POST-STRATIFICATION WEIGHTING IN ORGANIZATIONAL SURVEYS: A CROSS-DISCIPLINARY TUTORIAL

JOHN T. KULAS, DAVID H. ROBINSON, JEFFREY A. SMITH, AND DONALD Z. KELLAR

Post-stratification weighting is a technique used in public opinion polling to minimize discrepancies between population parameters and realized sample characteristics. The current paper provides a weighting tutorial to organizational surveyors who may otherwise be unfamiliar with the rationale behind the practice as well as "when and how to do" such weighting. The primary reasons to weight include: (1) reducing the effect of frame, sampling, and nonresponse bias *in point estimates, and, relatedly, (2) correcting for aggregation error resulting from over- and underrepresentation of constituent groups. We briefly compare and contrast traditions within public opinion and organizational polling contexts and present a hybrid taxonomy of sampling procedures that organizational surveyors may find useful in situating their survey efforts within a methodological framework. Next, we extend the existing HRM literature focused on survey nonresponse to a broader lens concerned with population misrepresentation. It is from this broadened methodological framework that we introduce the practice of* weighting *as a remedial strategy for misrepresentation. We then provide sample weighting algorithms and standard error corrections that can be applied to organizational survey data and make our data and procedures available to individuals who may wish to use our examples as they learn "how to weight." © 2018 Wiley Periodicals, Inc.*

Keywords: organizational survey, quantitative research methodology, research design

The typical response rate in organizational surveying may be reasonably expected to hover around 50 percent (Baruch, 1999; Baruch & Holtom, 2008), with different categories of respondents (Anseel, Lievens, Schollaert, & Choragwicka, 2010; Rogelberg et al., 2003) or different survey topics (Martin, 1994; Tourangeau, Groves, & Redline, 2010) contributing to variability around this estimate. Rogelberg (2006) solicited vendor figures to obtain a higher estimated average response rate of 78 percent for *intra*organization surveys. Regardless of whether the lens is applied within or across organizations, it is apparent that nonresponse is a reasonable expectation in organizational surveying. Rogelberg, Luong, Sederburg, and Cristol (2000) propose several proactive and one post hoc strategy—weighting—for dealing

Correspondence to: John T. Kulas, head handyman, Corporate Mr Fixit, 1046 Voyageur St., St. Cloud, MN 56303, Phone: 651-216-3353, E-mail: kulas@corporatemrfixit.com

Human Resource Management 2018

Published online in Wiley Online Library (wileyonlinelibrary.com).
DOI:10.1002/hrm.21796

with nonresponse. There are not, however, training resources widely available to the organizational surveyor (e.g., when and *how* to do the weighting). This article is intended to be such a resource that guides the organizational surveyor regarding (1) when to weight, (2) how to weight, and (3) further resources if more specific training is desired. We do this via methodological extraction from the public opinion polling literature, integration with the existing HRM surveying literature, and demonstration of the weighting procedure with a data set constructed to represent simplified but typical organizational survey data.

Post-stratification weighting (also known simply as *sample weighting*) alters the relative contribution of individual responses within a data set. With this procedure, some respondents' data is assigned greater relative influence and others' is assigned less. The practice is commonplace in the summary of polling data (e.g., elections and politics; Rivers & Bailey, 2009; prevalence rates of psychological disorders; Kessler et al., 2009; feelings of physical safety; Quine & Morrell, 2008). The practice is, however, largely absent within the HRM domain or published organizational literature. Both Macey (1996) and Kraut (1996b) make brief mention of sample weighting as an option in organizational survey contexts, but aside from similar textbook mentions or casual referrals (see, e.g., Rogelberg et al., 2000, p. 291), there are no illustrations of the application of sample weighting to organizational surveys. This is possibly reflective of a broader lack of attention paid to sampling *methodology* in organizational surveying. For example, the Society for Industrial and Organizational Psychology's Professional Practice Series text on organizational surveys (Kraut, 1996a) consists of 15 chapters—none of which directly address sampling methodology.

Post-stratification weighting (also known simply as sample weighting) alters the relative contribution of individual responses within a data set. With this procedure, some respondents' data is assigned greater relative influence and others' is assigned less.

This void is perhaps, however, at the beginning stages of being addressed within the organizational sciences. A forthcoming commentary in the "treatise → rejoinder" format of *Industrial and Organizational Psychology: Perspectives on Science and Practice* (Landers & Behrend, 2015) echoes many of our same observations regarding sampling methodology. In this article, we explicitly define and address this methodological void and stress the importance of organizational pollster adherence to a broader procedural framework. We do not, however, provide a full background on sampling methodology as developed and evolved within other disciplines (primarily marketing, political science, and statistics). We rather summarize and synthesize sampling concepts and procedures that we deem most relevant for HRM and note how these broader considerations can benefit the organizational surveyor.

Important Sampling Methodology Concepts for the Organizational Surveyor

The orientation of this presentation and tutorial is consistent with that of several survey researchers (including Cook, Heath, & Thompson, 2000; Krosnick, 1999; and Rogelberg & Stanton, 2007) who note that although response rate is a commonly cited criterion for survey quality, it is sample representativeness that is actually much more important. The organizational surveying literature *does* widely acknowledge and investigate response rate (generally referred to as survey nonresponse; see the opening paragraph citations as an example of this focus), but there are other threats to sample representativeness that are largely left unconsidered within the organizational surveying domain. Representativeness can certainly be affected by nonresponse, but it can also be affected by these other factors (discussed further below) that are only identifiable if a broader survey methodological framework is considered.

Methods of Pursuing a Representative Sample

Within public opinion polling applications, there is a very strong emphasis placed on careful adherence to appropriate sampling *procedures* with an end goal of obtaining a desired representative sample. A census is an extreme case of the general sampling spectrum where representativeness is guaranteed (a census refers to a situation where every member of a population is sampled and is not typical in public opinion polling). More commonly, when each member of a population has a known likelihood of being selected (into the targeted sample), the approach is classified as one of several possible *probability sampling* procedures. Under this broad umbrella, the sampling strategy may be characterized as simple random sampling (every possible sample of a certain size has an equal probability of being selected), stratified (random samples within preidentified strata), systematic (a random sample of equally spaced members from the sampling *frame* [described further below]), clustered (a random sample of preidentified clusters of individuals), or a more complex combination of two or more of these procedures. Sampling is typically performed without replacement (e.g., the individual is removed from the eligible sampling pool once he/she has been selected).

Nonprobability sampling procedures refer to situations in which respondents are allowed to opt in or are solicited by the researcher(s) but *a priori* probabilities of inclusion for all population elements are unknown (e.g., the forecasted constituency of the target sample is not reasonably well predicted). Realized sample inclusion here is most appropriately attributable to simple accessibility or availability of the individual. Kish (1965) offers a sardonic metaphor for nonprobability sampling procedures: inferring the quality of "… a basket of grapes by tasting one of them" (p. 18). There are implicit and often unacknowledged conditions that limit the generalizability of information gathered with such sampling procedures, regardless of resulting sample size. Kish wisely notes, for example, that individuals who employ such sampling procedures must be operating under a tacit assumption of uniformly distributed characteristics (e.g., extreme homogeneity [all grapes in Kish's metaphor must be similar to each other in quality if there is to be confidence placed on the sample-to-population inferences made from such a sampling procedure]).

Probability samples are desired. However, deviations from a probability sampling ideal (every member of a population having a known probability of being included in the sample) occur within both organizational and public opinion contexts. Mallett (2006), in fact, makes a strong distinction between probability and nonprobability sampling procedures in market research applications, but concludes that "most samples are a mix of probability and nonprobability components" (p. 160)—most commonly due to nonresponse. Consequently, both public opinion and organizational surveying contexts share common concerns and threats to sample representativeness. The interdisciplinary concepts of primary relevance to organizational surveyors interested in sample representativeness are (1) the previously mentioned issue of nonresponse (not all invited respondents choose to participate), (2) frame undercoverage (the list of invited individuals fails to include some population members), and (3) sampling error (when pulse surveying or engaging in nonprobability sampling procedures). Each of these threats is further described below.

Fundamental Survey Concepts (Shared across Disciplines)

The sampling *frame* is the tangible list of population members from which the target sample is drawn. Ideally, the frame would exactly match the population. In practice, however, the frame may be inaccurate in two ways. First, there may be individuals listed in the frame who are not part of the population. This is referred to as *overcoverage*. Second, there may be individuals in the population who are not listed in the frame. This results in *undercoverage*. Overcoverage should not be considered a serious a problem in either public opinion or organizational surveying contexts, as individuals who are not members of the intended population may be screened out in the process of collecting the sample or summarizing results. Undercoverage is a more difficult problem to deal with, and it holds greater potential to severely damage sample representativeness. This general danger acknowledgment in public opinion polling contexts dates at least to 1936, when over 2 million homes (an extremely large sample) famously responded to a frame characterized by undercoverage, leading to the confident political prediction of Alfred Landon defeating Franklin Roosevelt.

Probability samples are desired. However, deviations from a probability sampling ideal (every member of a population having a known probability of being included in the sample) occur within both organizational and public opinion contexts.

The *target sample* is the set of individuals or sampling units that are drawn from the frame with the intent of inclusion in the sample (effectively the number of individuals contacted or asked to provide a response). In a noncensus survey, this number will be less than the total number of individuals in a frame, which introduces the concept of sampling rate (percentage of the frame that is selected for participation [the target sample size divided by the frame size]). Organizational surveying efforts where "everyone" is offered the opportunity to participate (sometimes referred to as an *organizational census*) would have a sampling rate of 100 percent, although even with a 100 percent sampling rate, misrepresentation is possible due to errors in frame coverage (either over- or undercoverage) or nonresponse. So-called pulse surveys will also typically be characterized by sampling rates less than 100 percent, again introducing the possibility of sample-population discrepancy.

In Figure 1, we present our basic but essential concepts for any survey project (organizational or public opinion polling) in visual form. Figure 2 demonstrates that the concepts of coverage and sampling rate are ignored or implicitly assumed to be 100 percent when the frame and target sample are not explicitly acknowledged as important sampling concepts, resulting in an undeserved exclusivity granted to response rate as the only

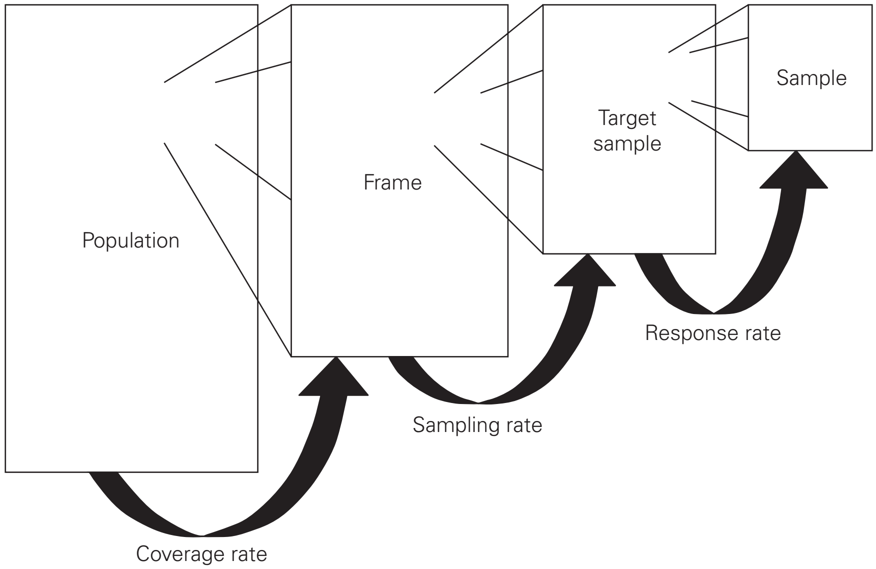

FIGURE 1. The Relationship between Sample and Population in Survey Research (Organizational or Public Opinion Polling)

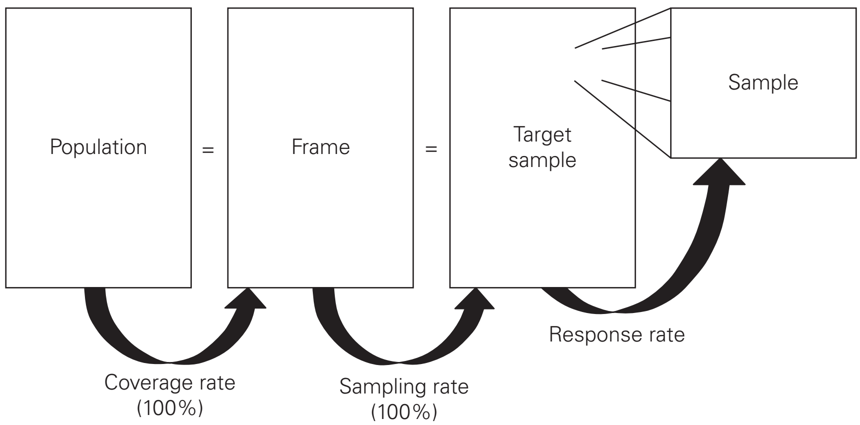

FIGURE 2. Tacit Assumption in the Organizational Science Application of Surveying Via Organizational "Census"

potential source of misrepresentation. In Table I, we present several commonly encountered public opinion sampling procedures with noted comparisons to HRM applications. Specifically, Table I describes common methodological relationships between the Figure 1 concepts of frame and target sample (but note that our conceptualization of nonprobability procedures could alternatively be presented as sampling procedures that have no explicit frame identification). Organizational surveyors may find Table I helpful in planning an organizational survey sampling project strategy or in situating their procedure within a methodological framework. (As noted further below, the acknowledgment of such a procedure can help the surveyor better identify potential contributing sources of sample misrepresentation.)

Causes of Sample Misrepresentation

Nonrepresentative samples are expected with nonprobability sampling procedures (consider, e.g., Kish's grape analogy above). There are also, however, three common scenarios with probability sampling procedures that can result in nonrepresentative realized samples. The first is undercoverage of the population by the frame (as mentioned above) and highlighted in the Alfred Landon anecdote. Organizational surveyors are most likely to encounter undercoverage in organizations with rapidly transitioning employees, high

TABLE I Classification of Sampling Procedures and Description as Applied within Organization Surveying Contexts (Possible Frame-Target Sample Relationships [Nonprobability Procedures Do Not Explicitly Acknowledge Frame])

Sampling Procedure	Description/Example	Determinant of Sample Inclusion	Appropriate Weighting Procedure		
			Design	Scale	Proportional
Census*	Every organizational member is sampled	Pollster (Conscripted)			X†
Probability (simple random)*	A random selection algorithm is applied to the organizational roster, and these individuals (who constitute the target sample) are invited to participate	Pollster (Conscripted)		X	X†
Probability (stratified proportionate)*	Desired sample size is first estimated, and the sampling ratio (n_k/N_k) is applied to strata of interest to determine a number of participants within each strata	Pollster (Conscripted)		X	X†
Probability (stratified disproportionate)*	Desired sample size is estimated, and sampling ratios (n_k/N_k) are adjusted to oversample small groups	Pollster (Conscripted)	X	X	X†
Probability (partial)*	Only targeted subgroups are invited to respond (e.g., nonmanagement; this is a special case of the stratified disproportionate sampling procedure when at least one sampling ratio is zero)	Pollster (Conscripted)		X	X†
Nonprobability (fortuitous)	Survey is made available (for instance at a common kiosk) and employees invited to respond	Respondent (Volunteered)		X	X
Nonprobability (quota sampling)	Survey administrators identify a small number of variables (e.g., location, function) and construct a target sample that closely matches population percentages of these variables	Respondent (Volunteered)		X	X

*Procedure technically requires 100 percent response rate from target sample.
†= appropriate only if any Figure 1 rate < 100 percent.

turnover rates, or poor internal tracking systems. The second scenario is too small of a sampling rate (or *rates* if proportional sampling procedures are pursued). Organizational surveyors are most likely to encounter sampling rate issues when they employ nonprobability sampling procedures (e.g., make a survey available and invite interested employees to participate) or pulse surveys are conducted. The third possibility is that sample misrepresentation arises due to nonresponse.[1] These threats to sample representativeness exist and apply to all survey procedures, whether considered probability sampling or nonprobability sampling.

Nonrepresentative samples are expected with nonprobability sampling procedures. There are also, however, three common scenarios with probability sampling procedures that can result in nonrepresentative realized samples.

Misrepresentation and Bias

In our presentation, misrepresentation is simply a colloquial term used to describe the statistical concept of *bias*. Bias is a deviation between a sample statistic and population parameter, and the *potential for bias* exists in all surveying applications when any of our above-mentioned rates deviates from 100 percent. Whether or not bias emerges in these situations depends on the dissimilarity of individuals included and excluded from any of our Figure 1 elements (frame, target sample, and realized sample). The specific expression of bias presented within our current broadened framework is therefore:

$$Bias_{\bar{y}} = P_{NF}(\mu_{NF} - \mu_F) + P_{NS}(\mu_{NS} - \mu_S) + P_{NR}(\mu_{NR} - \mu_R) \quad (1)$$

Here, bias in the sample mean ($\bar{y}$) is a function of the previously mentioned elements: P_{NF} is the proportion of the population not covered by the frame, μ_F is the mean response from the individuals within the frame, μ_{NF} is the mean response from the individuals not covered by the frame, P_{NS} is the proportion of the frame not included in the target sample, μ_S is the mean response from the entire target sample, μ_{NS} is the mean response from the nonsampled individuals, P_{NR} is the proportion of the target sample not responding to the survey, μ_R is the mean response from the respondents in the realized sample, and μ_{NR} is the mean response from the nonrespondents. Note that bias does not exist in the sample statistic if *either* the proportional element *or* the mean difference is zero for any Equation (1) component. This is why previous researchers correctly note that a low response rate may or may not result in poor survey results. The lack of absolute value specification across elements also acknowledges that these various sources of bias can theoretically have a summative canceling effect. The last elements of our above bias equation are, by tradition, the lone explicit focus of organizational researchers (e.g., P_{NR} [$\mu_{NR} - \mu_R$] is the common mathematical expression for nonresponse bias).

Full incorporation of our presentation of bias and "new" sampling methodology concepts within the HRM and organizational surveying domains is attempted in Table II, where we reference the nonresponse bias impact assessment strategy (N-BIAS) advocated by Rogelberg and Stanton (2007). N-BIAS is a "… series of techniques that when used in combination, provide evidence about a study's susceptibility to bias and its external validity" (Rogelberg & Stanton, 2007, p. 195). The focus is in measuring and mitigating *nonresponse* bias with, again, an acknowledgment that simple nonresponse does not necessarily result in threats to external validity. That is, Rogelberg and Stanton (2007) propose that researchers interested in quantifying the quality of a survey should perform an assessment of nonresponse bias impact rather than relying on simple response rates. We agree and extend their procedure here to include (1) additional potential sources of bias and (2) identification of groups to retain for purposes of post-stratification weighting.

Note here that the deciding factor for which grouping/demographic variables are included in organizational surveys has traditionally been based on the needs of the organizational survey client, and a typical projection of, "What do we want to do with this survey data?" When post-stratification weighting is employed, there is an additional consideration that needs to be made prior to survey administration: "Is there an anticipated pattern of error or bias (e.g., through attending to resources such as Tables I and II), and can this pattern be captured via group-level information?" The answer to this question leads the organizational surveyor to additional grouping information to capture—not for purposes of feedback delivery, but rather for sampling methodology (to align realized sample characteristics with known population parameters via sample weighting).[2]

Types of Weights and Weighting Procedures—an Overview

Differences between population and sample can be grossly categorized as reflecting differences of scale (e.g., the sample contains fewer people than the population) or proportion (e.g., certain identifiable groups may be under- or overrepresented within the sample [relative to the population]).

TABLE II Focus of Attention and Weighting Targets within the Broader Methodological Framework

Focus of Attention	How to Investigate	When to Weight	What to Weight
Population–Frame	Multiframe comparison	Obtained sample is over- or under represented relative to proper frame	Grouping variables that exhibit obtained sample—proper frame misspecification
	*Archival analysis	Respondent demographic constituencies differ from employee roster	Variables for which frame-obtained sample differences exist
Frame–Target Sample	Sampling procedure identification: Nonprobability sampling	Volunteer, opt-in, or convenience sampling	Grouping variables with significant differences between obtained sample and frame
	Sampling procedure identification: Intentional nonproportional sampling	Oversampling or undersampling within a grouping variable	Grouping variables which have nonproportional obtained sample allocations
	*Demonstrate generalizability	Additional obtained samples differ significantly from original sample	Grouping variables that exhibit attitudinal scale differences across samples
Target Sample–Obtained Sample	*Active nonresponse analysis	Active nonresponse within any group or across groups anticipated to be greater than 15%	Grouping variables that capture greatest variability in active nonresponse
	*Benchmarking analysis	Significant differences exist between sample characteristics and norms	Grouping variables that exhibit significant differences
	*Follow-up approach	Follow-up nonrespondent sample attitudes differ significantly from realized sample	Variables associated with respondent/nonrespondent attitude differences
	*Interest-level analysis	Interest-level items exhibit associations with substantive survey content	Grouping variables significantly associated with interest items
	*Wave analysis	Late respondent attitudes differ from early respondents	Grouping variables associated with significant attitudinal differences
	*Worst-case resistance	Number of nonrespondents needed to reverse conclusions is possible	Grouping variables about which the conclusions can be reversed

*N-BIAS technique presented in Rogelberg and Stanton (2007). Possible sources of misrepresentation arise whenever the coverage, sampling, or response rate is less than 100 percent.

Common weighting procedures parallel these forms of difference.

Scale Weight (v)

A scale (also called *base*) weight addresses size differences between population and sample and its computation is fairly straightforward: the scale weight is the reciprocal of the sampling ratio (the sampling ratio is the sample size [n] divided by the population size [N]):

$$v = \frac{N}{n} \tag{2}$$

When applied to a data set, the scale weight transforms output to the population metric (e.g., instead of reporting that n respondents [individuals in the sample] are happy/unhappy with their level of pay, application of the scale weight would enable an estimate to be reported regarding the inferred number of organizational members who are happy/unhappy with their level of pay). Other types of weighting are then interpreted as adjustments to the base scale weight.

Proportional weighting is concerned with sampling ratios within identifiable strata. There may be as many proportional weights as there are identifiable (and important) strata

Proportional Weight Adjustment (π_k)

Proportional weighting is concerned with sampling ratios within identifiable strata. There may be as many proportional weights as there are identifiable (and important) strata. These weights are technically adjustments to the base weights, characterized as ratios of the proportion of strata members within a population to the proportion of strata members within the sample and constitute our focus in this article:

$$\pi_k = \frac{N_k/N}{n_k/n} = \frac{N_k/n_k}{N/n} \tag{3}$$

Oversampling elements of a strata (k) results in proportional weights less than one, while undersampling (relative to the population) results in proportional weights greater than one. The proportional weighting can be done before or after the sample is drawn. When applied before the sample is drawn, the proportional weights are simply a special type of design weight, as described next. However, more often the proportional weight adjustments are applied after the sample is drawn to adjust for imperfections such as nonresponse and are in that special case called *post-stratification weights*.

Design Weight Adjustment (D_k)

Scale and post-stratification proportional weights can be computed without any necessary consideration given to sampling methodology (see Table I). Design weight adjustments are different. They are generalizations of presampling proportional weight adjustments and are fully determined by the sampling procedure. Here, the reciprocal of selection probabilities is applied to strata, and these reciprocal probabilities constitute design weight adjustments to the base scale weights. These adjustments account for different presampling probabilities of selection from the population:

$$D_k = \left(\frac{1}{p_k(\text{selection into sample})}\right) / \left(\frac{N}{n}\right) \tag{4}$$

For example, in a public opinion polling context, such weights may be constructed based on cell phone and landline ownership due to different likelihoods of being contacted. In an organizational context, the need for design weights would also most likely be attributable to frame multiplicity (where an employee had, for instance, multiple appointments, e-mails, or contracts). If frame multiplicity is a concern, the organizational surveyor can apply design weight adjustments in addition to proportional weight adjustments. For the majority of organizational surveying applications, however, design weight adjustments will be unnecessary. If the reader is interested in learning more about these types of weights, contact the corresponding author to request supplementary material to this article.

Mixed Weight (W_k)

The three *types* of weights can be combined, such that a sample is weighted to account for differences in sampling ratios across strata in addition to differences in sampling probability as well as being transformed to the population metric. Such indices are known as *mixed*, *combined*, or *integrated* weights and are computed by multiplying the base scale weight with the appropriate design weight adjustment, or more generally multiplying the base scale weight, the design weight adjustment, and the post-stratification weight adjustment:

$$W_k = vD_k\pi_k \tag{5}$$

Which Type of Weighting to Apply

The most common use of weighting in organizational sampling would be for purposes of post-stratification adjustment. For most software applications, the scale weight v is built into all calculations of means and estimates of population totals, so this weight need not be explicitly used in practice. Furthermore, design weights D_k are generally not needed, unless there is a predetermined plan to oversample some strata and undersample other strata (e.g., using a stratified proportionate sampling procedure as described in Table I). Barring this kind of situation, the primary use of weighting for organizational samplers would be in the post-stratification weights π_k.

It is for this reason that the weighting procedure description that follows is based on the proportional weight adjustment, and the specific scenario in which they are implemented as post-stratification weights. Default base scale weights as described by Equation (2) are assumed as a starting point, so to simplify the manuscript presentation the proportional adjustments will be referred to as proportional weights. These weights vary from being less than 1 to greater than 1. In this context, an *average* weight of 1 is desirable so that the overall weighted sample size (found by summing the proportional weights) is the same as the original sample size n. These post-stratification weights correct for the simple discrepancy between realized sample characteristics (along predetermined strata) and known (or very well estimated) population parameters. These are also the "weights" that are implicated by authors who have previously mentioned weighting as an option in organizational surveying (e.g., Macey, 1996; Kraut, 1996b; Rogelberg et al., 2000) and are most appropriate for integration within the organizational surveying literature that historically focuses exclusively on nonresponse (e.g., Rogelberg et al., 2000, 2003; Rogelberg, Spitzmüeller, Little, & Reeve, 2006; Rogelberg & Stanton, 2007).

Weighting Procedures

There are two common post-stratification weighting *procedures* (methods of applying the weights to a data set). First, if population cell frequencies (joint probabilities across strata dimensions) are known, weighting may be applied at this level. This is the desired/less-complex scenario and is fully described by direct application of Equation (3). Throughout our presentation, we refer to this procedure as *cell-level weighting*. Second, when joint probabilities are not known, an iterative process referred to as *raking* is employed.

Raking

When joint probabilities (across strata) are unknown and there is more than one strata that is taken into consideration, the standard procedure for multiple weight estimation is an iterative process that may be referred to by multiple names. *Rim weighting*, *iterative proportional fitting*, and *raking* are all terms that may be implicating this same general process first advocated by Deming and Stephan (1940). The general procedure applies each weight in progressive independent stages and consists of the following steps:

1. Determine the design weight adjustment D_k (this weight is simply a default value of "1" if design weights are not specified)
2. Determine proportional weights for one strata (a proportional weight is the population percent divided by the sample percent). Assign these weights to cases: $D_k\pi_{k1}$.
3. Determine proportional weights for the second strata, after the step 2 weights have been applied (the current sample percentages will be affected by the step 2 weighting procedure). Multiply previous (step 2) weights by the proportional weights for this second strata: $D_k\pi_{k1}\pi_{k2}$.
4. Determine proportional weights for the third strata (which will once again require inspection of the *current* sample percent). Multiply the previous step 3 weights by the third strata proportional weights: $D_k\pi_{k1}\pi_{k2}\pi_{k3}$.
5. Repeat steps 2, 3, and 4 (or more if more than three strata are considered) in sequence until the sample characteristics closely match the population characteristics.

These weights vary from being less than 1 to greater than 1. In this context, an average weight of 1 is desirable so that the overall weighted sample size (found by summing the proportional weights) is the same as the original sample size n.

Procedurally, Kish (1965) recommends that each cell or variable used for weighting contain at least 10 observations. Similarly, for both raking and cell-level weighting, the distribution of weights should be investigated, with a focus on extremely large or small values. When extreme values are found, the practice of "trimming" is fairly common in public opinion polling, whereby extreme weights are limited (e.g., to a value of 0.5 for oversampled groups and 2 for undersampled groups).

Weighting and Sampling Error

In a typical unweighted survey, rough estimates of descriptive statistic accuracy are commonly estimated via the standard error of the sample mean (the denominator in a one-sample z- or t-test; $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$; $s_{\bar{x}} = \frac{s}{\sqrt{n}}$)). Building confidence intervals of specified range (e.g., 95 percent confidence interval) entails multiplying the standard error estimate by a tabled two-tailed critical value associated with the desired confidence level (e.g., $1.96(\sigma_{\bar{x}})$ or $t_{cv}(s_{\bar{x}})$).[3] These standard error bands are applied to point estimates (e.g., $\bar{X} \pm 1.96(\sigma_{\bar{x}})$) to indicate the range within which there is a (95 percent) probability that the population mean, μ, resides (e.g., across 95 percent of sampling occasions, the identified band will include the population mean). This basic approach to the estimation of sampling error differs with the specification of weights.

When weights are utilized, the appropriate approximation of sampling error is not the typical standard error of the sample mean. Mallett (2006) provides one broad correction (multiplying the typical standard error by a constant "design factor," [K], where $se_{\text{weighted}} = K \times s_{\bar{x}}$). The design factor is a function of the average weight and average squared weight:

When weights are utilized, the appropriate approximation of sampling error is not the typical standard error of the sample mean.

$$K = \frac{\sqrt{\overline{W^2}}}{\overline{W}}\ ;\ \overline{W} = \frac{\Sigma w}{n}\ ;\ \overline{W^2} = \frac{\Sigma(w)^2}{n}.$$

The use of sample weights therefore results in larger standard error estimates than the unweighted estimates. How much larger depends on the variation in the weights. Mallett's (2006) correction is not the only option, and different standard error estimates can be generated based on a multitude of sampling designs. It should also be noted that if standard error estimates are generated via software programs, it is important that only proportional weights and not scale weights be specified when requesting the uncorrected standard error (or for that matter *any* inferential index). If the reader is interested in learning more about different standard error estimates (or alternative corrections), good resources are Maletta (2007) and Kish (1965).[4]

When only one variable (or cell information) is used for weighting, and proportional weighting is applied, the sum of the weights will be equal to the sample size. When multiple strata are considered (e.g., raking), the resulting sum of the weights may not necessarily equal the sample size. As mentioned above, when raking, we recommend "centering" the weights so that the average weight $\overline{W}$ is equal to 1. This is accomplished through rescaling the final weights such that if, for example, the average weight is 2, every weight is divided by this number (2). If this rescaling is performed, the denominator of the formula for the design factor becomes a value of 1 and the design factor is therefore simplified to the square root of the average squared weight.

A Simplified Example—the Impact of Weights

Table III presents a simplified fictional situation to demonstrate the impact of proportional weights on scale-level point estimates. Here, three respondents provide single-item evaluations of (a) satisfaction with their boss, and (b) satisfaction with their pay. These three individuals represent three different functional areas that are (relatively) over- or underrepresented within the survey sample. Specifically, the weights reflect a relative undersampling of human resource employees (compared to marketing) and research and development (compared to both marketing and human resources). Aggregating the raw sample information across functional areas (to get a snapshot of overall satisfaction with supervision or pay within the organization) would result in identical satisfaction ratings (3.3 on the original 5-point scale metric) without weighting, but reveal dissatisfaction with supervision under the weighted aggregation (2.7 on the 1 [dissat] → 5 [sat] scale).

Note that the mean score differences between the unweighted and weighted data result from the *frequency* of response being altered. In the unweighted scenario, responses of "1," "4," and "5" all occur with equal (33%) frequency, and represent a standard error estimate of 1.2. In the weighted scenario, for "Sat w/ Boss," the frequencies of occurrence of each rating are "1" = 50% (3/6), "4" = 33% (2/6), and "5" = 17% (1/6), with the adjusted

TABLE III Sample Weighting of Two Satisfaction Scale Scores across Functional Area

		Sat (5-point ordinal scale)	
Functional Area of Respondent	**Weight**	**Sat w/ Boss**	**Sat w/ Pay**
Human Resources	2	4	1
Marketing	1	5	4
Research and Development	3	1	5
Sat w/ Boss vs. Pay (Unweighted):		10/3 = 3.3	10/3 = 3.3
Sat w/ Boss vs. Pay (Weighted):		16/6 = 2.7	21/6 = 3.5

standard error estimate for the aggregate mean at a value of 1.3.[5] This rather simple scenario can quickly become complex with the addition of just a few more strata of respondents (e.g., under/over-representation by shift, geographic location, or function).

The remainder of our presentation entails practical applications of the above material, demonstrating the effect of weighting under differing sampling circumstances and walking the reader through a procedural tutorial.

Method

We use an artificially constructed data set for two purposes: (1) to demonstrate the effect of weighting on point estimates, and (2) to provide a tutorial on "how to weight." Our intent was to make the data and procedural applications *available* as well as making them prototypical. All data (SPSS data format), weight calculations (Excel document), and procedural applications (SPSS syntax) for the tutorial are therefore available as online supplements to this article, and further below we describe how the reader can use these files to help learn "how to weight."

Both the demonstration and tutorial use the same data set (and this is also the data set shared via online supplements). We constructed this static data set via WinGen 3.0 (Han, 2007) with the following characteristics (region, function, and salaried status are fictitious and specified only to provide some context to the data): 120 MidWest respondents, 100 NorthEast respondents, and 90 SouthEast respondents; 95 IT respondents, 85 HR respondents, and 130 customer service representatives; 245 hourly and 65 salaried workers (total survey, $n = 310$).

The data consists of ordinal ($1 \rightarrow 5$) responses to 7 items collectively defining a survey scale fictionally labeled "Communication." Across these items, SouthEast Customer Service workers (both salaried and hourly workers) exhibited suppressed scores. Cronbach's alpha for the 7-item scale (across all 310 respondents) was .89, with corrected item total correlations ranging from .51 to .81. The unweighted scale average across all individuals was 3.2 ($s = 1.3$). This estimate reflects equal (unit) weighting of individual respondents. Detailed characteristics of the data set are presented in Table IV.

Demonstration of the Effect of Weights

Procedure and Results

To demonstrate the "effect" of the weighting procedure, we conducted multiple random samplings of our tutorial data, effectively treating the 310

TABLE IV Prescribed (Desired Theta) and Realized (Empirical Mean) Sample Characteristics of Simulated Data

Geographic Region	Functional Area	Salaried Status	n	Simulated Theta	Realized Mean
MidWest	IT	Hourly	30	0	3.0
		Salaried	10	0.5	3.8
	Human Resources	Hourly	30	1.5	4.3
		Salaried	5	0.5	3.6
	Customer Service	Hourly	35	0	3.2
		Salaried	10	1	3.8
NorthEast	IT	Hourly	25	0.5	3.6
		Salaried	5	0	2.7
	Human Resources	Hourly	20	1.5	4.4
		Salaried	10	1	4.1
	Customer Service	Hourly	35	.5	3.3
		Salaried	5	0	2.9
SouthEast	IT	Hourly	20	0	2.7
		Salaried	5	1.5	4.4
	Human Resources	Hourly	15	0.5	3.4
		Salaried	5	0	4.1
	Customer Service	Hourly	35	–2.5	1.2
		Salaried	10	–2.5	1.3

individuals as a population (as noted below, the *tutorial* treats these 310 individuals differently). To simulate the presence or absence of bias within our unweighted sample means, we crossed random samplings of 20 percent and 80 percent of our "disgruntled" (SouthEast customer service employees) as well as "satisfied" (all other employees) groups. We chose these numbers of 20 and 80 to reflect fairly extreme conditions of either high or low Figure 1 rates rather than attempting to replicate a standard (as mentioned in the introduction, the literature reports an average *response* rate of 50 percent; Baruch, 1999; Baruch & Holtom, 2008). Specifically, across the samplings, the South Eastern Customer Service employees (n = 45) were randomly sampled at rates of 20 percent or 80 percent. Similarly, all other employees (n = 265) were also randomly sampled at either 20 percent or 80 percent.

The presence or absence of bias in these simulations could be attributable to any of our identified bias elements (e.g., coverage, sampling, or nonresponse, see Equation (1) above). That is, the simulations model any situation where a rate deviates from zero (in our cases, deviations of 20 percent and 80 percent) and means differ inside and outside of the *frame*, *target sample*, or *realized sample*. Subsequent to the samplings, weighting was performed (either raking or cell-level weighting)[6] and both weighted and unweighted means were computed at overall organization, regional, functional, and salaried status levels. Eighty thousand total random samplings of the n = 310 were performed (10,000 random samplings for the four combinations of 20 percent and 80 percent performed twice [40,000 with cell-level weighting and 40,000 with raking]).

Tables V and VI present the results of our simulations, showing that (1) when bias exists, the weighting makes a corrective adjustment; and (2) when bias does not exist, the weighting does not appreciably "hurt" the estimate. Specifically, to use nonresponse as an example (although low coverage or sampling rates are equally plausible illustrations), with rampant but fully random nonresponse (random samplings at 20 percent), the weighting procedures may result in slightly deviant point estimates from the parameter. With higher response rates (80 percent random samplings), these deviations fully disappear. In the biased sampling scenarios (Table VI), we direct the reader's attention to the weighted correction at the level of the overall mean. At this highest level of aggregation, the unweighted estimate exhibits a bias of +0.24 or –0.59 on our 5-point scale (with the erroneous conclusion that employees are, on average, dissatisfied in the –0.59 scenario). Weighting correctly mitigates this bias, with the raking procedure resulting in improvements from $-0.59_{\text{unweighted}}$ to -0.11_{weighted} and $+0.24_{\text{unweighted}}$ to $+0.11_{\text{weighted}}$, and cell-level weighting resulting in improvements of $-0.59_{\text{unweighted}}$ to -0.02_{weighted} and $+0.24_{\text{unweighted}}$ to 0.00_{weighted}. Similar patterns are found in the subgroup means (particularly the "salaried status" aggregations, where bias corrections are substantial, and again correct the erroneous unweighted conclusion of subgroup dissatisfaction [unweighted estimates indicate dissatisfaction, weighted estimates indicate satisfaction]). The results here demonstrate that the procedure shines when bias is present in point estimates (when bias-relevant strata are identified and applied). When an unweighted bias exists, it is corrected. When an unweighted bias is absent,

TABLE V Absence of Bias Simulations (Mean and [Standard Deviation])

		20% non-CS SE–20% CS SE			80% non-CS SE–80% CS SE		
	Parameter	Unweighted	Raking	Cell	Unweighted	Raking	Cell
Overall Mean	3.20	3.20 (0.11)	3.20 (0.11)	3.19 (0.11)	3.20 (.03)	3.20 (0.03)	3.20 (0.03)
Regional							
MidWest	3.53	3.53 (0.19)	3.53 (0.19)	3.52 (0.19)	3.53 (.05)	3.53 (0.05)	3.53 (0.05)
NorthEast	3.63	**3.64** (**0.21**)	3.64 (0.21)	3.66 (0.22)	3.63 (.05)	3.63 (0.05)	3.63 (0.05)
SouthEast	2.27	2.26 (0.21)	2.26 (0.22)	2.20 (0.18)	2.27 (.05)	2.27 (0.05)	2.27 (0.03)
Functional							
IT	3.24	3.24 (**0.21**)	3.25 (0.22)	3.22 (0.22)	3.24 (.05)	3.24 (0.05)	3.24 (0.05)
Human Resources	4.08	4.08 (0.18)	4.08 (0.18)	4.09 (0.19)	4.08 (**.04**)	4.08 (0.04)	4.08 (0.04)
Customer Service	2.59	2.58 (0.18)	2.58 (0.19)	2.58 (0.17)	2.59 (.04)	2.59 (0.05)	2.59 (0.04)
Salaried Status							
Hourly	3.16	3.16 (0.13)	3.15 (0.13)	3.15 (0.12)	3.16 (.03)	3.16 (0.03)	3.16 (0.03)
Salaried	3.36	**3.37** (0.32)	3.36 (0.34)	3.33 (0.31)	3.36 (.08)	3.36 (0.08)	3.36 (0.06)

Note: **Bold** indicates simulation estimate deviations of one one-hundredth across the two *sets* of simulations (one set applied cell weights, the other set raked).

TABLE VI Bias Present Simulations (Mean and [Standard Deviation])

		20% non-CS SE–80% CS SE			80% non-CS SE–20% CS SE		
	Parameter	Unweighted	Raking	Cell	Unweighted	Raking	Cell
Overall Mean	3.20	2.61 (0.08)	3.09 (0.10)	3.18 (0.11)	3.44 (0.03)	3.31 (0.03)	3.20 (0.03)
Regional							
MidWest	3.53	3.53 (0.19)	3.55 (0.19)	3.52 (0.19)	3.53 (0.05)	3.51 (0.05)	3.53 (0.05)
NorthEast	3.63	**3.64** (**0.21**)	3.70 (0.21)	3.66 (0.22)	3.63 (0.05)	3.57 (0.05)	3.63 (0.05)
SouthEast	2.27	1.65 (0.11)	1.81 (0.15)	2.19 (0.16)	2.89 (0.07)	2.76 (0.07)	2.29 (0.07)
Functional							
IT	3.24	3.24 (**0.21**)	3.27 (0.23)	3.22 (0.22)	3.24 (0.05)	3.22 (0.05)	3.24 (0.05)
Human Resources	4.08	4.08 (0.18)	4.15 (0.18)	4.09 (0.19)	4.08 (**0.04**)	4.01 (0.05)	4.08 (0.04)
Customer Service	2.59	1.90 (0.11)	2.27 (0.15)	2.57 (0.16)	3.06 (0.05)	2.92 (0.05)	2.60 (0.06)
Salaried Status							
Hourly	3.16	2.58 (0.10)	3.08 (0.12)	3.15 (0.12)	3.39 (0.04)	3.24 (0.04)	3.16 (0.03)
Salaried	3.36	2.69 (0.24)	3.15 (0.28)	3.30 (0.26)	3.64 (0.09)	3.57 (0.13)	3.39 (0.14)

Note: **Bold** indicates simulation estimate deviations of one one-hundredth across the two *sets* of simulations (one set applied cell weights, the other set raked).

the weighted correction does not appreciably deviate, although it will introduce a small amount of error when the coverage, sampling, or response rate is very low.

Tutorial ("How to Weight")

Procedure

For the tutorial, the n = 310 data set was treated as a *realized sample* within which groups were relatively over- or underrepresented. This treatment mimics the typical organizational surveyor data set. Weights were applied to the data at both cell and marginal (variable) levels, and were each estimated three times representing targeted misrepresentations of small, moderate, and large effect. The implications of these different effects are quite limited—they are specified here only to demonstrate the impact such deviations can have on weighted point estimates. For the cell-level (joint probabilities known) estimates, the SouthEast Customer Service employee group was treated as being underrepresented by 10 percent, 30 percent, and 50 percent. Procedurally, the other individuals were treated as being relatively overrepresented (relative to the SouthEast Customer Service employees), and this overrepresentation effect was held constant across groupings (e.g., the equivalent relative oversampling effect was applied to each grouping).

For the demonstration of the raking procedure, the SouthEast region, Customer Service function, and Salaried status were each treated as being underrepresented by 10 percent, 30 percent, and 50 percent. Again, the other marginal categories were all treated as being equally oversampled (e.g., all comparative groups were relatively overrepresented to the same extent). This "all other groups exhibiting equal sample constituency" effect would not typically be considered representative of real-world data, but was specified here to facilitate explanation on the application of post-stratification weights.

How to Use the Supplementary Files

The data itself ("sample data.sav") and weighting algorithms ("weight calculations.xlsx"; "how to weight.sps") used to generate the sets of results reported in Table VII are available by request from the corresponding author.

The implications of these different effects are quite limited—they are specified here only to demonstrate the impact such deviations can have on weighted point estimates.

Weight Calculations

The "weight calculations.xlsx" file is a repository and workspace for the sample statistics and population parameters. A file such as this needs to be created as a first step when "using" weights. There are two tabs in this Excel file. The first "cell weighting" tab demonstrates the calculation of weights when joint population constituencies are known (these are presented in, for example, column C). The population percent (column C) divided by the sample percent (column B) yields the weights (e.g., column G) used for cell weighting. These are simply applications of Equation (2) presented above.

The "raking" tab contains information pertinent for the calculation of *starting* weights (e.g., column G) when joint probabilities are unknown. These weights are calculated in the same manner as above (population percentage [column C] divided by sample constituency [column B]), however, when incorporated into the raking procedure, only one set of these weights will be directly applied to the data (these weights are highlighted in *yellow* in the "weight calcluations.xlsx" file).

How to Weight

The "how to weight.sps" SPSS syntax applies the Excel-specified weights to our sample data. The syntax consists of ten sets of commands, is heavily annotated, and is intended to be run in segments. The first three commands demonstrate cell-level weighting. The first two sets of commands identify the sample data file and apply cell-level weights to the appropriate employee groups via a series of *do if/else if* statements. Note that the numbers contained in these *do if/else if* statements are the values contained in column G of the "cell weighting" Excel file tab. The third command requests the weighted communication scale estimate to be reported separately for MidWest, NorthEast, and SouthEast regions (the resulting means are identified in Table VII via shading).

Commands 4 through 10 demonstrate the raking procedure. Command 4 identifies the raw data file and command 5 requests sample constituencies for the three grouping variables. The sixth command initiates the raking procedure via application of the weights presented in column G of the Excel file's "raking" tab, and commands 7 and 8 continue the raking procedure. The ninth and tenth commands request weighted descriptive information (means and standard errors), and these estimates are again identified via shading in Table VII.

Results

Table VII presents scale averages and standard errors aggregated at overall organization, regional, functional, and salaried status levels. These averages reflect either raw individual aggregation or weighted individual aggregation (and standard error correction) based on relative underrepresentation of either the SouthEast Customer Service workers (n = 45; 14.5 percent of the realized respondent sample) or SouthEast (29.0 percent), Customer Service (41.9 percent), and Salaried (21.0 percent) employee groups. The primary interest in the table lies in the deviation of point estimates in the unweighted versus weighted columns, and how these deviations increase as the undersampling becomes more extreme. As the simulations demonstrate, standard errors also increase with the application of weights, and the magnitude of difference between the unweighted and weighted estimates can be quite large. It should

TABLE VII Mean (Corrected Standard Error) Communication Estimates across Organizational Strata with Relative Undersampling of the SouthEast Customer Service Salaried Employees (SPSS-Calculated Supplementary File Estimates Shaded)

		Cell Weighting			Rim Weighting		
	No Weight	10% Undersampling	30% Undersampling	50% Undersampling	10% Undersampling	30% Undersampling	50% Undersampling
Overall Mean	3.20 (0.072)	3.16 (0.072)	3.10 (0.072)	3.03 (0.074)	3.12 (0.072)	2.97 (0.076)	2.81 (0.084)
Regional							
MidWest	3.53 (0.096)	3.53 (0.096)	3.53 (0.096)	3.53 (0.096)	3.52 (0.096)	3.52 (0.100)	3.51 (0.107)
NorthEast	3.63 (0.106)	3.64 (0.106)	3.64 (0.106)	3.64 (0.107)	3.60 (0.107)	3.54 (0.110)	3.48 (0.117)
SouthEast	2.27 (0.133)	2.21 (0.133)	2.09 (0.135)	1.98 (0.137)	2.20 (0.134)	2.06 (0.139)	1.92 (0.148)
Functional							
IT	3.24 (0.105)	3.24 (0.105)	3.24 (0.105)	3.23 (0.106)	3.25 (0.106)	3.27 (0.108)	3.29 (0.112)
Human Resources	4.08 (0.089)	4.08 (0.089)	4.09 (0.089)	4.09 (0.090)	4.06 (0.090)	4.03 (0.092)	4.01 (0.096)
Customer Service	2.59 (0.115)	2.54 (0.115)	2.44 (0.117)	2.36 (0.119)	2.53 (0.116)	2.43 (0.119)	2.33 (0.124)
Salaried Status							
Hourly	3.16 (0.080)	3.12 (0.080)	3.06 (0.080)	3.01 (0.082)	3.07 (0.080)	2.91 (0.084)	2.74 (0.090)
Salaried	3.36 (0.162)	3.32 (0.162)	3.23 (0.164)	3.13 (0.170)	3.29 (0.163)	3.13 (0.170)	2.95 (0.184)

be noted that, quite different from the Table V and VI results, there is no "correct" estimate in these Table VII results (i.e., there is no standard to evaluate whether the weighted or unweighted estimates are "more accurate"). The impact of Table VII is therefore quite limited—the information in the table primarily serves as a standard to be replicated by the reader who is interested in applying the information contained in "weight calculations.xlsx" via the processes outlined in "how to weight.sps." Any reader who would prefer the tutorial in other software formats is asked to contact the first author, who can provide the same tutorial information in *R* script and also can likely accommodate additional requests for other commonly used software platforms.

Discussion

Several researchers have noted potentially misplaced emphasis on survey response rates, with Cook et al. (2000), Krosnick (1999), Rogelberg and Stanton (2007), and Visser, Krosnick, Marquette, and Curtin (1996) articulating the point of this current article: representativeness of the sample is much more important than response rate. Representativeness refers to the relationship between sample and population, whereas "response rate" has a more limited focus on the relationship between the realized sample and target sample (which obscures additional potential sources of error [the most notable of which being frame undercoverage; see Figure 1]). Krosnick (1999) also specifically comments that, even when probability sampling is employed, response rate does not necessarily implicate either good or poor sample representativeness. One aim of this article, therefore, is to reinforce this primary *representativeness* orientation to the organizational surveyor who may be otherwise inclined to focus on response rate as a sufficient index of data quality. We do this via a hybridized approach, taking what we deem to be the more important concepts from multiple disciplines to develop a framework with discipline-free fidelity (e.g., mutually acceptable terms and concepts) and recommend weighting as a corrective action to fix misrepresentation when it appears.

Because our intended consumer audience is organizational surveyors, we retained the N-BIAS model as the integrator of our interdisciplinary framework. N-BIAS is focused solely on *nonresponse*, and obviously organizational researchers will and should continue to be concerned with nonresponse (and below we use nonresponse to illustrate the ramifications of our simulations and tutorial), but explicit attendance to other potential sources of error (such as frame undercoverage/out-of-date employee records) is also an important exercise for increasing sample representation in organizational survey data. Through leveraging the familiarity with terminology and procedural recommendations of N-BIAS, we therefore hope to have contributed positively to an expansion of the conceptualization of error in organizational surveys. Rogelberg and Stanton (2007) refer to their presentation as a first attempt: N-BIAS version 1.0. We consider our presentation to be an extension of this entailing (1) considerations of sampling procedure, coverage, and sampling rates; and (2) the application of corrective weights.

One aim of this article is to reinforce this primary representativeness orientation to the organizational surveyor who may be otherwise inclined to focus on response rate as a sufficient index of data quality. We do this via a hybridized approach, taking what we deem to be the more important concepts from multiple disciplines to develop a framework with discipline-free fidelity (e.g., mutually acceptable terms and concepts) and recommend weighting as a corrective action to fix misrepresentation when it appears.

Tables I and II along with Figure 1 should be helpful in pursuit of this extension. As our simulations demonstrate, systematic patterns of nonresponse (again using the response rate illustration) can result in biased sample estimates, and post-stratification weighting can mitigate this bias (if the contributing groups can be identified and incorporated into the weighting algorithm). This requirement is also the central purpose of the N-BIAS strategy, outlining, "… what a survey researcher can do to test for and examine whether obtained data are bias susceptible" (Rogelberg & Stanton, 2007, p. 196). Our extension advocates the application of weighting algorithms as a corrective action when data is determined to be "bias susceptible" (note that "bias susceptible" requires both coverage, sampling, or response rates less than 100 percent and different *attitudes* of groups inside and outside of the frame, target sample, or realized sample). Although our simulations did result in the introduction of a small amount of error when weights are applied to *unbiased* sample estimates (obtained with very low coverage, sampling, or response rates but nondifferent frame, target sample, or realized sample attitudes), the trade-off between this slight addition of error in the low response-rate condition and the large advantage gained when a bias *does* exist in the data should embolden the organizational surveyor to weight.

We did not characterize the typical organizational survey as being represented by either probability or nonprobability sampling procedures. Rather, we simply provided a taxonomy of sampling procedures and ask that the surveyor refers to this taxonomy as well as Figure 1 and Table II as guides to help identify potential sources of error in survey applications. For example, if an organizational "census" is pursued (100 percent sampling rate), Figure 1 and Table II would direct the organizational surveyor to attend to coverage rate in addition to response rate. If smaller pulse surveys are conducted, the organizational pollster would be wise to apply a probability sampling procedure (such as stratified random sampling) and consider all three potential sources of misrepresentation (coverage, sampling, and response rates). Although some of these rates may seem nonapplicable to organizational surveying applications, we believe this is due to a tradition of surveying that largely ignores sampling methodology. We also note here that our presentation extracted concepts that we deemed most important and germane for organizational surveyors from a vastly larger and more extensive literature. The reader who is interested in learning more regarding sampling methodologies is therefore encouraged to reference comprehensive sampling texts such as, for example, Cochran (1977), Kish (1965), or Lohr (1999).

Our extension advocates the application of weighting algorithms as a corrective action when data is determined to be "bias susceptible."

Summary, Limitations, and Recommendations

It was alluded to in the introduction but should be explicitly articulated that a nonprobability sample does not technically yield a known standard error. It can be calculated in the usual manner, but the estimate is not based on strict properties of sampling distributions. We acknowledge that surveyors may find such estimates useful in a descriptive or comparative sense (e.g., compared with other standard errors from the same or previous samples of a similar nature). However, these estimates should not technically be used in inferential contexts (e.g., the construction of confidence intervals). This recommendation is obviously restrictive, and as pointed out by an anonymous reviewer, perhaps a bit quixotic within the realities of research within the organizational sciences. However, there may be middle ground to be found—again through attendance to interdisciplinary advancements. Survey methodologists have, within the past decade, been recognizing the concept of *total survey error* (as opposed to margin of error; see, e.g., Weisberg, 2005). Total error incorporates many different types of sample errors and may be a better index of precision for organizational surveyors (than traditionally estimated standard errors [particularly if the procedure/sample is nonprobability]).

Our simulations were facilitated by an SPSS macro developed by Vladimír Sedláček (we would not have been able to do the iterative simulations without such a tool). There is also an *R* package that performs raking: "anesrake." Either of these resources will automate weighting procedures, although we recommend that it initially be done manually via the method outlined in our tutorial (and online materials) until the practitioner fully understands "what the procedure is doing." Certainly, more comprehensive and detailed simulations could also further investigate the impact of weights under different conditions than the four that we evaluated. Our purpose in the current article was not to provide such a comprehensive set of simulations but rather to demonstrate that weights can correct biased estimates and that they can have an impact on estimates even in the weak extreme (e.g., a very small constituent group with discrepant opinions being underrepresented [cell weighting] or a larger constituent group with less discrepant opinions [rim weighting]). Moving forward, we hope we have convinced organizational surveyors to (1) think representativeness, not response rate; (2) consider additional potential sources of error beyond nonresponse; and (3) include weighting as remediation.

Author's Note

State Farm Mutual Automobile Insurance Company, its subsidiaries, and affiliates were not involved with the creation of this paper and no State Farm information was used in its development. The conclusions and opinions expressed in this document are solely those of the authors, and State Farm neither approves nor endorses the conclusion and opinions expressed in the document.

Notes

1. Note that in this article and the proposed framework, nonresponse is operationally defined as incongruence between target and realized samples.
2. Note that weighting is an alternative correction to Rogelberg and Stanton's (2007) suggestion of control

variable specification (e.g., with N-BIAS technique 5: interest-level analysis).

3. Note that these standard errors apply most appropriately to means sampled with a probability sampling procedure, and are therefore likely to be biased estimates in survey applications that employ nonprobability sampling. Although it is common practice to report such estimates, with nonprobability samples they are not based on strict properties of sampling distributions and we are not advocating their application (including the specification of confidence intervals). This circumstance again highlights the danger of failing to adhere to sampling methodology terms and procedures (for further detail, see Ludbrook & Dudley, 1998; Sterba, 2009).

4. Organizational surveyors, for example, may be interested in an additional correction that is appropriate with probability sampling procedures where respondents are likely to represent a significant percentage of the population. This is known as the finite population correction (fpc). The fpc corrects for the reduction in variation across all *possible* samples when sampling from a fixed size population. This correction is for a different purpose than Mallett's weighting correction. Both may be used simultaneously in estimation of a standard error in our context, but because our focus is on weighting, we only present the formula for Mallett's correction.

5. Note that this standard error adjustment must be computed by hand, as software program estimates will result in biased standard error estimates (.8 in the current example based on the weight-induced doubling of the sample size). If weights of "1," "0.5," and "1.5" were utilized instead of "2," "1," and "3," the sample size would have remained unchanged, but the reported standard error would not *necessarily* provide an accurate estimate.

6. Due to computing limitations (e.g., chronic crashes), we were restricted to performing two independent sets of simulations—one that performed cell-level weighting (and unweighted estimates) and one set that performed the raking procedure (as well as unweighted estimates). Tables V and VI both note the uncommon cases where the unweighted values differed very slightly across the two sets of simulations.

JOHN T. KULAS is head handyman at Corporate Mr Fixit. His work is focused on developing solutions for systematic nonfocal construct influences on psychological inventory response and applications of interdisciplinary research methodologies in organizational contexts.

DAVID H. ROBINSON has been a professor of statistics at the university level since 1979, teaching at St. Cloud State University (SCSU) since 1985. Early in his academic career, he published articles in the area of nonparametric statistics and coauthored a book on computer simulation of statistical distributions. In the past ten years he has devoted his career to applications of statistics in other fields, especially sample surveys and higher education data analysis projects. His current job title is faculty director of data analytics at SCSU.

JEFFREY A. SMITH is a research analyst for State Farm Mutual Automobile Insurance Company in the Department of Strategic Resources. He is currently a member of the Organizational Performance Measurement Team dedicated to understanding customer and employee experiences. He leverages a background in industrial/organizational psychology and computer science designing research to better understand matters impacting organizational outcomes.

DONALD Z. KELLAR graduated from St. Cloud State with degrees in statistics and mathematical economics. He is employed by an insurance company in St. Paul as an actuarial associate and is working to attain a designation from the Society of Actuaries. He lives in downtown St. Paul with his wife, Amy.

References

Anseel, F., Lievens, F., Schollaert, E., & Choragwicka, B. (2010). Response rates in organizational science, 1995–2008: A meta-analytic review and guidelines for survey researchers. Journal of Business and Psychology, 25, 335–349.

Baruch, Y. (1999). Response rates in academic studies—a comparative analysis. Human Relations, 52, 421–434.

Baruch, Y., & Holtom, B. C. (2008). Survey response rate levels and trends in organizational research. Human Relations, 61, 1139–1160.

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

Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in web- or Internet-based surveys. Educational and Psychological Measurement, 60, 821–836.

Deming, W. E., & Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Annals of Mathematical Statistics, 11, 427–444.

Han, K. T. (2007). WinGen: Windows software that generates IRT parameters and item responses. Applied Psychological Measurement, 31(5), 457–459.

Kessler, R. C., Avenevoli, S., Costello, E. J., Green, J. G., Gruber, M. J., Heeringa, S., … Zaslavsky, A. M. (2009). National comorbidity survey replication adolescent supplement (NCS-A): II. Overview and design. Journal of the American Academy of Child & Adolescent Psychiatry, 48, 380–385.

Kish, L. (1965). Survey sampling. New York, NY: Wiley.

Kraut, A. I. (Ed.). (1996a). Organizational surveys. San Francisco, CA: Jossey-Bass.

Kraut, A. I. (1996b). Planning and conducting the survey: Keeping strategic purpose in mind. In A. Kraut (Ed.), Organizational surveys: Tools for assessment and change (pp. 149–203). San Francisco, CA: Jossey-Bass.

Krosnick, J. (1999). Survey research. Annual Review of Psychology, 50, 537–567.

Landers, R. N., & Behrend, T. S. (2015). An inconvenient truth: Arbitrary distinctions between organizational, Mechanical Turk, and other convenience samples. Industrial and Organizational Psychology: Perspectives on Science and Practice, 8, 142–164.

Lohr, S. L. (1999). Sampling: Design and analysis. Pacific Grove, CA: Duxbury Press.

Ludbrook, J., & Dudley, H. (1998). Why permutation tests are superior to *t* and *F* tests in biomedical research. American Statistician, 52, 127–132.

Macey, W. H. (1996). Dealing with the data: Collection, processing, and analysis. In A. Kraut (Ed.), Organizational surveys: Tools for assessment and change (pp. 204–232). San Francisco, CA: Jossey-Bass.

Mallett, D. (2006). Sampling and weighting. In R. Grover & M Vriens (Eds.), The handbook of marketing research: Uses, misuses, and future advances (pp. 159–177). Thousand Oaks, CA: Sage.

Maletta, H. (2007). Weighting. Retrieved from http://www.spsstools.net

Martin, C. L. (1994). The impact of topic interest on mail survey response behavior. Journal of the Market Research Society, 36, 327–338.

Quine, S., & Morrell, S. (2008). Feeling safe in one's neighbourhood: Variation by location among older Australians. Australian Journal of Rural Health, 16, 115–116.

Rivers, D., & Bailey, D. (2009). Inference from matched samples in the 2008 U.S. national elections. Proceedings of the Survey Research Methods Section of the American Statistical Association, 627–639.

Rogelberg, S. G. (2006). Understanding nonresponse and facilitating response to organizational surveys. In A. I. Kraut (Ed.), Getting action from organizational surveys: New concepts, methods, and applications (pp. 312–325). San Francisco, CA: Jossey-Bass.

Rogelberg, S. G., Conway, J. M., Sederburg, M. E., Spitzmüller, 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.

Rogelberg, S. G., Spitzmüeller, C., Little, I., & Reeve, C. L. (2006). Understanding response behavior to an online special topics organizational satisfaction survey. Personnel Psychology, 59, 903–923.

Rogelberg, S. G., & Stanton, J. M. (2007). Understanding and dealing with organizational survey nonresponse. Organizational Research Methods, 10, 195–209.

Sterba, S. K. (2009). Alternative model-based and design-based frameworks for inference from samples to populations: From polarization to integration. Multivariate Behavior Research, 44, 711–740.

Tourangeau, R., Groves, R. M., & Redline, C. D. (2010). Sensitive topics and reluctant respondents: Demonstrating a link between nonresponse bias and measurement error. Public Opinion Quarterly, 74, 413–432.

Visser, P. S., Krosnick, J. A., Marquette, J., & Curtin, M. (1996). Mail surveys for election forecasting? An evaluation of the Columbus Dispatch poll. Public Opinion Quarterly, 60, 181–227.

Weisberg, H. F. (2005). The total survey error approach: A guide to the new science of survey research. Chicago, IL: University of Chicago Press.