1

Missing Data Techniques and Low Response Rates

The Role of Systematic Nonresponse Parameters

Daniel A. Newman

This chapter attempts to debunk two popular misconceptions (or legends) about missing data: Legend #1, low response rates will necessarily invalidate study results; and Legend #2, listwise and pairwise deletion are adequate default techniques, compared with state-of-the-art (maximum likelihood) missing data techniques. After reviewing general missingness mechanisms (i.e., MCAR, MAR, MNAR), the relevance of response rates and missing data techniques is shown to depend critically on the magnitude of two systematic nonresponse parameters (or SNPs: labeled d_{miss} and f^2_{miss}). Response rates impact external validity only when these SNPs are large. Listwise and pairwise deletions are appropriate only when these SNPs are very small. I emphasize (a) the need to explicitly identify and empirically estimate SNPs, (b) the connection of SNPs to the theoretical model (and specific constructs) being studied, (c) the use of SNPs in sensitivity analysis to determine bias due to response rates, and (d) the use of SNPs to establish inferiority of listwise and pairwise deletion to maximum likelihood and multiple imputation approaches. Finally, key applications of missing data techniques are discussed, including longitudinal modeling, within-group agreement estimation, meta-analytic corrections, social network analysis, and moderated regression.

Organization of the Chapter

The material that follows is organized into six sections. First, I distinguish three *levels* of missing data (item level, scale level, and survey level), two *problems* caused by missing data (bias and low statistical power), and three *mechanisms* of missing data (MCAR, MAR, and MNAR). Second, I present a fundamental principle of missing data analysis ("use all the available information") and review four missing data techniques (listwise deletion, pairwise deletion, maximum likelihood, and multiple imputation) in light of this fundamental principle. Third, I introduce two systematic nonresponse parameters (SNPs: d_{miss} and f^2_{miss}) and illustrate how response rate bias depends entirely on the interaction between SNPs and response rates, rather than on response rates alone. Fourth, I present a theoretical model of survey nonresponse, highlighting how SNPs and response rate bias vary with the substantive constructs being studied. Fifth, I use the aforementioned information to redress two popular legends about missing data. Sixth, I review several prominent data-analytic scenarios for which the choice of missing data technique is likely to make a big difference in one's results.

Levels, Problems, and Mechanisms of Missing Data

Missing data is defined herein as a statistical difficulty (i.e., a partially incomplete data matrix) resulting from the decision by one or more sampled individuals to not respond to a survey or survey item. The term *survey nonresponse* refers to the same phenomenon, at the level of the individual nonrespondent. Missing data is a problem from the perspective of the data analyst, whereas survey nonresponse is an individual decision made by the potential survey participant. Although nonresponse decisions may vary in how intentional they are (e.g., forgetting about the survey vs. discarding the survey deliberately), the above definition of survey nonresponse assumes that a potential respondent saw the survey invitation and made a de facto choice whether to complete the measures.

Three Levels of Missing Data

The missing data concept subsumes three levels of nonresponse: (a) *item-level nonresponse* (i.e., leaving a few items blank), (b) *scale-level nonresponse* (i.e., omitting answers for an entire scale or entire construct), and (c) unit- or *survey-level nonresponse* (i.e., failure by an individual to return the entire survey). The *response rate*, which is a ratio of the total number of completed surveys to the number of solicited surveys, is an aggregate index of survey-level nonresponse.

Two Problems Caused by Missing Data (External Validity and Statistical Power)

There are two primary problems that can be caused by low response rates. The first problem is poor external validity (i.e., response rate bias), which in this case means that the results obtained from a subsample of individuals who filled out the survey may not be identical to results that would have been obtained under 100% response rates. In other words, a respondents-based estimate (e.g., respondents-based correlation: r_{resp}) can sometimes be a biased (over- or underestimated) representation of the complete-data estimate (e.g., complete-data correlation: $r_{complete}$).

The second problem caused by missing data is low statistical power, which means that—even when there is a true nonzero effect in the population—the sample of respondents is too small to yield a statistically significant result (i.e., Type II error of inference). I clarify that *power is a function of the sample size, and not a direct function of response rate*. For example, attempting to sample 1,000 employees and getting a 15% response rate yields more statistical power (N = 150) than attempting to sample 200 employees and getting a 60% response (N = 120). After controlling for sample size, response rates have negligible effects on power.

Missingness Mechanisms (MCAR, MAR, and MNAR)

Data can be missing randomly or systematically (nonrandomly). Rubin (1976) developed a typology that has been used to describe three, distinct missing data mechanisms (see Little & Rubin, 1987):

MCAR (***missing completely at random***)—the probability that a variable value is missing does not depend on the observed data values or on the missing data values. The missingness pattern results from a completely random process, such as flipping a coin or rolling a die.

MAR (***missing at random***)—the probability that a variable value is missing partly depends on other data that are observed in the data set but does not depend on any of the values that are missing.

MNAR (***missing not at random***)—the probability that a variable value is missing depends on the missing data values themselves.

Of the three missingness mechanisms, only MCAR would be considered "random" in the usual sense, whereas MAR and MNAR would be considered "systematic" missingness (note the unusual label, *missing at random* [MAR], to describe a particular type of *systematic* missingness). For a helpful example of the MAR and MNAR mechanisms, consider two variables X and Y, where some of the data on variable Y are missing (Schafer & Graham, 2002). Missing data would be MAR if the probability of missingness on Y is related to the observed values of X but unrelated to the values of Y after X is controlled (i.e., one can predict whether Y is missing based on the observed values of X). The data would be MNAR if the probability of missingness on Y is related to the values of Y itself (i.e., related to the missing values of Y). Note that in practice, it is usually considered impossible to determine whether missing data are MNAR, because this would require a comparison of the observed Y values to the missing Y values, and the researcher does not have access to the missing Y values.

Why do missing data mechanisms matter? *Missing data mechanisms determine the nature and magnitude of missing data bias and imprecision* (see Table 1.1). In general, systematic missingness will lead to greater bias in parameter estimates (e.g., correlations and regression weights) than will completely random missingness. That is, MCAR is harmless in that it does not bias the means, standard deviations, and estimated relationships between variables. Systematic missingness (MAR or MNAR), on the other hand, will often bias parameter estimates.

TABLE 1.1 Parameter Bias and Statistical Power Problems of Common Missing Data Techniques

Missing Data Technique	Missingness Mechanism: MCAR	MAR	MNAR
Listwise deletion	Unbiased, low power	Biased, low power	Biased, low power
Pairwise deletion	Unbiased, inaccurate power	Biased, inaccurate power	Biased, inaccurate power
Maximum likelihood	**Unbiased, accurate power**	**Unbiased, accurate power**	Biased, accurate power
Multiple imputation	**Unbiased, accurate power**	**Unbiased, accurate power**	Biased, accurate power

Note. Recommended techniques are in boldface.

Missing Data Treatments

A Fundamental Principle of Missing Data Analysis

Across missing data conditions, the best data-analytic methods for dealing with missing data follow a simple yet fundamental principle: *use all of the available data.* This principle characterizes all of the recommended missing data techniques shown in Table 1.2. However, the principle is not found in many of the more commonly applied missing data techniques, such as listwise and pairwise deletion.

In general, *item-level nonresponse* can be redressed through $\text{mean}_{\text{item}}$ imputation (Roth, Switzer, & Switzer, 1999), meaning that a researcher can average across the subset of scale items with available responses to calculate a scale score. This approach works especially well when scale items are essentially parallel. Unfortunately, there is a relatively common practice of setting an arbitrary threshold number of items that must be completed in order to calculate a scale score (e.g., if 4 or more items from an 8-item scale are complete, then those items can be averaged into a scale score; otherwise, set the respondent's scale score to "missing"). Setting such an arbitrary threshold violates the fundamental principle of missing data analysis, because it *throws away real data* from the few items that were completed. Dropping an entire scale from analysis simply because some of its items were omitted will typically produce worse biases, in comparison

TABLE 1.2 Three Levels of Missing Data and Their Corresponding Missing Data Techniques

Level of Missing Data	Recommended Missing Data Technique	Favorable Condition for Technique
Item-level	Use mean$_{item}$ imputation.	Essentially parallel items
Scale-level	Use maximum likelihood (ML) or multiple imputation (MI).	Probability of missingness is correlated with observed variables (i.e., MAR mechanism)
Survey-level (i.e., person-level, as reflected in overall response rate)	Use systematic nonresponse parameters (d_{miss} and f^2_{miss}).	Data are available from previous studies that compare respondents to nonrespondents on the constructs of interest (i.e., local d_{miss} and f^2_{miss} can be estimated)

to assuming that the few completed items appropriately reflect the scale score.

Next, *scale-level nonresponse* can be treated through maximum likelihood or multiple imputation techniques (ML and MI techniques; Dempster, Laird, & Rubin, 1977; Enders, 2001; Schafer, 1997), in which a researcher estimates the parameters of interest (e.g., correlations, regression weights) using a likelihood function (or alternatively using a Bayesian sampling distribution) based on observed data from all of the measured variables. (ML and MI will be discussed in more detail later.) In other words, if a respondent omits an entire scale, then using ML or MI techniques to recover the parameter estimates will typically produce less bias than using ad hoc techniques, such as listwise deletion, pairwise deletion, and single imputation (Newman, 2003). ML and MI techniques work especially well when missing data are systematically missing according to the common MAR mechanism.

Finally, *survey-level nonresponse*—in which the entire survey is not returned—can be addressed using nonlocal meta-analytic estimates that describe respondent-nonrespondent differences on the constructs of interest. These respondent-nonrespondent differences are captured by two SNPs, labeled d_{miss} and f^2_{miss}. The use of SNPs to address survey-level missingness (i.e., low response rates) is a primary focus of this chapter. SNPs are particularly useful for addressing the response rate issue, because some of the more-developed missing data approaches (e.g., ML and MI) are not currently capable of addressing survey-level (i.e., person-level) nonresponse, in which the data set contains absolutely no data on the nonrespondents. For handling *survey-level nonresponse* (i.e., low response rates), SNP methods reflect an attempt to *use all of the available data* (including nonlocal data on respondent-nonrespondent differences).

Missing Data Techniques (Listwise and Pairwise Deletion, ML, and MI)

Table 1.1 summarizes relationships between the missingness mechanisms (MCAR, MAR, MNAR) and parameter estimation bias. As seen in Table 1.1, the problems attributable to different mechanisms of missingness (i.e., missing data bias and low statistical power) depend on the missing data technique that is used. Four missing data techniques are covered here: listwise deletion, pairwise deletion, maximum likelihood (ML), and multiple imputation (MI). Listwise deletion involves analyzing data exclusively from individuals who provide complete data for all of the variables surveyed (i.e., partial respondents' data are discarded). Pairwise deletion involves estimating correlations between two variables (X and Y) using all of the respondents who reported data for both X and Y (i.e., and ignoring data from respondents who did not report on both X and Y). ML and MI approaches both involve estimating the relevant parameters (e.g., correlations, regression weights) by using *all of the available data on all of the variables from all of the respondents*, regardless of partial data incompleteness. For example, ML and MI techniques estimate the correlation between two variables (X and Y) while accounting for the linear dependencies of X's and Y's missingness on the observed values of X, Y, Z, Q, and all other variables in the observed data set (see Enders, 2001, for a lengthier description of ML and MI techniques).

As seen in Table 1.1, listwise and pairwise deletion are unbiased only when data are MCAR, whereas ML and MI techniques are unbiased under both MCAR and MAR conditions. This is why ML and MI approaches have been advocated as generally superior to listwise and pairwise deletion (Graham, Cumsille, & Elek-Fiske, 2003; Little & Rubin, 2002; Schafer & Graham, 2002). ML and MI techniques (e.g., FIML, EM algorithm, and multiple imputation; now available in most statistical packages) perform well under MAR because they use all the available data to estimate parameters, whereas ad hoc techniques (e.g., listwise deletion) discard or ignore some of the available data.

As for statistical power, I note that missing data reduce power regardless of the missingness mechanism. Some missing data techniques, however, are far worse than others when it comes to power. Listwise deletion typically will be far less powerful than other missing data techniques (Table 1.1), because listwise deletion discards all data from partial respondents, thereby greatly reducing sample size. Pairwise deletion, in contrast, suffers from its inability to account for the differential sample sizes across correlation estimates (Marsh, 1998). Although some correlations are based on more data than others (i.e., some correlations have more power than others), pairwise deletion uses a single sample size to estimate all the standard errors, providing overestimates of power for some parameters and underestimates for others (Newman, 2003). This problem is avoided under full information maximum likelihood (FIML) and MI approaches, which use the more appropriate standard errors for each estimate (and therefore give accurate estimates of statistical power).

Finally, there are currently few if any available missing data techniques that perform well under the common scenario of MNAR missingness (see Collins, Schafer, & Kam, 2002; Newman, 2003). This is the context within which SNPs are introduced, as a way of characterizing respondent-nonrespondent differences, which can be used to better understand and deal with response rate bias (resulting from the MNAR mechanism).

Systematic Nonresponse Parameters (d_{miss} and f^2_{miss})

In this chapter, I propose a way to index the nature and magnitude of missingness mechanisms. It is suggested that, for any given variable that a researcher is interested in studying, SNPs can be estimated that characterize the differences between respondents and nonrespondents on the constructs of interest. Two such nonresponse parameters are the focus here: d_{miss} and f^2_{miss}.

The parameter d_{miss} is defined as the standardized respondent-nonrespondent mean difference on a variable

$$[\text{i.e., } d_{miss} = \left(\bar{X}_{non} - \bar{X}_{resp}\right) / s_{pooled}]$$

(Newman & Sin, in press). In other words, if individuals with low job satisfaction are less likely to respond to a job satisfaction survey (Rogelberg, Conway, Sederburg, Spitzmuller, Aziz, & Knight, 2003), then d_{miss} will be negative. A nonzero d_{miss} suggests that missing data on a job satisfaction survey are missing systematically (MNAR), whereas $d_{miss} = 0$ suggests that the missingness mechanism is completely random (MCAR). Also, when d_{miss} is large and negative, paying attention to the respondents only will lead to an upward bias in estimates of mean job satisfaction, where the bias increases in magnitude as response rates drop. So the SNP d_{miss} is a useful way of describing the extent to which missingness is systematic (not random) for a particular variable, and it also *determines the extent to which a parameter estimate (in this case, the mean) is biased by low response rates.*

The relationships among d_{miss}, response rate, and missing data bias in estimated means are illustrated in Figure 1.1a. In Figure 1.1a, we see that—when d_{miss} is negative—the respondent-based mean is an *overestimate* of the complete-data mean. Further, this positive bias increases as response rates fall (e.g., at $d_{miss} = -.4$, the mean is overestimated by 11.8% when the response rate is 10%). Importantly, when $d_{miss} = 0$, there is no missing data bias in the mean, regardless of the response rate. That is, *low response rates only threaten external validity (i.e., lead to missing data bias) to the extent the SNP (d_{miss}) is large.*

Next, Figure 1.1b shows how the relationship between bias and response rate for the standard deviation (SD) also depends entirely on d_{miss}. There is a negative response rate bias in SD (i.e., an underestimation of SD) that increases nonlinearly as response rates drop. At $d_{miss} = -.4$ and response rate = 10%, the SD is underestimated by 15.3% (see Newman & Sin, in press, for derivation of formulae that produced Figures 1.1a and 1.1b).

A second systematic nonresponse parameter, f^2_{miss}, is defined as the standardized respondent-nonrespondent difference in the *relationship* between two variables, *X* and *Y*. This parameter can be thought of as an *effect size* for a categorical moderator of *response status* (see Appendix for derivation). When "$f^2_{miss(+)}$" is large, it means the correlation between *X* and *Y* among *nonrespondents* is larger than the *XY* correlation for *respondents* (and when "$f^2_{miss(-)}$" is large, the

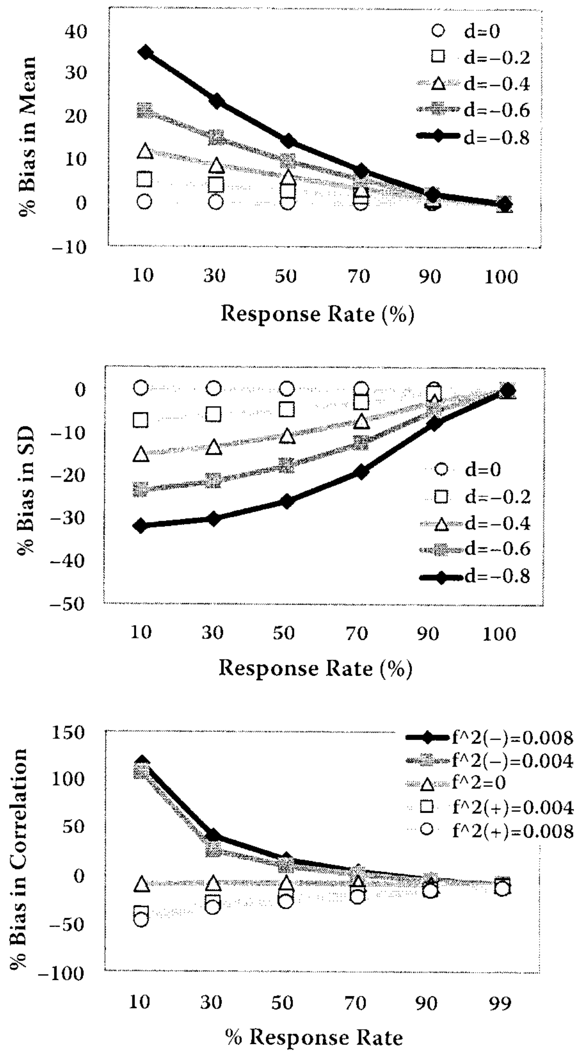

Figure 1.1 (a) Response rate bias in the mean. (b) Response rate bias in the standard deviation. (c) Response rate biases in the correlation. *Note.* Mean bias evaluated at $\bar{X}_{resp} = 4$; correlation bias at $r_{resp} = .3$; $d_{miss\_x} =$ and $d_{miss\_y} = -.4$.

nonrespondent correlation is *smaller* than the respondent correlation). In Figure 1.1c, we see that at $f^2_{miss(+)} = .004$, $d_{miss\_x} = d_{miss\_y} = -.4$, and response rate = 10%, the XY correlation is underestimated by 41.6% due to missing data.

As can also be observed in Figure 1.1 (panels a, b, and c), there is no magical response rate below which an observed mean, standard deviation, or correlation becomes automatically invalid. Further, for a given, arbitrary amount of "tolerable" bias (say 10%), the corresponding response rate that produces this amount of bias depends entirely on the SNPs (d_{miss} and f^2_{miss}).

To help the reader in gauging the representativeness of the range of values presented in Figure 1.1, we summarize empirical estimates of SNPs (d_{miss} and f^2_{miss}) as found in previous studies of nonrespondents (see Table 1.3). The estimates in Table 1.3 are taken from nonrespondent studies that employed two types of designs: (a) follow-up studies that tracked down nonrespondents after they were observed to not respond (e.g., Rogelberg et al., 2003), and (b) studies based on *self-reported* response behavior to past surveys and intentions to respond to future surveys (Rogelberg et al., 2000). As shown in Table 1.3, estimates of d_{miss} that are based on respondent *self-reported* intentions toward future survey responding (as well as self-reported retrospective histories of responding) offer large overestimates of d_{miss} when compared to the d_{miss} values obtained from observing actual response behavior (e.g., for the construct "satisfaction with management": $d_{miss} = -.59$ for self-reported survey response, but $d_{miss} = -.15$ for actual, observed response behavior). The largest d_{miss} estimate for actual response behavior involved the construct of "procedural justice" ($d_{miss} = -.44$), suggesting that employees are much less likely to respond to a survey solicited by a company they believe has treated them unfairly. The important message of Table 1.3 is that systematic nonresponse parameters (d_{miss} and f^2_{miss}) vary depending on the psychological constructs that are being studied.

Theory of Survey Nonresponse

Although survey nonresponse is often thought of as a *methodological* problem, it can also be connected to substantive, theoretical concepts. The individual decision to respond (or not respond) to a survey is a behavioral construct, which results from underlying attitudes, motives, dispositions, and norms. As with research on absenteeism (Martocchio & Harrison, 1993), studies of nonresponse behavior face the difficulty of modeling what individuals are *not* doing, rather than what they are actually doing. Rogelberg et al. (2000) described

TABLE 1.3 Empirical Estimates of Systematic Nonresponse Parameters (d_{miss} and f^2_{miss})

Construct	d_{miss}[a]	d_{miss}[b]	f^2_{miss}[a] Satisfaction (with Management)	f^2_{miss}[a] Turnover Intentions	f^2_{miss}[a] Agreeable
Organizational commitment	—	−.59 (183)	—	—	—
Job satisfaction	—	−.62 (182)	—	—	—
Satisfaction (work)	—	−.68 (183)	—	—	—
Satisfaction (pay)	—	−.13 (183)	—	—	—
Satisfaction (promotion)	—	−.24 (183)	—	—	—
Satisfaction (management/ supervision)	−.15 (399)	−.59 (180)	0	—	—
Turnover intentions	.13 (399)	.60 (181)	.0028$^{(+)}$	0	
Agreeableness	−.35 (399)	—	.0027$^{(-)}$	.0042$^{(-)}$	0
Conscientiousness	−.38 (399)	—	.0014$^{(+)}$	.0074$^{(-)}$	.0096$^{(+)}$
Procedural justice	−.44 (608)	—	—	—	—
Perceived organizational support	−.13 (608)	—	—	—	—

Note. All estimates uncorrected. Corresponding sample sizes (N) in parentheses.
[a]Based on actual response behavior (Rogelberg et al., 2001; Spitzmuller et al., 2006); estimates compare respondents to pooled active-intentional and passive-unintentional nonrespondents.
[b]Based on self-rated response intentions and retrospective response reports only (Rogelberg et al., 2000).

response to at-work surveys as an organizational citizenship behavior, and research consistent with this idea shows that nonrespondents have lower average job satisfaction, organizational commitment, conscientiousness, agreeableness, and intentions to remain with the company (see Table 1.3).

In developing a *Theoretical Model of Survey Nonresponse*, I focus on predictors at multiple levels of analysis. That is, individual nonresponse behavior may theoretically result from individual attributes (e.g., dissatisfaction), group attributes (e.g., group trust and

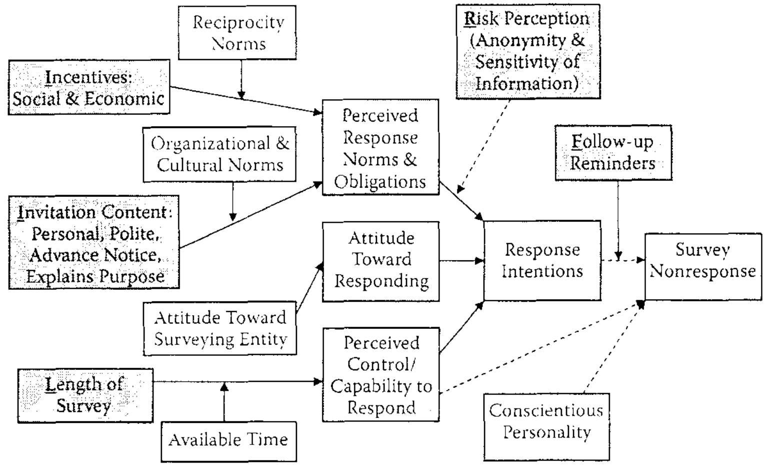

Figure 1.2 Theoretical model of survey nonresponse. *Note.* Dotted lines represent negative relationships. Light gray boxes are Theory of Planned Behavior Constructs. Dark gray boxes are Methodological Choices under the researcher's control.

support), and organizational and cultural attributes (e.g., company norms for survey participation, or Dillman's [1978] cultural norms of willingness to do a small favor for a stranger who asks you to fill out a survey). According to the Theory of Planned Behavior (Ajzen, 1988), a behavior such as survey nonresponse will be predicted by (a) favorable or disfavorable attitudes toward responding to the survey at hand, (b) subjective norms reflecting whether important referent others would likely respond to the survey, and (c) perceived confidence in one's capability to respond to the survey. These three antecedents (attitudes, norms, and perceived control) influence survey response behavior through a causal mechanism of survey response intentions (see Figure 1.2; cf. Rogelberg et al., 2000). Onto this Theory of Planned Behavior model for survey nonresponse, I have overlain several antecedents and moderating conditions, including some proactive steps a researcher can take to increase response rates (see Figure 1.2).

Past research has highlighted several design features that help in securing higher response rates (see dark gray boxes in Figure 1.2; largely consistent with Dillman, 1978; Fox, Crask, & Kim, 1988; Roth & BeVier, 1998; Yammarino, Skinner, & Childers, 1991; Yu & Cooper, 1983). This research shows survey response rates are higher when participants are given advance notice, the survey is personalized, follow-up reminders are sent, and monetary incentives are offered. However, not all these techniques are equally effective. Below, I briefly summarize distinctions among techniques and speculate on their theoretical mechanisms.

In Roth and BeVier's (1998) integrative meta-analysis, response rates were most strongly affected by survey *invitation* factors (i.e., advance notice, more personalized [nonmailed] survey distribution, and distribution within one's own company [rather than across many companies]). *Follow-up* reminders (e.g., postcards) had a smaller but still important unique effect on response rates. I conjecture that follow-up survey reminders offer additional opportunities for response intentions to be converted into actual response behavior (Figure 1.2). That is, follow-up reminders do not directly act to generate response intentions—rather they simply provide more chances to manifest these intentions. (The importance of distinguishing response intentions from actual response behavior is illustrated in the first two columns of Table 1.3.)

Contrary to popular belief, survey *length* had only a meager effect on response rates (Roth & BeVier, 1998). I explain this by suggesting that survey length is moderated by individual differences in available time to complete surveys (Figure 1.2). Also, survey length may have a nonlinear association with response intentions, such that potential respondents lose interest after about 4 pages (Yammarino et al., 1991)—although the exact threshold for length is unknown. Monetary *incentives* for survey participation have their basis in exchange theory (Foa & Foa, 1980). Contrary to previous research (Yammarino et al., 1991), Roth and BeVier (1998) showed that monetary incentives may have virtually no effect on response rates to organizational surveys. I suggest that monetary incentives rely on reciprocity norms (Gouldner, 1960) in order to change response intentions (Figure 1.2) and thus may not uniformly result in more responses.

Finally, norms for survey response can be made more salient when participants are placed at *risk*, due to sensitive content of the survey questions or perceived lack of confidentiality. Roth and BeVier (1998) showed that when anonymity is compromised, survey response rates actually *increase* substantially (probably due to fear of reprisal for nonparticipation). Despite the fact that compromising anonymity increases response rates, doing so violates research *ethics* and should therefore be staunchly avoided—survey response *must* be voluntary.

Why is a *Theoretical Model of Survey Nonresponse* (Figure 1.2) important for choosing a missing data strategy or, for that matter, for determining whether a given study's response rate is "too low"? The answer is straightforward: Figure 1.2 gives rise to the SNPs (d_{miss} and f^2_{miss}). Stated differently, *nonresponse behavior is related to many social and psychological variables*. For example, the Figure 1.2 box labeled "attitude toward the surveying entity" includes such concepts as organizational commitment and procedural justice, which have been shown to differ between respondents and nonrespondents (Table 1.3). The reason missing data can bias results of research studies is that *the concepts being studied are related to individual survey response decisions*. If we assume that a single cutoff response rate (e.g., below 20%) applies to all studies, regardless of the constructs being studied, then we have ignored Figure 1.2 and assumed nonresponse is related equally to all constructs. But—as shown in Table 1.3 and Figure 1.1—SNPs (a) vary across constructs being studied and (b) directly determine the extent of nonresponse bias. The above facts are useful in debunking two popular missing data legends, as explained below.

Missing Data Legends

Legend #1: "Low Response Rates Invalidate Results"

As with most legends, the above statement contains a kernel of truth: As response rates decrease, results calculated from respondents only will (a) increasingly suffer from Type II error (low power) and (b) increasingly threaten bias in estimated means, standard deviations, and correlations, *conditional upon the systematic missingness mechanism*. The first myth associated with this kernel of truth is that it is possible to define heuristic response rates (e.g., 20%) below which results automatically fail to generalize. A related, false belief is that all nonresponse is the same—that is, results from a study with 40% response rate are more valid than results from a study with a 15% response rate (without explicitly considering the constructs and magnitude of substantive missingness mechanisms [d_{miss} and f^2_{miss}]).

To debunk this legend, I note that low response rates create no bias when data are MCAR. Likewise, low response rates often create only modest biases when data are missing systematically (MNAR).

Further, these biases depend entirely on the SNPs (d_{miss} and f^2_{miss}; see Figure 1.1). Finally, the issue of low statistical power is really an issue of respondent sample size (N) and not a *response rate* issue per se. As such, power-based criticisms of low-response-rate studies should focus on sample size and not on the response rate itself.

A third, related myth is that response rates are a methodological issue only and are unrelated to the theory being tested. In fact, the response rate problem is an explicit function of the SNPs (d_{miss} and f^2_{miss}) *that correspond to the specific constructs being studied*. Studies on topics like conscientiousness and procedural justice perceptions will be far more affected by response rates, in comparison to studies on satisfaction and turnover intentions (Table 1.3). Nonresponse is a behavioral indicator of one or more latent constructs, and these constructs can be substantive forces in empirical models, to varying degrees.

What Should the State of Practice Be? Rather than relying on the above legend to parse studies into "inadequate" versus "adequate" categories based on their response rates, there may be another—more graduated and empirical—approach. The first step in understanding response rate bias is to identify SNPs germane to the model being tested in a particular study (i.e., d_{miss} and f^2_{miss} for each construct or pair of constructs). Empirical estimates of these nonresponse parameters can be sought in the extant literature, especially from studies using follow-up designs that solicit information from initial nonrespondents (see Rogelberg et al., 2003, for a review of such designs). Ultimately, researchers can meta-analyze d_{miss} and f^2_{miss} across many primary follow-up studies, in order to more precisely estimate the local respondent-nonrespondent differences. With basic information about d_{miss} and f^2_{miss}, the researcher can then conduct a sensitivity analysis to determine the response rate at which inferences break down, given the data set at hand and the SNPs identified.

Take the following example. In a single-sample empirical study, we want to test whether the effect of conscientiousness on turnover intentions is mediated by job satisfaction. The mediation model is conscientiousness (C) → satisfaction (S) → turnover intentions (T). Let the respondent-based correlation matrix be r_{CS} = .20 (Judge, Heller, & Mount, 2002), r_{CT} = −.14 (Zimmerman, 2006), and r_{ST} = −.48 (Tett & Meyer, 1993). Assume the number of respondents for this sample is N = 200, but the response rate is only 10%. Our objective is to calculate a Sobel (1982) test for the indirect effect of conscientiousness on turnover, via satisfaction (i.e., Sobel $z = \beta_{CS}\beta_{ST} / \sqrt{\beta^2_{CS} SE^2_{\beta_{ST}} + \beta^2_{ST} SE^2_{\beta_{CS}}}$). (Note that $\beta_{CS} = r_{CS}$, $SE_{\beta_{CS}} = \sqrt{(1 - r^2_{CS})/(N-2)}$, $\beta_{ST} = (r_{ST} - r_{CT} r_{CS})/(1 - r^2_{CS})$, and $SE_{\beta_{ST}} = \sqrt{(1 - R^2)/[(N-3)(1 - r^2_{CS})]}$.) After running the Sobel test on this sample, we find that Sobel z = 1.97 ($p < .05$), indicating a statistically significant indirect effect of conscientiousness on turnover intentions, mediated by satisfaction.

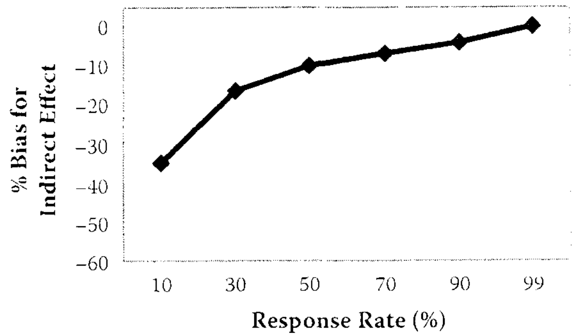

Figure 1.3 Response rate bias in indirect effect ($\beta_{CS}\beta_{ST}$).

Now, suppose a reviewer of the above study offers the following criticism: "With a response rate of only 10%, your observed positive result could very likely be due to missing data bias." Such critical claims are commonplace but are founded on particular assumptions about the underlying pattern of nonresponse parameters, d_{miss} and f^2_{miss}. That is, low response rates can lead to *either* overestimation *or* underestimation of the mediated effect, depending on d_{miss} and f^2_{miss}. The corresponding empirical estimates of d_{miss} for this mediation analysis example can be found in the first column of Table 1.3, and the needed f^2_{miss} parameter estimates can be found in columns 3 and 4 of Table 1.3. Using the above formulae and the formula for $\hat{r}_{xy_{complete}}$ from the Appendix, we get Figure 1.3. What Figure 1.3 shows is that—given the available empirical evidence for d_{miss} and f^2_{miss} involving the constructs of conscientiousness, satisfaction, and turnover intentions (Table 1.3)—at 10% response rates, the indirect effect $\beta_{CS}\beta_{ST}$ is likely to be *underestimated* by 34.9%. If the response rate had been higher, then the observed effect size would have been *larger* (not smaller) due to response rate bias, and N would have also been larger. Therefore, Sobel z would have been *much larger* (not smaller) at higher response rates.

At this point, a caveat is in order—the d_{miss} and f^2_{miss} estimates found in Table 1.3 are too tentative as yet to support a universal call for response rate corrections. Rather, I recommend a more limited use of SNPs, as follows. *When a critic proposes, in the absence of supportive data, that an observed sample effect is positively biased due to low response rates, prior empirical estimates of respondent-nonrespondent differences should be brought to bear on the question.* If prior d_{miss} and f^2_{miss} estimates suggest that the observed effect is unbiased or downwardly biased by nonresponse (see example above), then the low response rate is no longer a legitimate criticism of the study's conclusions. To restate, under the MNAR mechanism (i.e., when d_{miss} or f^2_{miss} is nonzero), the appropriate analytic strategy is to conduct a sensitivity analysis to see whether the obtained result can be explained away by known systematic nonresponse biases (see Table 1.2). This strategy follows the fundamental principle of missing data analysis: *Use all of the available data* (including nonlocal data on respondent-nonrespondent differences).

Legend #2: "When in Doubt, Use Listwise or Pairwise Deletion"

This belief also contains a (very small) kernel of truth: *Listwise and pairwise deletion are unbiased techniques, but only when data are missing completely at random* (MCAR; Table 1.1). The first myth associated with this kernel of truth is simply, "If one does not know the systematic missingness mechanism, it is OK to assume missingness is completely random." This myth equates *ignorance* of systematic biases with *absence* of systematic biases. The myth is debunked by Table 1.3, which shows that commonly studied psychological constructs (e.g., attitudes, personality) are subject to sizable respondent-nonrespondent differences. A second and related myth is, "Missing data techniques that have been most used in the past are the best ones to use in the future." This myth equates the *familiarity/popularity* of a technique with the *accuracy/robustness* of the technique. This (flawed) line of thinking is consistent with a Darwinian model of research methods (only the strongest methods survive over time). Perhaps a truer model of research methods is the

convenience model (only the easiest methods survive). Also, there is a tendency for students and professors to learn which methods are appropriate through imitation of what appears in scholarly journals. (Top journal articles in psychology and management still typically employ listwise and pairwise deletion.) Although this imitation strategy can sometimes enable helpful diffusion of methodological innovations, it also stymies progress by reinforcing the dominant methodological paradigm. There is a further *technological* element of resistance to methodological change, as revealed by the lack of availability of modern missing data techniques in popular statistical software packages (e.g., for many years lagging the development of ML and MI approaches, SPSS software offered only listwise and pairwise deletion options).

A third myth surrounding Legend 2 is that ML and MI approaches are based on shaky assumptions, compared with listwise and pairwise deletion. Although it is true that the ML approach was derived under the assumption of multivariate normality, listwise and pairwise deletion are ad hoc approaches, with no strong statistical basis at all. Departures from multivariate normality do not harm ML estimates as much as they harm estimates from ad hoc approaches (Gold & Bentler, 2000), and corrections are being developed to help the ML approaches become even more robust to nonnormality (see Gold, Bentler, & Kim, 2003). When it comes to comparing ML estimates against listwise and pairwise deletion, it is the deletion techniques that are founded on shaky assumptions (i.e., the MCAR assumption; Table 1.1).

What Should the State of Practice Be? Researchers and editors should begin by understanding that—short of achieving 100% response rates (which may be unethical)—one must choose a missing data technique. Listwise and pairwise deletion are no more safe or natural than ML and MI techniques. Whether one uses listwise, pairwise, or ML techniques, the choice must be based on weighing the pros and cons of each technique. When weighing the pros and cons, ML and MI techniques are always as good as (under MCAR), and usually better than (under MAR), listwise and pairwise deletion, on the criteria of obtaining unbiased parameter estimates and accurate standard errors (Newman, 2003).

When results from an ML or MI missing data technique differ from results obtained through an ad hoc procedure (e.g., listwise deletion, pairwise deletion, mean imputation), then the burden of proof should be placed on the ad hoc technique, not the state-of-the-art technique. That is, ML and MI techniques were designed to provide superior parameter and standard error estimates under a wider range of conditions than listwise and pairwise deletion can handle (summarized in Table 1.1). A biased approach (e.g., listwise or pairwise deletion) should not be used to "double-check" the accuracy of a less-biased approach (ML or MI). Further, maximum likelihood (EM algorithm, FIML) and multiple imputation approaches can now be variously implemented in SAS, SPSS, LISREL, MPlus, and other popular software packages. The number of good excuses for using listwise and pairwise deletion is quickly shrinking.

Applications

Longitudinal Modeling

When sampling the same individuals across time points, a large portion of the missing data comes from attrition, or dropouts. Interestingly, dropouts are usually MAR (i.e., a dropout's missing scores on *X* and *Y* at Time 2 are correlated with her/his observed scores on *X* and *Y* at Time 1). The propensity for MAR mechanisms in longitudinal designs gives ML and MI approaches a major advantage over ad hoc techniques (Table 1.1; Newman, 2003).

Longitudinal designs are also sensitive to compounded missingness. If the response rate is 60% at each wave of measurement, the compounded response rate is Response $\text{Rate}_{\text{compounded}} = (.60)^W = 21.6\%$, where $W = 3$ waves (Newman, 2004). Also, when the response rate rises over consecutive waves (e.g., 40% response rate for first wave, then 80% response rates in subsequent waves), missing data can create a regression-to-the-mean phenomenon, resulting in upward bias in estimated slopes of growth models (Newman, 2004). For longitudinal studies, it is important to continually attempt to sample those who dropped out from earlier waves.

Finally, longitudinal designs hold a special role in the study of SNPs (d_{miss} and f^2_{miss}), because they enable the estimation of respondent-nonrespondent differences (see Rogelberg et al., 2003). That is, one way to estimate d_{miss} and f^2_{miss} is to compare Time 2 respondents versus Time 2 nonrespondents, based on their responses from Time 1.

Within-Group Agreement Estimation

Missing data (MNAR in particular) can lead to overestimation of agreement among members of a group. If an agreement index is used to assess whether group-level aggregation is justified (e.g., $r_{WG(J)}$; James, Demaree, & Wolf, 1984), then missing data can lead to a false conclusion that aggregation is justified, when in fact it is not (Newman & Sin, in press). Further, when group agreement represents a substantive construct (Chan, 1998, e.g., climate strength; Schneider, Salvaggio, & Subirats, 2002), missing data can bias tests of whether agreement predicts other, group-level outcomes. Specifically, tests of dispersion hypotheses are prone to bias whenever there is between-groups variability in response rates (Newman & Sin, in press). One way to address these problems is to conduct a sensitivity analysis, assessing whether response rates and levels of d_{miss} shown in Table 1.3 would lead to large enough changes in estimates that the conclusions of one's study will change. Such sensitivity analyses are reviewed by Newman and Sin (in press).

Meta-analysis

Meta-analyses suffer mainly from two types of missing data problems: (a) unreported artifact information (e.g., scale reliabilities) and (b) publication bias. For missing reliability estimates, Hunter and Schmidt (2004) recommend using artifact distributions based on reported reliability estimates. One important question is, "Are the unreported reliability estimates missing completely at random (MCAR), or are low reliability estimates less likely to be reported than high reliability estimates (MNAR)?" In the latter case, corrections based on observed reliability estimates will lead to overestimation of reliability and therefore undercorrection of the primary study effects. Another common practice is mean imputation from the reported reliabilities (e.g., Harrison, Newman, & Roth, 2006), although substituting a mean for the missing values will artificially reduce the variance of the artifact distribution. Given the above discussion, it seems that a better approach to correcting for unreported artifacts (which are probably MNAR) would involve incorporating SNPs into artifact distributions (e.g., based on a d_{miss} parameter comparing reported versus unreported reliability estimates).

Publication bias, another missing data problem in meta-analysis, is a particular form of MNAR missingness, wherein smaller effects are less likely to be published and thus more likely to be missing from the meta-analytic database (see Lipsey & Wilson, 1993). Methods conceptually similar to the SNP approach advocated in the current chapter have been recommended, in order to estimate what the meta-analytic effect size would have been in the absence of publication bias (Duvall & Tweedie, 2000; Vevea & Woods, 2005).

Social Network Analysis

Several types of social network analyses (e.g., calculating connectedness, indirect friendships, etc., across the entire network) can be extremely sensitive to missing data (see Burt, 1987). In general, social network studies are held to a high standard of data completeness, with journal reviewers regularly requiring response rates of 90% or higher. Costenbader and Valente (2003) and Borgatti, Carley, and Krackhardt (2006) have offered early demonstrations that missing data influence individual network centrality scores in a predictable fashion. However, these analyses only simulate the MCAR pattern, which is potentially problematic because network data missingness is likely systematic, not random (i.e., missingness is associated with the strength of ties and with demographic factors; Burt, 1987).

One reasonable strategy for reducing the negative impact of the missing data on network analyses is to impute respondent-to-nonrespondent ties in place of missing nonrespondent ties (Stork & Richards, 1992). In other words, if person A (a respondent) nominates person B (a nonrespondent) as a friend, then we can assume that person B would have nominated person A as a friend (i.e., friendship symmetry assumption). Consider an example network analysis of 100 individuals, of which only 70 respond to the network survey (individual response rate = 70%). At the network-tie level, there are 100 × 100 = 10,000 potential network ties (e.g., friendships vs. nonfriendships) that could be reported. Getting data from only 70% of the network members results in a network-tie-level response rate of (70 × 70)/10,000 = 49%. Using the strategy advocated above (assuming friendship symmetry) would increase the response rate from 49% up to [10,000 – (30 × 30)]/10,000 = 91%! That is, by *using all the available data* (i.e., by not listwise deleting nonrespondents), we observe a dramatic improvement in the dyadic tie-level response rate. Another approach to modeling respondent *and* nonrespondent ties—which also uses all available data—is exponential random graph modeling (Robins, Pattison, & Woolcock, 2004).

Moderated Regression

When conducting tests for statistical interaction effects (i.e., testing whether the *relationship* between X and Y depends on a third variable, M), listwise deletion increases Type II errors of inference (i.e., failures to detect true effects). Pairwise deletion, on the other hand, leads to elevated Type I error (i.e., concluding there is a moderator effect, when in fact there is not; Dawson & Newman, 2006). ML and MI should be the preferred missing data techniques for testing moderator hypotheses.

Conclusions

This chapter offers three contributions. First, it identifies two SNPs (d_{miss} and f^2_{miss}) that capture the differences between respondents and nonrespondents. Second, it illustrates how response-rate biases in the mean, standard deviation, and correlation depend on an interaction of these SNPs with the response rate. Third, it points out that Type II error (low power) is a function of number of respondents and not the response rate per se. These contributions together demonstrate that low response rates (e.g., below 20%) need not invalidate study results. Rather, the robustness of results to low response rates is an empirical question, driven by d_{miss} and f^2_{miss}.

In theory, survey response is part of a social exchange, wherein the respondent contributes a limited amount of time and effort in exchange for inducements of satisfaction, perceived organizational support, trust, and the promise of anonymity (Figure 1.2). As such, any psychological variable that is related to the nonresponse decision (especially attitudes and personality) will demonstrate a nonzero d_{miss} parameter estimate.

How are d_{miss} and f^2_{miss} parameters estimated? Shafer and Graham (2002) note that it is very difficult to determine whether missing data are missing-not-at-random (MNAR), because this requires actually collecting data from the nonrespondents. Rogelberg et al. (2003) suggest four strategies for gathering data from nonrespondents (e.g., follow-up designs). Using these designs, Rogelberg and colleagues (2000, 2003) show that there exist mean differences between respondents and nonrespondents in terms of job satisfaction, organizational commitment, conscientiousness, and agreeableness (d_{miss} estimates vary from –.1 to –.6, suggesting that nonrespondents are less satisfied and less conscientious than respondents, on average; Table 1.3).

It is the precise sizes of these d_{miss} estimates that determine bias due to low response rates. Researchers should not rely on a heuristic response rate (e.g., below 20%) to automatically invalidate results. Rather, it should be acknowledged that "response rate bias" is an explicit, interactive function of response rate with d_{miss} and f^2_{miss}, *for the constructs at hand.* When d_{miss} and f^2_{miss} are nil, there is no response rate bias. By the same token, when d_{miss} and f^2_{miss} are large, results can be rendered invalid even at higher response rates (e.g., 50%). Response rate bias is not merely a function of response rate—SNPs also play a fundamental role (Figure 1.1). To answer the question, "Is my response rate high enough to support the conclusions of my study?" it will be useful to conduct a sensitivity analysis, using representative SNPs (Table 1.3) and formulae found in Newman and Sin (in press) and the Appendix of this chapter.

Future Research on d_{miss} and f^2_{miss}

At present, relatively little is known about the magnitudes of SNPs (i.e., d_{miss} and f^2_{miss}) for many psychological constructs. As such, our confidence in the biasing effects of low response rates will grow as more follow-up studies are conducted, and mean respondent-nonrespondent differences are cataloged (through meta-analyses of d_{miss} and f^2_{miss}) for a variety of well-known psychological constructs (e.g., Big Five personality traits, affectivity, self-esteem, cognitive ability, job satisfaction, job performance).

It would further be useful to investigate actions that can be taken to potentially alter the sizes of these SNPs. For instance, sending out survey reminders may result in more responses from *passive nonrespondents* (i.e., those who have response intentions but just have not responded yet) but may do little to attract responses from *active nonrespondents* (i.e., those who deliberately choose not to respond; Rogelberg et al., 2003; Spitzmuller et al., 2006). Thus, sending out survey reminders may increase response rates, while simultaneously *increasing* d_{miss}. The diagrams in Figure 1.1 assumed d_{miss} was orthogonal to the response rate, which may or may not hold up under empirical scrutiny.

Missing Data Techniques

A final advantage of considering SNPs is that these parameters indicate the extent to which popular missing data techniques (listwise and pairwise deletion) will result in biased estimates (low external validity). In specific, listwise and pairwise deletion are appropriate only under MCAR (i.e., where $d_{miss} = 0$ and $f^2_{miss} = 0$). As such, the inferiority of listwise and pairwise deletion can be empirically demonstrated by looking at the SNPs. Because missing data are very rarely MCAR (Table 1.3), it can be expected that listwise and pairwise deletion strategies will routinely create nonresponse bias.

What should be done when $d_{miss} \neq 0$ and/or $f^2_{miss} \neq 0$? Low response rates (i.e., survey-level nonresponse) create an MNAR pattern whenever $d_{miss} \neq 0$ or $f^2_{miss} \neq 0$. This MNAR missingness cannot be well addressed through listwise, pairwise, ML, or MI techniques (Table 1.1; see Collins et al., 2001). To deal with low response rates, then, the most appropriate (least biased) missing data treatment will be a sensitivity analysis based on SNPs (see Table 1.2).

References

Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. *Journal of Applied Psychology, 90,* 94–107.

Ajzen, I. (1988). *Attitudes, personality, and behavior.* Homewood, IL: Dorsey Press.

Borgatti, S. P., Carley, K. M., & Krackhardt D. (2006). On the robustness of centrality measures under conditions of imperfect data. *Social Networks, 28,* 124–136.

Burt, R. S. (1987). A note on missing network data in the General Social Survey. *Social Networks, 9,* 63–73.

Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology, 83*, 234–246.

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*, 330–351.

Costenbader, E., & Valente, T. W. (2003). The stability of centrality measures when networks are sampled. *Social Networks, 25*, 283–307.

Dawson, J. F., & Newman, D. A. (2006, May). Pairwise deletion problems with moderated multiple regression. In D. A. Newman (Chair), *Testing interaction effects: Problems and procedures.* Symposium presented at the SIOP Annual Convention, Dallas, TX.

Dempster, A. P., Laird, N. H., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B39*, 1–38.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method.* New York: Wiley.

Duvall, S., & Tweedie, R. (2000). Trim and fill: A simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*, 276–284.

Enders, C. K. (2001). A primer on maximum likelihood algorithms for use with missing data. *Structural Equation Modeling, 8*, 128–141.

Foa, E. B., & Foa, U. G. (1980). Resource theory: Interpersonal behavior as exchange. In K. Gergen, M. S. Greenberg, & R. Willis (Eds.), *Social exchange: Advances in theory and research* (pp. 77–94). New York: Plenum Press.

Fox, R. J., Crask, M. R., & Kim, J. (1988). Mail survey response rate: A meta-analysis of selected techniques for inducing response. *Public Opinion Quarterly, 52*, 467–491.

Gold, M. S., & Bentler, P. M. (2000). Treatments of missing data: A Monte Carlo comparison of RBHDI, iterative stochastic regression imputation, and expectation-maximization. *Structural Equation Modeling, 7*, 319–355.

Gold, M. S., Bentler, P. M., & Kim, K. H. (2003). A comparison of maximum-likelihood and asymptotically distribution-free methods of treating incomplete nonnormal data. *Structural Equation Modeling, 10*, 47–79.

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review, 25*, 161–178.

Graham, J. W., Cumsille, P. E., & Elek-Fiske, E. (2003). Methods for handling missing data. In J. A. Schinka & W. F. Velicer (Eds.), *Research methods in psychology* (pp. 87–114). Vol. 2 of *Handbook of psychology* (I. B. Weiner, Editor in Chief). New York: Wiley.

Harrison, D. A., Newman, D. A., & Roth, P. L. (2006). How important are job attitudes? Meta-analytic comparisons of integrative behavioral outcomes and time sequences. *Academy of Management Journal, 49*, 305–325.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Newbury Park: Sage.

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology, 69*, 85–98.

Judge, T. A., Heller, D., & Mount, M. K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology, 87*, 530–541.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48*, 1181–1209.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York: Wiley.

Marsh, H. W. (1998). Pairwise deletion for missing data in structural equation models: Nonpositive definite matrices, parameter estimates, goodness of fit, and adjusted sample sizes. *Structural Equation Modeling, 5*, 22–36.

Martocchio, J. J., & Harrison, D. A. (1993). To be there or not to be there? Questions, theories and methods in absenteeism research. In K. Rowland & G. Ferris (Eds.), *Research in personnel and human resources management* (Vol. 11, pp. 259–329). Greenwich, CT: JAI Press.

Newman, D. A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods, 6*, 328–362.

Newman, D. A. (2004, April). Missing data in longitudinal designs: Enhancing imputation with auxiliary variables. In D. A. Newman & J. L. Farr (Cochairs), *Assumptions and conventions in data analysis.* Symposium presented at the SIOP Annual Convention, Chicago, IL.

Newman, D. A., & Sin, H. P. (2009). How do missing data bias estimates of within-group agreement? Sensitivity of SD_{WG}, CV_{WG}, $r_{WG(J)}$, $r_{WG(J)}{}^*$, and ICC to systematic nonresponse. *Organizational Research Methods.*

Ostroff, C. (1993). Comparing correlations based on individual-level and aggregated data. *Journal of Applied Psychology, 78*, 569–582.

Robins, G., Pattison, P., & Woolcock, J. (2004). Missing data in networks: Exponential random graph (p*) models for networks with nonrespondents. *Social Networks, 26*, 257–283.

Robinson, W. S. (1950). Ecological correlations and the behaviour of individuals. *American Sociological Review, 15*, 351–357.

Rogelberg, S. G., Conway, J. M., Sederburg, M. E., Spitzmuller, C., Aziz, S., & Knight, W. E. (2003). Profiling active and passive nonrespondents to an organizational survey. *Journal of Applied Psychology, 88*, 1104–1114.

Rogelberg, S. G., Luong, A., Sederburg, M. E., & Cristol, D. S. (2000). Employee attitude surveys: Examining the attitudes of noncompliant employees. *Journal of Applied Psychology, 85*, 284–293.

Roth, P. L., & BeVier, C. A. (1998). Response rates in HRM/OB survey research: Norms and correlates, 1990–1994. *Journal of Management, 24*, 97–117.

Roth, P. L., Switzer, F. S., & Switzer, D. M. (1999). Missing data in multiple item scales: A Monte Carlo analysis of missing data techniques. *Organizational Research Methods, 2*, 211–232.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*, 581–592.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* New York: Chapman & Hall.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177.

Schneider, B., Salvaggio, A. N., & Subirats, M. (2002). Climate strength: A new direction for climate research. *Journal of Applied Psychology, 87*, 220–229.

Spitzmuller, C., Glenn, D. M., Barr, C. D., Rogelberg, S. G., & Daniel, P. (2006). "If you treat me right, I reciprocate": Examining the role of exchange in survey response. *Journal of Organizational Behavior, 27*, 19–35.

Stork, D., & Richards, W. D. (1992). Nonrespondents in communication network studies: Problems and possibilities. *Group & Organization Management, 17*, 193–209.

Tett, R. P., & Meyer, J. P. (1993). Job satisfaction, organizational commitment, turnover intention, and turnover: Path analyses based on meta-analytic finding. *Personnel Psychology, 46*, 259–293.

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using *a priori* weight functions. *Psychological Methods, 10*, 428–443.

Yammarino, F. J., Skinner, S. J., & Childers, T. L. (1991). Understanding mail survey response behavior. *Public Opinion Quarterly, 55*, 613–629.

Yu, J., & Cooper, H. (1983). A quantitative review of research design effects on response rates to questionnaires. *Journal of Marketing Research, 20*, 36–44.

Zimmerman, R. D. (2006). *Understanding the impact of personality traits on individuals' turnover decisions.* Unpublished doctoral dissertation, University of Iowa.

Appendix

Derivation of Response Rate Bias for the Correlation (Used to Generate Figure 1.1c)

Beginning with Aguinis, Beaty, Boik, and Pierce's (2005, p. 105) *modified* f^2, I derive the following:

$$f^2 = \frac{pNr_{non}^2 a^2 s_{y_{resp}}^2 + (1-p)Nr_{resp}^2 s_{y_{resp}}^2 - \dfrac{\left[pNr_{non} a s_{y_{resp}} b s_{x_{resp}} + (1-p)Nr_{resp} s_{y_{resp}} s_{x_{resp}}\right]^2}{pNb^2 s_{x_{resp}}^2 + (1-p)Ns_{x_{resp}}^2}}{pNa^2 s_{y_{resp}}^2 (1-r_{non}^2) + (1-p)Ns_{y_{resp}}^2 (1-r_{resp}^2)},$$

where N is the total number of surveys distributed (at response rate = 100%), $p = n_{non}/N$ (i.e., nonresponse rate), $(1 - p) = n_{resp}/N$ (i.e., response rate), $a = s_{y_{non}}/s_{y_{resp}}$ and $b = s_{x_{non}}/s_{x_{resp}}$ (i.e., standard deviation ratios for y and x, modeling variance heterogeneity), and n_{resp} approximates $(n_{resp} - 1)$ and $(n_{resp} - 2)$. Rearranging and then solving for r_{non} via the Quadratic Formula yields the following equation:

$$r_{non} = \frac{pbr_{resp} \pm \sqrt{p^2 b^2 r_{resp}^2 - p\left\{1 + f^2\left[1 + \dfrac{pb^2}{(1-p)}\right]\right\}\left\{\left\{f^2\left[pb^2 + (1-p)\right] + pb^2\right\}r_{resp}^2 - f^2\left[pa^2 + (1-p)\right]\left[1 + \dfrac{pb^2}{(1-p)}\right]\right\}}}{pa\left\{1 + f^2\left[1 + \dfrac{pb^2}{(1-p)}\right]\right\}}$$

The presence of "$\pm$" in the Quadratic Formula suggests that r_{non} can be either *larger* or *smaller* than r_{resp}, for a given level of f^2. As such, the new notation $f_{miss(-)}^2$ means that the nonrespondent correlation (r_{non}) is smaller than the respondent correlation (r_{resp}), whereas $f_{miss(+)}^2$ means that r_{non} is larger than r_{resp}.

Finally, the complete-data individual-level correlation (at 100% response rates) can be estimated as

$$\hat{r}_{xy.complete} = r_{group}\eta_x\eta_y + r_{pooled}\sqrt{(1-\eta_x^2)(1-\eta_y^2)}$$

(see Ostroff, 1993; Robinson, 1950). Substituting alternative expressions for r_{group}, η_x, η_y, and r_{pooled}, the above equation expands to

$$\hat{r}_{xy_{complete}} = \left(\frac{2d_{miss\_x}d_{miss\_y}}{(d_{miss\_x}^2 + d_{miss\_y}^2)} \right) \left(\frac{d_{miss\_x}}{2\sqrt{1 + d_{miss\_x}p(1-p)}} \right) \left(\frac{d_{miss\_y}}{2\sqrt{1 + d_{miss\_y}p(1-p)}} \right)$$

$$+ \sqrt{\left[pr_{non}^2 + (1-p)r_{resp}^2 \right] \left(1 - \frac{d_{miss\_x}^2}{4\left[1 + d_{miss\_x}p(1-p)\right]} \right) \left(1 - \frac{d_{miss\_y}^2}{4\left[1 + d_{miss\_y}p(1-p)\right]} \right)} \quad .$$

2

The Partial Revival of a Dead Horse? Comparing Classical Test Theory and Item Response Theory

Michael J. Zickar and
Alison A. Broadfoot

Advances in psychometric theory over the last 30 years have introduced many new tools and techniques to researchers interested in measuring psychological constructs. The revolution of item response theory (IRT) has raised questions about the relevance of its predecessor, classical test theory (CTT). In fact, some writers have suggested that CTT has been made obsolete by its successor. For example, Rojas Tejada and Lozano Rojas (2005) discussed how recent research has been used to "displace the CTT in favour of the use of Item Response Theory–based models" (p. 370), and Harvey and Hammer (1999) predicted that "IRT-based methods . . . will largely replace CTT-based methods over the coming years" (p. 354). Samejima, in critiquing CTT, describes its "*fatal* deficiency [italics added]," which relates to how CTT models measurement precision (Samejima, 1977, p. 196). Borsboom argues that "few, if any, researchers in psychology conceive of psychological constructs in a way that would justify the use of classical test theory as an appropriate measurement model" (Borsboom, 2005, p. 47). We have heard people dismiss CTT as irrelevant and antiquated, more worthy of history books than contemporary psychometric classes. Often these same individuals treat IRT as a panacea for all psychometric woes. In short, CTT is treated as an old racehorse that is nice to have around, though everyone is expecting it to perish soon. According to this argument, IRT is the new steed that has won a few races and is expected to abolish its predecessor's triumphs. We believe that this urban legend is just plain myth and