?

We propose a typology of response propensity models that may help provide possible explanations for the large variations in the correlations observed in practice. Before presenting the typology, we emphasize that response propensities as discussed in Brick (2013) are a function of both the sampled person's characteristics and design factors, such as the number of contact attempts and interviewer characteristics. Some of these characteristics may be constant (e.g., sampled person's sex) and others may change over time (e.g., sampled person's employment status or health).

First, it could be that some or most of the variation in response propensities is due to factors that have a large random error component. If response propensities are a function of a large number of transient influences (whether the person happens to be sick at the time of the initial contact, his or her mood on that day, how busy he or she is at work, who in the household answers the door or telephone, his or her travel schedule, whether the person read the advance letter or threw it away, whether he or she happens to hit it off with the interviewer, whether the television is on when the telephone rings, etc.), then whether a given person takes part in a study may, in effect, be the outcome of a kind of multidimensional coin flip. Typically, the response outcome is the result of one or two interactions with the study and the random component (sick on the date of contact) is not averaged or smoothed over multiple events. If so, nonresponse is unlikely to produce many estimates with large nonresponse bias, and the average bias should be small. The average bias would be low because the response propensities are unlikely to be related to any stable respondent characteristics. We refer to this as the *random propensities* model.

A second model postulates that most of the variation in response propensities is determined by study-level design features that are largely unrelated to characteristics of the sample members. We refer to this as the *design-driven propensities* model. A

mandatory survey (like the American Community Survey, or ACS) that features repeated callbacks and offer opportunities to respond by web, mail, telephone, or face-to-face is likely to produce a high response rate with limited variation in propensities across different sample members. In such a survey, nearly all groups of people are very likely to respond eventually. The high response rate does not prevent some estimates from having substantial biases (McQuillian et al. 1994), but it would be highly unlikely that the average bias would be large. Similarly, a survey conducted over a very short time period, say two- or three-days, with no advance letter or incentive is (in the U.S., at least) unlikely to produce a high response rate. Although it may be unlikely that *anyone* has a very high response propensity for such a study or that response propensities would vary much across people, there may be subgroups such as the retired persons who are more likely to be respondents. Thus, the bias of some estimates for this subgroup might be relatively large and the average of the biases should be expected to be greater than if the response rate were substantially higher. Nevertheless, if response propensities are mostly driven by features of the data collection protocol rather than by characteristics of the sample members, then most estimates are likely to have relatively low average nonresponse bias. This model may help explain why many low response rate telephone surveys that have been examined still have relatively modest biases.

A third possibility is that propensities are determined, at least in part, by household or person-level characteristics, but the characteristics reliably related to response propensities are only weakly related to most survey variables. For example, age, sex, urban residence, education, home ownership, and having a listed telephone number are consistently related to response propensities across a range of surveys (e.g., Groves and Couper 1998). In general, in the U.S., older people are more likely to respond than younger people, women more likely to respond than men, residents of suburban and rural areas more likely to respond than residents of central cities, well-educated persons more likely to respond than less-educated persons, home owners more likely to respond than renters, people with listed telephone numbers than people without listed numbers, and so on. We refer to this as the *demographic-driven propensities* model. Again, when this model holds, we might expect most estimates to have relatively low levels of nonresponse bias because the survey variables are weakly correlated to these demographic variables. The estimates based on the few variables that are highly correlated with the household or person-level characteristics are likely to have large biases, but the average bias would still be relatively low.

A fourth model is that response propensities vary both with survey design features and with other characteristics of the sample members. This model – the *correlated propensities* model – implies that some groups are consistently overrepresented in the respondent set. Biases for the variables highly correlated to these characteristics can be large. This may be true regardless of the overall response rate if the survey protocol produces a high level of overrepresentation of certain groups among the respondents. This corresponds to the missing not at random assumption.

Both the random and design-driven propensities models imply that nonresponse biases should be small on average. These models roughly correspond to the missing completely at random assumption, in which the estimate is unbiased. In the Groves and Peytcheva data, the percentage estimates are off by 2.1 percent on average (this is the mean absolute difference across the 804 proportion estimates, re-expressed as percentages). The

demographic-driven propensities model implies small biases except on survey variables that are strongly related to one or more of the person-level characteristics that often affect response propensities. Even in these cases, if the person-level covariates are observable and used in nonresponse adjustment, the biases in the estimates when adjusted for the demographics should be small (this is similar to the missing at random assumption). Responsive design procedures that increase the response rate should, in these cases, cause average study-level nonresponse bias to decrease, especially if the increase in response rate produces better balance on the demographic variables.

Because estimates with a large average nonresponse bias are most likely to occur when the correlated response propensities model holds, the question is when this arises in practice. A number of studies show that a cluster of variables related to a sense of civic obligation, such as estimates on voting and volunteering, are highly related to both survey participation and these survey outcomes. If a survey attempts to estimate these types of variables, there is substantial potential for large average biases. [Tourangeau et al. \(2010\)](#) find very high levels of nonresponse bias in a survey of voters and nonvoters in the State of Maryland; the voters were strongly overrepresented among the survey respondents, producing large overestimates of the level of political participation. Similarly, [Abraham et al. \(2009\)](#) find that volunteers are strongly overrepresented in the American Time Use Study, introducing systematic distortions in the estimates. Still, in any survey, it is likely that essentially chance factors will affect the propensities, attenuating their relationship to the survey variables and reducing the risk of large biases.

4. Implementing Responsive Design to Reduce Nonresponse Bias

[Table 1](#) shows that the 59 surveys have relatively low average rebias for most of the estimates, but, even though the overall average is low, there are some estimates with very large rebias. This is consistent with the typology of response propensity models, where the correlated response propensity model is not likely to hold for many survey variables. This has two implications for survey designers considering responsive design to reduce nonresponse bias. First, the average nonresponse bias can be reduced somewhat by increasing overall response rates, but to be effective the response rates must be increased substantially. Second, estimates with high levels of nonresponse bias in a survey that has an average overall level of bias can be addressed only by directing responsive design efforts to counteract specific imbalances in the respondent set. Frame data could be used for this purpose, but auxiliary or frame data are generally not available for many of the items, such as civic participation, that are likely to have large biases. [Tourangeau et al. \(2017\)](#) point out that an alternative is to curtail further efforts with high propensity cases in the hope that this will increase the sample balance with respect to the variables not available during data collection. If this is not an acceptable approach, then the responsive design protocol at least should employ methods that are very different in terms of engaging the nonrespondents. Ideally, the second phase or responsive design stage of data collection should use methods that produce conditional propensities (for the latter stage) that are inversely related to the propensities before the change.

The traditional approach to conducting surveys is to design a data collection protocol and apply it uniformly to all sampled units. This “equal effort” approach is well

established in data collection organizations and can be effective for achieving reasonably high response rates. However, unequal effort methods are central in responsive designs to reduce nonresponse bias.

Three unequal effort methods that have the potential to boost response rates are: (1) changing the mode of data collection, (2) providing higher monetary incentives, and (3) exerting higher levels of data collection effort. Other methods such as using more experienced interviewers and modifying the length of the instrument have been used, but the changes in response rate for these alternatives tend to be modest in most cases (Tourangeau et al. 2017; Groves and Cooper 1998). We briefly describe these three unequal effort methods and discuss their potential for reducing biases in a responsive design.

4.1. *Switching Modes*

Changing the mode of data collection can result in relatively large increases in response rates. The ACS (U.S. Census Bureau 2014) is an example of a survey that transitions from Internet to mail to telephone and then to face-to-face for a subsample of the remaining nonrespondents. The initial response rates to Internet and mail increase by over 30 percentage points (when weighted for the subsampling) by using the multiple modes. The final weighted response rates are around 95 percent. Messer and Dillman (2011) provide another example of a survey in which switching modes increases response rates. They conducted experiments that began with a low-cost mail invitation to complete a survey on the Internet and then switched to a mail questionnaire. This mode switch increased response rates from the Internet-only approach by 12 to 20 percentage points, depending on other experimental factors.

Implementing a responsive design that switches modes may substantially increase response rates and thus lower average nonresponse bias. Messer and Dillman (2011) also show that the large biases in estimates from the Internet-only data were reduced significantly by switching modes. This implies that, even when mode switches do not target specific groups, the mode switch provides a different response stimulus for some groups of cases with low response propensities to the Internet mode. Switching modes is a form of responsive design that appears to have potential to reduce nonresponse bias in some surveys.

4.2. *Incentives*

Incentives, especially non-contingent cash incentives, consistently result in higher response rates (Singer 2002). Mercer et al. (2015) conducted a meta-analysis showing that incentives increase response rates, but quickly reach a point of diminishing returns. This finding is consistent with previous research on incentives (e.g., Trussell and Lavrakas 2004). Mercer et al. (2015) find that the increase in response rate due to offering an incentive compared to no incentive ranges from about ten percentage points for face-to-face and telephone surveys to 20 percentage points for mail surveys.

Groves et al. (2000) suggest that incentives might operate to increase the response propensities of those who are otherwise least likely to respond, while having little effect on those with higher response propensities. Under this theory, nonresponse bias should be

reduced because the variation in response propensities are reduced. Although some experiments have shown the theorized effects, many experiments have not (Groves et al. 2006). In these latter studies, incentives increase response propensities overall without reducing the variation in the propensities. Consequently, responsive designs that add incentives are likely to raise overall response rates and may reduce average nonresponse bias somewhat. However, it is unclear when incentives substantially reduce the variation in response propensities and the risk of large nonresponse biases in the estimates.

4.3. *Level of Effort*

As noted earlier, some surveys adopt an equal effort approach and apply more of the same type of effort to increase response rates. Keeter et al. (2000) report on an experiment in which extra efforts in a telephone survey raised response rates from 36 percent to 60 percent. This increase of 24 percentage points is among the highest in this type of study. Särndal and Lundquist (2014) show substantial increases in sample balance are possible by limiting the number of contact attempts.

Several studies examine the effects of higher levels of effort by computing response rates at various stages of the data collection process rather than manipulating the data collection itself (Bartholomew 1961; Curtin et al. 2000; Särndal and Lundquist 2014). For example, Särndal and Lundquist (2014) look at the estimates that would have been computed if the data collection ended earlier in terms of the number of contact attempts. In these studies, large differences in response rates are simulated.

The effects on nonresponse bias in most of these studies have been relatively small, consistent with the empirical and theoretical evidence presented earlier. Simply changing the level of effort may increase response rates and thereby slightly lower average nonresponse bias, but there is no reason to expect that it will reduce the variation in response propensities. In fact, the findings of Särndal and Lundquist (2014) suggest that doing more of the same type of effort may actually increase the variance of the response propensities in some cases. This implies that responsive designs that simply increase the level of effort are unlikely to reduce nonresponse biases despite increasing the response rate.

4.4. *Two-Phase Sampling*

Two-phase sampling is a key component in many responsive designs. The first use of the two-phase approach for nonresponse is Hansen and Hurwitz (1946), who proposed a face-to-face follow-up of a subsample of nonrespondents to a mail survey. In their application, the follow-up obtained essentially complete response; the goal of subsampling was to reduce the total cost of the survey. The efficiency is achieved by getting more completes from the low-cost first phase sample.

Building upon this idea, Groves and Heeringa (2006) show that two-phase sampling could be a major tool in responsive designs. However, it is important to realize that two-phase sampling is really about reducing data collection *cost*, as discussed by Hansen and Hurwitz (1946). Nonresponse bias is not likely to be very different whether all cases or a subsample of cases (appropriately weighted for the subsampling) are included in the second phase if the second phase protocol is similar to the first phase protocol. Two-phase sampling could leave nonresponse bias unchanged, reduce it, or even increase it. To

reduce biases substantially for items with large biases after the first phase, the data collection protocol for the second phase must be effective in obtaining responses from cases with low response propensities in the first phase. As discussed above, the type of targeting in the second phase to reduce the variation in response rates often requires auxiliary data that are not available. In this case, two-phase sampling may help at least to reduce costs, but its effect on nonresponse bias may not be very large.

5. Discussion

As cost pressures continue to mount, alternatives that reduce costs while controlling nonresponse bias are needed and responsive design can be a valuable tool in accomplishing this. Our reanalysis of the Groves and Peytcheva data and our theoretical results show that raising response rates can help reduce average nonresponse bias in the estimates. Responsive design approaches such as switching modes, giving monetary incentives, and increasing the level of effort can be used in some settings to increase response rates. Two-phase sampling, where the more intensive and expensive methods are applied to only a subsample of the cases, can help reduce data collection costs. These methods, when they are effective, help by raising the mean of the response propensities, reducing their variation, or both.

Increasing overall response rates is far from a panacea. The slopes of the regression lines in [Figure 1](#) are not very steep, so the response rate needs to be increased substantially to lower the average nonresponse bias. Much of the variation in the bias of the estimates is within-study and is driven by the correlation between the characteristics being estimated and the response propensities. Thus, the targeting of responsive design features to address those estimates that are expected to have the largest biases is extremely important. Mode switches and incentives sometimes have this effect, but just increasing the level of effort is not likely to be very effective even if it increases the overall response rates. With some characteristics such as volunteering and voting, even mode switches and incentives are unlikely to have a substantial effect on the magnitude of the nonresponse bias.

Weighting for nonresponse is an alternative that may also reduce biases ([Särndal and Lundström 2005](#)). Recent research suggests that decreasing the variation in response propensities during the data collection stage may provide additional protection against nonresponse bias beyond what nonresponse weighting can achieve alone ([Schouten et al. 2016](#); [Moore et al. 2016](#); [Sturgis et al. 2017](#); [Luiten and Schouten 2013](#); [Särndal et al. 2016](#)). These two approaches to control nonresponse bias are not in conflict, but are complimentary. This is especially true in the United States and other countries where few auxiliary variables related to nonresponse propensities and estimates of interest are known prior to data collection. In this situation, a few geographic characteristics and paradata may be the only data available for responsive design implementation. Other auxiliary data, such as those from the ACS or other large surveys with high response rates, can only be used in nonresponse weight adjustments after data collection. When data are available that could be used for responsive design and for weighting, researchers must evaluate whether investing in reducing imbalance through responsive design is worth the additional cost. Nonresponse weighting should be done regardless of whether a responsive design is implemented.

In practice, we believe data collection and weighting adjustments are not equivalent. Consider a simple example. Suppose the respondent set is not well balanced with respect to income and this variable is available on the sampling frame. Targeting low income nonrespondents by giving them higher incentives may be effective in increasing their propensity to respond. Targeting the data collection effort in this way may also produce a respondent set that is more balanced with respect to other variables that are not on the frame, such as citizenship status. Adjusting the weights of the initial data set might give roughly the same bias reduction for income as using the final, more balanced, respondent set with the same weighting adjustment. However, for other variables, such as citizenship status, the bias may be different for the two weighted data sets. The effect depends on whether the incentive reduces or increases imbalance with respect to citizenship.

Reducing imbalance incurs data collection costs but it may provide a more robust data set for producing a variety of estimates. The decision to allocate additional efforts to reduce bias should recognize both these costs and the potential benefits of the additional data collection with a responsive design. Clearly, a low response rate survey with large imbalances has the most to gain from a responsive design. In this case, large biases may be mitigated and the costs of data collection could be controlled by using two-phase sampling. On the other hand, implementing a responsive design that is aimed primarily at an arbitrary response or yield rate, without targeting or otherwise reducing the variation in the response propensities, is unlikely to be effective.

6. References

- Abraham, K., S. Helms, and S. Presser. 2009. "How Social Processes Distort Measurement: The Impact of Survey Nonresponse on Estimates of Volunteer Work in the United States." *American Journal of Sociology* 114: 1129–1165.
- Bartholomew, D.J. 1961. "A Method of Allowing for 'Not-at-Home' Bias in Sample Surveys." *Applied Statistics* 10: 52–59. Doi: <http://dx.doi.org/10.2307/2985408>.
- Bethlehem, J.G. 1988. "Reduction of Nonresponse Bias Through Regression Estimation." *Journal of Official Statistics* 4: 251–260.
- Brick, J.M. 2013. "Unit Nonresponse and Weighting Adjustments: A Critical Review." *Journal of Official Statistics* 29: 329–353. Doi: <https://doi.org/10.2478/jos-2013-0026>.
- Chun, Y. 2009. "Nonparticipation of the 12th graders in the National Assessment of Educational Progress: Understanding Determinants of Nonresponse and Assessing the Impact on NAEP Estimates of Nonresponse Bias According to Propensity Models." University of Maryland. Available at: <http://hdl.handle.net/1903/9916> (accessed June 4, 2017).
- Curtin, R., S. Presser, and E. Singer. 2000. "The Effects of Response Rate Changes on the Index of Consumer Sentiment." *Public Opinion Quarterly* 64: 413–428. Doi: <https://doi.org/10.1086/318638>.
- Davern, M. 2013. "Nonresponse Rates are a Problematic Indicator of Nonresponse Bias in Survey Research." *Health Services Research* 48: 905–912. Doi: <http://dx.doi.org/10.1111/1475-6773.12070>.

- Groves, R. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70: 646–675. Available at: <http://www.jstor.org/stable/4124220> (accessed July 9, 2017).
- Groves, R.M., J.M. Brick, M. Couper, W. Kalsbeek, B. Harris-Kojetin, F. Kreuter, B.E. Pennell, T. Raghunathan, B. Schouten, T. Smith, and R. Tourangeau. 2008. "Issues Facing the Field: Alternative Practical Measures of Representativeness of Survey Respondent Pools." *Survey Practice* 1: 14–22.
- Groves, R.M. and M. Couper. 1998. *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons.
- Groves, R.M., M.P. Couper, S. Presser, E. Singer, R. Tourangeau, G.P. Acosta, and L. Nelson. 2006. "Experiments in Producing Nonresponse Bias." *Public Opinion Quarterly* 70: 720–736. Available at: <http://www.jstor.org/stable/4124223> (accessed July 9, 2017).
- Groves, R.M. and S.G. Heeringa. 2006. "Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 169: 439–457. Doi: <http://dx.doi.org/10.1111/j.1467-985X.2006.00423.x>.
- Groves, R. and E. Peytcheva. 2008. "The Impact of Nonresponse Rates on Nonresponse Bias." *Public Opinion Quarterly* 72: 167–189. Doi: <https://doi.org/10.1093/poq/nfn011>.
- Groves, R.M., E. Singer, and A. Corning. 2000. "Leverage-Saliency Theory of Survey Participation: Description and an Illustration." *Public Opinion Quarterly* 64: 299–308. Available at: <http://www.jstor.org/stable/3078721> (accessed July 9, 2017).
- Hansen, M.H. and W.N. Hurwitz. 1946. "The Problem of Nonresponse in Sample Surveys." *Journal of the American Statistical Association* 41: 517–529. Doi: <http://dx.doi.org/10.1080/01621459.1946.10501894>.
- Keeter, S., C. Kennedy, M. Dimock, J. Best, and P. Craighill. 2006. "Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey." *Public Opinion Quarterly* 70: 759–779. Available at: <http://www.jstor.org/stable/4124225> (accessed July 9, 2017).
- Keeter, S., C. Miller, A. Kohut, R. Groves, and S. Presser. 2000. "Consequences of Reducing Nonresponse in a Large National Telephone Survey." *Public Opinion Quarterly* 64: 125–148. Doi: <https://doi.org/10.1086/317759>.
- Little, R.J.A. and S. Vartivarian. 2005. "Does Weighting for Nonresponse Increase the Variance of Survey Means?" *Survey Methodology* 31: 161–168.
- Luiten, A. and B. Schouten. 2013. "Tailored Fieldwork Design to Increase Representative Household Survey Response: An Experiment in the Survey of Consumer Satisfaction." *Journal of the Royal Statistical Society A* 176: 169–189. Doi: <http://dx.doi.org/10.1111/j.1467-985X.2012.01080.x>.
- Lundquist, P. and C.-E. Särndal. 2013. "Aspects of Responsive Design with Applications to the Swedish Living Conditions Survey." *Journal of Official Statistics* 29: 557–582. Doi: <https://doi.org/10.2478/jos-2013-0040>.
- McQuillan, G.M., T.M. Ezzati-Rice, A.B. Siller, W. Visscher, and P. Hurley. 1994. "Risk Behavior and Correlates of Risk for HIV Infection in the Dallas County Household HIV

- Survey.” *American Journal of Public Health* 84: 747–753. Doi: <http://dx.doi.org/10.2105/AJPH.84.5.747>.
- Mercer, A., A. Caporaso, D. Cantor, and R. Townsend. 2015. “How Much Gets You How Much? Monetary Incentives and Response Rates in Household Surveys.” *Public Opinion Quarterly* 79: 105–129. Doi: <https://doi.org/10.1093/poq/nfu059>.
- Merkle, D. and M. Edelman. 2002. “Nonresponse in Exit Polls: a Comprehensive Analysis.” In *Survey Nonresponse*, edited by R. Groves, D. Dillman, J. Eltinge, and R. Little, 243–258. New York: John Wiley.
- Messer, B.L. and D.A. Dillman. 2011. “Surveying the General Public Over the Internet Using Address-Based Sampling and Mail Contact Procedures.” *Public Opinion Quarterly* 75: 429–457. Doi: <https://doi.org/10.1093/poq/nfr021>.
- Moore, J., G. Durrant, and P. Smith. 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, A*. Doi: <http://dx.doi.org/10.1111/rssa.12256>.
- Särndal, C.-E. 2011. “The 2010 Morris Hansen Lecture: Dealing with Survey Nonresponse in Data Collection, in Estimation.” *Journal of Official Statistics* 27: 1–21.
- Särndal, C.-E., K. Lumiste, and I. Traat. 2016. “Reducing the Response Imbalance: Is the Accuracy of the Survey Estimates Improved?” *Survey Methodology* 42: 219–238.
- Särndal, C.E. and S. Lundström. 2005. *Estimation in Surveys with Nonresponse*. New York: John Wiley & Sons.
- Särndal, C.-E. and P. Lundquist. 2014. “Accuracy in Estimation with Nonresponse: a Function of the Degree of Imbalance and Degree of Explanation.” *Journal of Survey Statistics and Methodology* 2: 361–387. Doi: <https://doi.org/10.1093/jssam/smu014>.
- Singer, E. 2002. “Use of Incentives to Reduce Nonresponse in Household Surveys.” In *Survey Nonresponse*, edited by R. Groves, D. Dillman, J. Eltinge, and R. Little, 163–177. Wiley.
- Schouten, B., F. Cobben, and J. Bethlehem. 2009. “Indicators for the Representativeness of Survey Response.” *Survey Methodology* 35: 101–113.
- Schouten, B., F. Cobben, P. Lundquist, and J. Wagner. 2016. “Does More Balanced Survey Response Imply Less Non-Response bias?” *Journal of the Royal Statistical Society, A* 179: 727–748. Doi: <http://dx.doi.org/10.1111/rssa.12152>.
- Silver, N. 2014. “Is the Polling Industry in Stasis or in Crisis?” Available at: <https://fivethirtyeight.com/features/is-the-polling-industry-in-stasis-or-in-crisis/> (accessed February 22, 2016).
- Sturgis, P., J. Williams, I. Brunton-Smith, and J. Moore. 2017. “Fieldwork Effort, Response Rate, and the Distribution of Survey Outcomes: A Multilevel Meta-Analysis.” *Public Opinion Quarterly* 81: 523–542. Doi: <https://doi.org/10.1093/poq/nfw055>.
- Tourangeau, R., J.M. Brick, S. Lohr, and J. Li. 2017. “Adaptive and Responsive Survey Designs: a Review and Assessment.” *Journal of the Royal Statistical Society A* 180: 203–223. Doi: <http://dx.doi.org/10.1111/rssa.12186>.
- Tourangeau, R., R.M. Groves, and C. Redline. 2010. “Sensitive Topics and Reluctant Respondents: Demonstrating a Link Between Nonresponse Bias and Measurement

- Error.” *Public Opinion Quarterly* 74: 413–432. Doi: <https://doi.org/10.1093/poq/nfq004>.
- Trussell, N. and P. Lavrakas. 2004. “The Influence of Incremental Increases in Token Cash Incentives on Mail Survey Response.” *Public Opinion Quarterly* 68: 349–367. Doi: <https://doi.org/10.1093/poq/nfh022>.
- U.S. Census Bureau. 2014. “American Community Survey Design and Methodology.” Available at: <https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html> (accessed February 19, 2016).

Received May 2016

Revised June 2017

Accepted June 2017