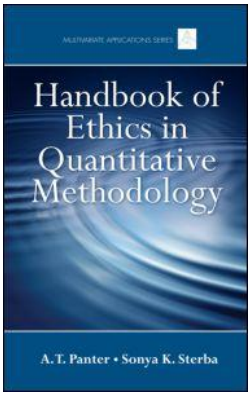
This article was downloaded by: *10.3.98.93*
On: 23 Oct 2018
Access details: *subscription number*
Publisher: *Routledge*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



Handbook of Ethics in Quantitative Methodology

A.T. Panter, Sonya K. Sterba

Beyond Treating Complex Sampling Designs as Simple Random Samples: Data Analysis and Reporting

Publication details
https://www.routledgehandbooks.com/doi/10.4324/9780203840023.ch10
Sonya K. Sterba, Sharon L. Christ, Mitchell J. Prinstein, Matthew K. Nock
Published online on: 20 Jan 2011

How to cite :- Sonya K. Sterba, Sharon L. Christ, Mitchell J. Prinstein, Matthew K. Nock. 20 Jan 2011, *Beyond Treating Complex Sampling Designs as Simple Random Samples: Data Analysis and Reporting from:* Handbook of Ethics in Quantitative Methodology Routledge
Accessed on: 23 Oct 2018
https://www.routledgehandbooks.com/doi/10.4324/9780203840023.ch10

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: https://www.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

Section IV

Ethics and Data Analysis Issues

TAF-Y101790-10-0602-S004.indd 265 12/3/10 10:10:49 AM

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

TAF-Y101790-10-0602-S004.indd 266 12/3/10 10:10:49 AM

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

10

Beyond Treating Complex Sampling Designs as Simple Random Samples: Data Analysis and Reporting

Sonya K. Sterba
Vanderbilt University

Sharon L. Christ
Purdue University

Mitchell J. Prinstein
University of North Carolina at Chapel Hill

Matthew K. Nock
Harvard University

This chapter addresses two issues: (a) how the method for selecting the sample ought to be reported in observational research studies, and (b) whether and when the sample selection method needs to be accounted for in data analysis. This chapter reviews available methodological and ethical guidelines concerning each issue and considers the extent to which these recommendations are heeded in observational psychological research. Discussion focuses on potential ethical implications of the gap between available methodological recommendations and current practice. A hypothetical case example and also a real world case example involving a daily diary study are used to demonstrate some alternative strategies for narrowing this gap.

It is important to note that both of the issues taken up in this chapter (reporting and accounting for sample selection in data analysis) arise after the sampling method has already been chosen. In contrast, a chapter on ethics and sampling in observational studies might have been expected to mainly concern the sample selection method itself—particularly whether a random (probability) or nonrandom (nonprobability) sample should

TAF-Y101790-10-0602-C010.indd 267 12/4/10 9:35:50 AM

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

be drawn.[1] The latter topic has long dominated informal discussions of ethics and sampling among social scientists, but has also often been misunderstood. Moreover, debate over choosing between probability versus nonprobability sampling has often led to an impasse, where observational researchers in particular fields (e.g., psychology) find only one sampling method pragmatically feasible (nonprobability sampling), and other fields (e.g., public health) find only one method statistically defensible (probability sampling; see Sterba, 2009). Our strategy is to begin with a brief overview of current and past perspectives on this controversial topic. The issues we address in this chapter are very general; they are relevant to whatever (probability or nonprobability) sample was selected. However, in discussing these issues in later sections, we periodically highlight relevant costs or benefits of using a probability versus nonprobability sampling method.

Random and Nonrandom Sample Selection

When sampling was first proposed as an alternative to census taking, a distinction was drawn between two different methods for selecting samples from populations: probability (or random) sampling and nonprobability (or nonrandom) sampling (Bowley, 1906; Kaier, 1895). In probability sampling, the probability of selection for all units in the target population is known and nonzero. In nonprobability sampling, the probability of selection for some units is unknown, and possibly zero, and the finite, target population may be only loosely identified. Whereas early methodological debates sought to establish one method as superior and the other as uniformly unacceptable (Neyman, 1934; Stephan, 1948), such definitive conclusions were never reached despite extensive dialogues on the topic (see Royall & Herson, 1973; Smith, 1983, 1994; Sugden & Smith, 1984).

To summarize this debate briefly, collecting a probability sample by definition requires that key selection variables are observed and that the *selection mechanism* (i.e., the mechanism by which sampling units get from a finite population into the observed sample) is well understood. Both aspects in turn reduce the risk that selection on unmeasured, unobserved variables will bias results. Furthermore, the randomness entailed

[1] Note that the issues that arise when deciding between random versus nonrandom *assignment* in treatment settings (e.g., Mark & Lenz-Watson, Chapter 7, this volume) are meaningfully different from those that arise when deciding between random versus nonrandom *selection* in observational (or experimental) settings, although there are certain parallels (Fienberg & Tanur, 1987).

TAF-Y101790-10-0602-C010.indd 268 12/4/10 9:35:50 AM

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

in a probability selection mechanism—specifically the fact that sampled and unsampled outcomes are assigned known probabilities—means that a distribution constructed from these probabilities can serve as the sole basis of inference to a finite population, without invoking strong modeling assumptions (e.g., Cassel, Sarndal, & Wretman, 1977). In contrast, nonprobability samples rely heavily on modeling assumptions to facilitate inference to a larger population, which is hypothetical. Nevertheless, should these modeling assumptions be met, there is a well-established statistical logic for inference from nonprobability samples (see Sterba, 2009, for a review of this logic). Hence both sampling methods have been recognized—initially at the 1903 Consensus Resolution of the International Statistical Institutes—and both are still frequently used.[2] Much attention has since turned to the two issues considered here: (a) what to report about sample selection, and (b) whether and when to account for sample selection in data analysis.

Reporting About Sample Selection

Methodological Guidelines

For the issue of reporting about sample selection, our review necessarily takes a historical perspective because reporting guidelines have been in existence for a long time, yet have evolved considerably. The first methodological recommendations on reporting practices appeared almost immediately after the practice of sampling was first introduced. The International Statistical Institute's 1903 Consensus Resolution called for "explicit account in detail of the method of selecting the sample" in research reports (Kish, 1996, p. 8). Similar recommendations were made in the proceedings of subsequent meetings, such as: "the universe from which the selection is made must be defined with the utmost rigour," and "exactness of definition" is needed for "rules of selection" (Jensen, 1926, pp. 62–63). The nonspecificity of these guidelines, however, led to inconsistent reporting practices.

By the 1940s, mounting dissatisfaction over inconsistent reporting practices led the United Nations (UN) Economic and Social Council to convene a Subcommission on Statistical Sampling that met throughout the decade to develop a common terminology for such reporting (UN, 1946, 1947, 1948, 1949a). This Subcommission resulted in the formalized

[2] This chapter pays specific attention to nonprobability (nonrandom) samples because they are most often used by psychologists.

TAF-Y101790-10-0602-C010.indd 269 12/4/10 9:35:50 AM

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

"Recommendations Concerning the Preparation of Reports of Sample Surveys" (UN, 1949b, 1949c). These recommendations highlighted the importance of reporting: (a) the *sampling units*; (b) the *frame*; and (c) the method of selecting (or recruiting) units—which may include (d) whether and how the frame was *stratified* before selection, (e) whether units were selected in *clusters*, (f) whether units were selected with *equal or unequal probabilities of selection*, and (g) whether units were selected in *multiple phases*. Also highlighted were reporting (h) *sample size*; (i) rates of refusals and attrition (see Enders & Gottschall, Chapter 14, this volume); (j) suspected areas of undercoverage of the frame; (k) methods undertaken after sample selection to gain insights into reasons for refusals and attrition; and (l) how the sample composition corresponds to preexisting survey data (e.g., census data). Table 10.1 provides definitions and brief examples of the italicized terms. Taken together, when a sample involves stratification, clustering, and/or disproportionate selection probabilities, it is conventionally called a *complex* sample, and those three key features are called *complex sampling features*. Sampling designs that lack all three features can be called *simple* (hence the term *simple random sample*).

In the 4 decades after their introduction, the UN guidelines had a limited impact on reporting practices, particularly in the social sciences. Indeed, a review of reporting practices from 1940–1979 found that instead of using the concrete terminology for describing sample selection as shown in Table 10.1, researchers often simply labeled their samples "representative" with little or no empirical substantiation (Kruskal & Mosteller, 1979a, b). That is, the descriptor "representative" was often used to provide "general, unjustified acclaim for the data," which Kruskal and Mosteller (1979b) equated to stating, "My sample will not lead you astray; take my word for it even though I give you no evidence … these data just happened to come to my hand, and I have no notion of the process that led to them or of relations between the target and sampled population" (p. 114–115). Moreover, Kruskal and Mosteller (1979b) found that the application of the term *representative* was itself ambiguous. Sometimes the term was meant to imply that sampling units were "typical cases" from a population; other times the term was used to convey that the sampling method provided "adequate coverage of population heterogeneity." In contrast to simply labeling a sample representative, the terms recommended by the UN Subcommission are less value-laden and communicate more precise information about the sample selection mechanism.

Ethical Guidelines

Guidelines for reporting about sample selection began to move from the purely methodological sphere to the ethical sphere in the 1980s.

TAF-Y101790-10-0602-C010.indd 270 12/4/10 9:35:50 AM

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

TAF-Y101790-10-0602-C010.indd 271

TABLE 10.1

Some Terms Useful for Reporting About Sample Selection

Term	Definition	Examples
Sampling units	The physical units that were selected.	Persons, schools, divorce records, accident reports.
Sampling frame	All sampling units that had a nonzero probability of being selected into the sample.	A list of daycare centers in a community; all persons with registered university e-mail addresses; birth records from a particular county within a 2-month period.
Stratified sampling	Independently selecting sampling units from mutually exclusive groups, or strata, which may be preexisting or artificially defined.	Schools could be stratified into public vs. private; patients could be stratified into inpatient vs. outpatient; Alzheimer facilities could be stratified into nursing homes vs. assisted-living centers.
Cluster sampling	Using entire groups as sampling units, in lieu of individual elements, at one or more stages of selection.	Schools might be sampling units at a *primary stage* of selection; classes within schools might be sampling units at a *secondary stage* of selection; students within class might be sampling units at a *tertiary stage* of selection. Here classes and schools represent clusters of the *ultimate sampling unit*: students. Or, schools might be sampling units at a primary stage of selection, but all classes and all students are included within selected schools. This constitutes one, not three, *stages of selection*.
Multiple phases of selection	Used when a frame containing values on desired selection variable(s) is unavailable. In a two-phase design, the first phase of selection entails collecting these values from a large sample of units, which in turn constitutes the frame for the second phase of the study.	See hypothetical case example, in a later section of the chapter, for a detailed example. (Note that phases of selection, described here, are different than stages of selection, described above.)
Equal or unequal probabilities of selection	Whereas equal or unequal selection probabilities can be achieved with a probability sample, in nonprobability samples, units are typically selected with unequal probabilities on observed and/or unobserved variables; the main question then becomes if the selection variable(s) are, for example, (a) independent variable(s), (b) dependent variable(s), or (c) design variables that conditionally correlate and/or interact with independent variables, while predicting the outcome; our shorthand is to refer to (b) and (c) as *disproportionate selection*.	Equal selection probabilities for mice in a one-stage cluster sample of $j = 1 \ldots J$ litters could involve selecting litter j with probability = (litter j size)/(total number of mice in the frame), and then including all mice within selected litters. Disproportionate selection could involve selecting parents based on parental income to study the effects of parental monitoring on child academic performance (where income is a design variable omitted from analyses, income and monitoring are correlated, and income predicts academic performance).

12/4/10 9:35:50 AM

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

Surprisingly, this shift was largely not spurred by methodologists wanting to speed the sluggish adoption of the UN guidelines, but was rather the result of external pressures. After a series of highly publicized research scandals in the late 1970s and early 1980s (chronologically reviewed by Mitcham, 2003), Congress held several hearings on research ethics. These hearings resulted in the creation of federal offices to oversee the promotion of research integrity, to facilitate the publication of research misconduct regulations, and to encourage scientific societies to pay greater attention to ethics—particularly in the form of ethics codes. Subsequently, some specifics from the UN's methodological guidelines for reporting about sample selection were incorporated into several societal standards or ethics codes (e.g., the Council of American Survey Research Organization's [CASRO] *Code of Standards and Ethics*, 2006; the American Psychological Association's [APA] *Statistical Methods in Psychology Journals: Guidelines and Explanations* [Wilkinson & the Task Force on Statistical Inference, 1999]; and the American Association for Public Opinion Research's [AAPOR] *Code of Professional Ethics & Practices*, 1991–2005). However, no specific UN guidelines were incorporated into other codes (e.g., the International Statistical Institute's [ISI] *Declaration on Professional Ethics*, 1985–2009; the APA's *Ethical Principles of Psychologists and Code of Conduct* [APA, 2002]; and the American Statistical Association's [ASA] *Ethical Guidelines for Statistical Practice*, 1983–1999).[3]

Current Practice

Between the more thorough methodological recommendations and less thorough ethical guidelines, resources on reporting about sample selection are now quite extensive. Still, a recent review of 10 observational studies in 2006 issues per each of four highly cited psychology journals (*Developmental Psychology*, *Journal of Personality and Social Psychology*, *Journal of Abnormal Psychology*, and *Journal of Educational Psychology*) found that 50 years of international methodological guidelines regarding how to report on sample selection (plus recent ethical guidelines) were not enough to routinely ensure adequate reporting practices in top-tier psychology journals (Sterba, Prinstein, & Nock, 2008). Of the 76% of studies that were nonprobability samples, only 23% described the method of selecting units (recruitment process), and only 52% reported anything about the

[3] For example, although the ASA's 1999 guidelines mentioned the general need to "explain the sample(s) actually used" (C5), the need to "include appropriate disclaimers" "when reporting analyses of volunteer data or other data not representative of a defined population" (C11) and the need to disclose consequences of failing to follow-through on an agreed sampling plan (C12), these guidelines still lack specifics about what sample selection features should be reported, and how.

TAF-Y101790-10-0602-C010.indd 272 12/4/10 9:35:50 AM

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

sampled population. For the 24% of studies that were probability samples, corresponding figures were better: 89% and 100%, respectively.

Hence although recommended reporting practices have been included in several societal ethics codes and standards, this has not ensured their adoption in practice. We suggest two potential reasons why. First, the presence of material on reporting about sample selection was inconsistent from one ethics code to the next. Efforts to standardize the inclusion of reporting recommendations could provide a more coherent reference source for applied researchers. Second, none of the codes or standards provided an explicit rationale for whether, and if so why, reporting is indeed an ethical issue, not simply a methodological issue. It is odd to expect an ethical imperative to improve reporting practices without providing a motivating explanation.

Is Reporting About Sample Selection an Ethical Issue?

There are several reasons why the gaps highlighted here between applied practice and methodological recommendations go beyond a purely methodological issue and into an ethical issue (e.g., negligence). These gaps are an ethical issue because researchers have the resources and ability to do something about them, but unintentionally have not, which leads to undesirable or even harmful consequences. This thesis is consistent with what is informally called the *ought implies can principle*: establishing that someone can do something is required before holding them accountable for doing it. Psychologists presently have the means to narrow the methods–practice gap regarding reporting about sample selection. Methodological recommendations and guidelines on reporting about sample selection have been available for an extremely long period—50 years—much longer than it typically takes a methodological advance to soak into applied practice. Additionally, the effort needed to implement recommended reporting practices is slim and does not require lengthy technical training. So the *ought implies can principle* is satisfied. Further, the consequences of adequate reporting about sample selection are important. Identifying and reporting complex sampling features that were used, such as those listed in Table 10.1, is a prerequisite first step before one can move on to determine whether these features need to be accounted for in data analyses. That is, if too little attention is paid to accurately reporting about sample selection, a researcher's ability to adequately account for the sample selection mechanism in data analysis is limited. Similarly, a reviewer's ability to crosscheck whether the analysis fully accounts for sample selection is limited. In turn, adequately accounting for the sample selection mechanism is necessary to ensure the validity of statistical inferences, as explained in the next section. When the validity of statistical inferences is in question, so are substantive conclusions based on those inferences.

TAF-Y101790-10-0602-C010.indd 273 12/4/10 9:35:50 AM

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

In the last section of this chapter we make suggestions for narrowing this methods–practice gap in reporting practices, with the aid of this ethical imperative.

Statistically Accounting for Sample Selection

Methodological Guidelines

In contrast to the first issue we considered (reporting), the second issue we consider (statistically accounting for sample selection in data analysis) has not been translated into accessible international methodological guidelines, nor even widely disseminated beyond the more technical statistical literature. Nonetheless, the topic of when and how to statistically account for sample selection is no less important than reporting—and arguably more so. Sample selection impacts inference because the particular sample selection mechanism chosen can constrain the population to which inferences can be made. However, analytic techniques that incorporate sample design features can broaden the population of inference. Hence this section provides an accessible introduction to existing recommendations on this topic from within the statistics literature.

The following practical guidance on when and how to account for the sample selection mechanism was gleaned from recommendations within the statistics literature. When the sample selection mechanism involves complex sampling features—(a) clustering, (b) stratification, and/or (c) disproportionate selection of sampling units (e.g., using selection variables that correlate or interact with independent variables and predict the outcome)—these features typically need to be accounted for in statistical analyses (Skinner, Holt, & Smith, 1989). To account for this kind of disproportionate selection, selection and recruitment variables can be entered as covariates in the model and allowed to interact with independent variables (and/or can be incorporated into the model estimation using sampling weights, if a probability sample was used). Biemer and Christ (2008), Pfeffermann (1993, 1996), and Sterba (2009) provide examples and procedures for this approach. Further, to account for stratification and clustering, stratum indicators can be entered as fixed effects and cluster indicators may be entered as random effects in a multilevel model (or incorporated into sandwich-type standard error estimation adjustments for a single-level model). Chambers and Skinner (2003), Lohr, (1999), and Skinner et al. (1989) give examples and procedures for this second approach.

Moreover, if these complex sampling features are not accounted for in data analysis, there can be direct consequences for the validity of

TAF-Y101790-10-0602-C010.indd 274 12/4/10 9:35:51 AM

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

statistical inferences. When stratification is not accounted for, standard errors are typically upwardly biased, and when clustering is not accounted for, standard errors are often downwardly biased (e.g., Kish & Frankel, 1974). When disproportionate selection is unaccounted for, point estimates and standard errors can both be biased (e.g., Berk, 1983; Smith, 1983; Sugden & Smith, 1984).

In the context of the complex sample features used in a given study, researchers and journal reviewers may find it helpful to try to mentally classify a study's sample selection mechanism according to a taxonomy developed by Little (1982) and Rubin (1983).[4] This taxonomy classifies sample selection mechanisms as ignorable, conditionally ignorable, or nonignorable. Each taxon poses different implications for the validity of inferences when the sample selection mechanism is or is not accounted for.

Ignorable Sample Selection

Any time the probability of selecting sampling units is proportionate to the rate at which those units appear in the frame,[5] and sampling units are neither stratified nor clustered, the sample selection mechanism is ignorable and does not need to be accounted for in the data analysis. One sampling mechanism that is always ignorable is a simple random sample.

Conditionally Ignorable Sample Selection

When some complex sampling features are used, but these features are properly accounted for in data analysis, as described previously, the sampling mechanism can be thought of as conditionally ignorable. A selection mechanism rendered conditionally ignorable by the data analysis will not result in biased parameter estimates or standard errors, and thus will not affect the validity of inferences.

Nonignorable Sample Selection

Consider instead the circumstance in which sampling units are again selected with (a) clustering, (b) stratification, or (c) disproportionate

[4] Closely related versions of this taxonomy have been used to describe not only sample selection mechanisms but also missing data mechanisms. All versions stem from Rubin (1976). That is, the criteria used to determine whether we need to statistically account for the process by which persons entered the sample (i.e., sample selection mechanism) are similar to the criteria used to determine whether we need to statistically account for the process by which persons or observations are missing from the sample (i.e., missing data mechanism; see Enders & Gottschall, Chapter 14, this volume).

[5] Here we are assuming no frame error (e.g., over- or undercoverage) that would make the frame systematically different than the inferential population.

TAF-Y101790-10-0602-C010.indd 275 12/4/10 9:35:51 AM

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

selection probabilities. Furthermore, suppose that some selection variables, stratum indicators, and/or cluster indicators are partially *un*observed, or *un*recorded. This would prevent their complete incorporation into the model (and/or complete incorporation into estimation-based weighting and standard error adjustments).[6] Or, suppose that selection variables, stratum indicators, and/or cluster indicators are fully observed but are simply omitted from the model specification and/or estimation. Under either circumstance, the sample selection mechanism is *nonignorable*, meaning that it may result in biased parameter estimates or standard errors in the data analysis, and thus may affect the validity of inferences.

It can be seen from this taxonomy that classifying a sample selection mechanism as *ignorable*, *conditionally ignorable*, or *nonignorable* depends partially on how the sample was selected at the data collection stage, and partially on how the sample selection mechanism was statistically accounted for at the data analysis stage. Fully *ignorable* sample selection is rare; as previously mentioned, simple random samples and their equivalent would fall into this category. Achieving *conditionally ignorable* sample selection and avoiding *nonignorable* sample selection is the typical goal.

Ethical Guidelines

We earlier mentioned that methodological recommendations on reporting are more widely disseminated than methodological recommendations on when and how to statistically account for sample selection. Similarly, many ethical guidelines that did describe desirable reporting practices in detail are silent on the topic of statistically accounting for the sample selection mechanism (e.g., AAPOR, 2005; APA, 2002; CASRO, 2006; ISI, 1985–2009). The ethical guidelines that do comment on when and how to statistically account for sample selection are in some cases vague, which can limit their practical use. For example, ASA's (1999) ethical guideline A2 is to "Employ data selection or sampling methods and analytic approaches that are designed to assure valid analyses," and ethical guideline B5 is to "Apply statistical sampling and analysis procedures scientifically, without predetermining the outcome." In other cases, available societal standards are misleading. For example, Wilkinson and

[6] As mentioned previously, for probability samples these estimation adjustments can involve probability-weighted point estimators and stratified, between-cluster sandwich variance estimators. Our focus here is on nonprobability samples, where probability weights are unavailable, but sandwich variance estimators are available (yet less often used). An overview of these estimation adjustments is given in du Toit, du Toit, Mels, and Cheng (2005).

TAF-Y101790-10-0602-C010.indd 276 12/4/10 9:35:51 AM

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

the Task Force on Statistical Inference (1999, p. 595) imply that stratification and clustering need to be accounted for only in statistical models for probability (i.e., random) samples. But the same requirement applies to nonprobability samples as well. Furthermore, they made no mention of needing to statistically account for other complex sampling features besides clustering and stratification (e.g., disproportionate probabilities of selection). To be sure, when and how to statistically account for sample selection is a less straightforward topic than reporting. This fact may have discouraged the incorporation of the former topic into societal standards and/or ethics codes. Nevertheless, it seems safe to say that more concrete, less misleading statements could be made without glossing over the complexities of deciding when to account for sample selection and without oversimplifying the alternative approaches for how to account for sample selection in data analysis.

Current Practice

A common perception is that so little is known about selection mechanisms for typical nonprobability samples in psychology that the possibility of following the aforementioned methodological guidelines is precluded (e.g., Jaffe, 2005; Peterson, 2001; Sears, 1986). That is, it is thought impossible for selection mechanisms from typical nonprobability samples in psychology to be rendered conditionally ignorable by statistically controlling for complex sample selection features. However, Sterba et al.'s (2008) article review indicated that this may not be the case. They found that 28% of studies based on nonprobability samples used one or more discernible (observed) complex sampling features (stratification, clustering, or disproportionate selection), and the authors accounted for all of them in the statistical model (potentially *conditionally ignorable sample selection*).[7] Another 58% of studies had one or more discernible complex sampling feature(s) but did not account for all of them in the statistical model (potentially *nonignorable sample selection*). The remaining 14% of studies had no discernible complex sampling features (potentially *ignorable sample selection*).[8] Corresponding percentages for probability samples were 56%, 33%, and 11%, respectively. This review tells us that there is a gap between the data available on known complex sample selection features on the one hand, and the subsequent use of those data in analyses to account for sample selection on the other.

[7] Instances of clustering solely as a result of time within person were not counted toward this total.

[8] It would have been useful if these authors had explicitly stated whether any complex sampling features were used so that their sample selection mechanisms could have been more cleanly classified.

TAF-Y101790-10-0602-C010.indd 277 12/4/10 9:35:51 AM

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

That is, samples are being treated as if they were simple random samples despite the fact that they include complex sampling features. Put another way, researchers are often not fully capitalizing on the potential to render their sample selection mechanisms conditionally ignorable in their data analyses.

Is Statistically Accounting for Sample Selection an Ethical Issue?

Not only are specific recommendations on statistically accounting for sample selection included in few ethics codes, but also a motivating explanation is typically absent. Without consistent inclusion and without justification, it is unsurprising that this ethical imperative seems not to have greatly affected practice. One potential two-part justification for considering accounting for sample selection an ethical issue is given here. First, psychologists often have the means to narrow the methods–practice gap regarding accounting for sample selection in data analysis. That is, more data on complex sampling features are often collected than are ultimately used in analyses (see previous section). Furthermore, multiple commercial software programs capable of accounting for complex sampling features are available; some have been available for more than 15 years. See the online appendix of Sterba (2009) for a software review. Second, the real world consequences of bias induced by unaccounted for, complex sampling features can affect substantive conclusions; this in turn misdirects scientific understanding and federal grant spending and can waste participants' time (Sterba, 2006).

But there is no denying that sometimes psychologists' means are limited; sometimes not enough is known about the sample selection mechanism in nonprobability studies to be able to fully control for it in the data analysis. This is less often the case in probability samples, where the logistics of the sampling design require that all stratum indicators, cluster indicators, and selection variable scores on the frame are observed so they can be used to assign probabilities of selection to units. This fact is certainly a strength of probability sampling and is reason to prefer it where possible. However, even in nonprobability samples, risk of biased inferences can be minimized in certain ways by recording more information on the sample selection mechanism during data collection. Also, the effects of sample selection features that were partially unobserved can sometimes still be investigated in statistical analyses to ascertain how much they may be impacting substantive conclusions. We next consider a short, hypothetical case example that illustrates how the recording of information about sample selection can be improved. We subsequently consider a longer, empirical case example that illustrates one way to investigate the effects of partially observed selection features in a common daily diary study design.

TAF-Y101790-10-0602-C010.indd 278 12/4/10 9:35:51 AM

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

Strategies for Narrowing the Gap Between Methodological Recommendations and Practice: Case Examples

Hypothetical Case Example: Recording More Information About Sample Selection

For this hypothetical case example, suppose a researcher intends to collect a nonprobability, convenience sample in a community setting, and suppose the researcher wants to oversample adolescents with sleep problems. Convenience samples are often used by psychologists when a specific, nonreferred subpopulation is desired but a frame or listing of sampling units that includes the selection variable(s) (e.g., sleep problems) is unavailable. For example, to oversample adolescents with sleep problems, study advertisements typically mention the variable to be oversampled (i.e., sleep problems). Would-be participants self-select into the study based on their interest, incentives, and/or their own perceived elevation on the variables mentioned in the advertisement. They then may be included or excluded based on additional study criteria or to meet quotas of youth with and without sleep problems. The problems are that it is unclear from this design (a) what variables persons (self-)selected on and (b) at what rate persons are being over- or undersampled. If unobserved self-selection variables are correlated with independent variables in the analysis and are predictive of the outcome, parameter bias may result.

Suppose, however, that this researcher is open to collecting additional data to more fully understand the selection mechanism. Here we consider one relatively simple and inexpensive method for collecting additional data about sample selection: conversion of a convenience sample into a two-phase sample (see Pickles, Dunn, & Vazquez-Barquero, 1995, for a review of two-phase samples). After describing how this convenience sample can be converted into a two-phase sample, we describe how the two-phase sample to some extent circumvents problems (a) and (b) mentioned above.

In phase 1 of a two-phase design, a brief screening questionnaire including questions about sleep problems and any other desired inclusion and exclusion criteria would be administered cheaply to a larger number of units than the desired sample size. This phase 1 sample might be taken from an institution such as a community health clinic (e.g., all well-child visits to a community health clinic in a given month) or from public records (e.g., all marriage records in a certain county for a certain duration of time). The phase 1 screened sample becomes the frame for the phase 2 sample. That is, the point of the screen is to record scores on selection variable(s) (e.g., sleep problem scores) for units who will then constitute the phase 2 frame. This in turn allows phase 1 individuals to

TAF-Y101790-10-0602-C010.indd 279 12/4/10 9:35:51 AM

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

be allocated to nonoverlapping strata based on their screen responses (e.g., high sleep problems stratum, low sleep problems stratum). Then, at phase 2, participants can be randomly sampled from the high sleep problems stratum at a higher rate (e.g., 80%) than the low sleep problems stratum (e.g., 20%). Furthermore, the inverse of the selection probabilities in each stratum can be used as sampling weights in the data analysis phase to ensure that the phase 2 sample is statistically generalizable to the phase 1 sample.

The key improvement of the two-phase sample over the convenience sample is that selection from phase 1 to phase 2 is now based mainly on observed variables under the control of the researcher. These observed selection variables can then be used as covariates in the analysis or entered into weight variables in the analysis. In so doing, problems (a) and (b) from the convenience sample have now been circumvented for inference from the phase 2 sample to the phase 1 sample, even if generalizability from the phase 1 sample to an undefined larger population is still uncertain.[9] The sample selection mechanism for the phase 2 sample is thus conditionally ignorable. Another way of looking at the added advantage of the two-phase sample is that, to a much greater degree, it disentangles interest in participating from eligibility to participate; the collection of screening information at phase 1 is *not* contingent on interest in participating in phase 2.

During the implementation of the two-phase sampling design, the following information needs to be collected and later reported: (a) the proportion of persons refusing the phase 1 screen and, if possible, the reasons for refusal, recruitment mode, and basic demographic information; (b) the mode of recruitment for persons completing the phase 1 screen (e.g., newspaper, e-mail, flier); (c) the proportions of persons who were excluded after phase 1 and the reasons for their exclusion; and (d) the proportion of persons recruited into phase 2 who refused and their reasons for refusal. It is often helpful to present information for items (a)–(d) in a flowchart (see Sterba, Egger, & Angold, 2007, p. 1007, for an example).

Empirical Case Example: Investigating the Effects of Partially Observed Selection Features

The previous case example considered the circumstance where data had not yet been collected, such that the data collection method could be modified to record more detailed information about sample selection. For samples that have already been collected, this option is not available. Consider now the situation in which a study has already been completed,

[9] In a later section, we discuss procedures that could be used to gain some insight into the correspondence between the phase I sample and a particular target finite population.

TAF-Y101790-10-0602-C010.indd 280 12/4/10 9:35:51 AM

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

but some complex selection features were partially observed or partially recorded, raising the possibility of a nonignorable selection mechanism with accompanying parameter and standard error bias. Specifically, we consider the situation in which some selection variables were unobserved, but any strata and cluster indicators used were observed. In this situation, there are several possible approaches for investigating whether the sample is systematically different from a particular target finite population of interest for inference.

One approach was briefly mentioned earlier: Find a large-scale probability sample collected from the target finite population (e.g., general population survey or census) and compare it with the sample on key variables—particularly variables that were hypothesized to be involved in selection and were included in both data sets. Another approach involves applying intensive effort to recruit a small subsample of persons in the target finite population who initially refused contact, participation, or screening. Then compare their responses with participants on key variables. Groves (2006, p. 655–656) and Stoop (2004) discuss the first and second approaches in greater detail. A third approach involves applying a model-based *sensitivity analysis* to find out the extent to which the suspected nonignorability of the sample selection mechanism impacts substantive conclusions. In this context, a sensitivity analysis involves specifying at least one alternative model in addition to the theorized model of substantive interest. These alternative model(s) relax certain assumptions about the sample selection mechanism. Those assumptions were potentially violated in the original theorized model of substantive interest. Comparing alternative model(s) with the original theorized model, the researcher can see whether their substantive conclusions are sensitive to different assumptions about the sample selection mechanism. This third approach can often involve some time and cost savings over the previous two approaches; thus, it is the one we empirically illustrate here.

Our empirical case example uses Nock and Prinstein's nonprobability experience sampling (or daily diary) study of nonsuicidal self-injury (NSSI) behaviors. In this case example, responses were solicited at repeated assessments, and our interest lies in the validity of inferences from the subset of persons selected at each repeated assessment to the full, originally recruited sample.[10] Thus, this case example differs from previously discussed examples in that sample selection occurs more than once. This case example also differs from previously discussed examples in that inference to the originally recruited sample, rather than to a wider population, is desired. Validity of inference from the

[10] Because of the manner of selection at each time point (to be described shortly), we find it more intuitive to characterize this case example in terms of a sample selection problem, but it is possible to alternatively think of it as a missing data problem.

TAF-Y101790-10-0602-C010.indd 281 12/4/10 9:35:51 AM

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

original sample to a wider population would entail other analyses (e.g., comparisons using external finite population data) that are outside the scope of the present discussion (see Nock, Prinstein, & Sterba, 2009, for more information).

Sample Selection Mechanism

In this empirical case example, the full, originally recruited sample consisted of 30 adolescents. For 14 days, these 30 adolescents were exogenously signaled to respond with their context, feelings, thoughts, and behaviors related to NSSI at several points throughout the day (called *signal-contingent selection*) and were told to also respond about these matters specifically when they were having an NSSI thought (called *event-contingent selection*). Signal-contingent selection, event-contingent selection, and their combination are widely used methods of soliciting responses at repeated assessments in daily diary studies (Bolger, Davis, & Rafaeli, 2003; Ebner-Priemer, Eid, Kleindienst, Stabenow, & Trull, 2009; Shiffman, 2007; Wheeler & Reis, 1991). Specifically, in signal-contingent selection, participants are prompted to respond by an external device that is preprogrammed to signal at fixed or varying time intervals. In contrast, in event-contingent selection, responses are solicited based on the current behavior, feelings, context, or thoughts of the participant. Event-contingent selection has been particularly recommended for rare or highly specific experiences, including interpersonal conflict, intimacy, alcohol consumption, and mood (Bolger et al., 2003; Ebner-Priemer et al., 2009). Event-contingent selection was used in the case example because NSSI is a rare experience.

In this case example, the "event" is the dependent variable itself, NSSI thought (which differs from Nock et al., 2009). Thus, the selection mechanism is suspected to be nonignorable. In this context, nonignorability practically means that the effects of independent variables on the propensity to have an NSSI thought are confounded with the effects of independent variables on the propensity to self-report. It may be the case that different covariates, or different levels of the same covariates, predict propensity to self-report versus propensity to have an NSSI thought, if we could tease those two processes apart. Yet, even in this worst-case scenario, a sensitivity analysis can be conducted to see whether this potentially nonignorable selection method meaningfully impacts results. We will see later that this sensitivity analysis capitalizes on the fact that a *combination* of event- and signal-contingent selection was used. The sensitivity analysis demonstrated here (proposed in Sterba et al., 2008, and used in Nock et al., 2009) adapts what has been termed a *shared parameter model* (Follmann & Wu, 1995; Little, 1995) or a *two-part model* (Olsen & Shafer, 2001) for the case of sample selection. These models have some similarities to traditional

TAF-Y101790-10-0602-C010.indd 282 12/4/10 9:35:52 AM

cross-sectional single-level selection models (e.g., Heckman, 1979) but are less restrictive.

Sensitivity Analysis Step 1

The first step in this sensitivity analysis is to specify our model of substantive theoretical interest as per usual; let us call it our outcome-generating model. This model ignores whether the response was self-selected (i.e., event-contingent) or signal-driven (i.e., signal-contingent). In our outcome-generating model, independent variables of interest at level 1 (observation level) are whether the participant was currently using drugs (*drug*), feeling rejected (*reject*), feeling sad (*sad*), feeling numb (*numb*), and whether they were with peers (*peer*). Independent variables of interest at level 2 (person level) are *age* and *sex*. In the specification of this outcome-generating model, the nesting of responses within an individual is accounted for using a multilevel model with a random intercept.[11]

Specifically, the outcome model predicting binary NSSI thoughts is:

$$\log\left[\frac{\Pr(thought_{ij}=1)}{1-\Pr(thought_{ij}=1)}\right] = \gamma^{o}_{00} + \gamma^{o}_{10}drug_{ij} + \gamma^{o}_{20}reject_{ij} + \gamma^{o}_{30}sad_{ij} + \gamma^{o}_{40}numb_{ij} + \gamma^{o}_{50}peer_{ij} + \gamma^{0}_{01}sex_{j} + \gamma^{o}_{02}age_{j} + u^{o}_{0j} \quad (10.1)$$

where the superscript *o* denotes the outcome equation, *i* denotes observation, and *j* denotes person. The γ represents a fixed effect, the *u* represents a random effect, and the random intercept variance is estimated $u^{o}_{0j} \sim N(0, \tau^{o})$. This multilevel model can also be portrayed graphically using Curran and Bauer's (2007) path diagrammatic notation, as in Figure 10.1. Drug use, rejection, sadness, and numbness were hypothesized to be positively related to NSSI thoughts, and being with peers was hypothesized to be negatively related to NSSI thoughts. Sex and age were control variables. Table 10.2, column 1, shows that only the hypotheses about rejection and sadness were supported.

Sensitivity Analysis Step 2

Estimates in Table 10.2, column 1, could be biased if the sample selection mechanism is not independent from the outcome-generating mechanism (i.e., if it is nonignorable, as we suspect). That is, if the selection and

[11] Checks for autocorrelation, cyclicity, and trend were described in Nock et al. (2009), and little evidence of each was found. For simplicity, these checks are not discussed here. A three-level model (responses nested within day nested within person) encountered estimation problems as a result of little day-to-day variability in NSSI thoughts; the day level was therefore dropped.

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

TAF-Y101790-10-0602-C010.indd 283 12/4/10 9:35:53 AM

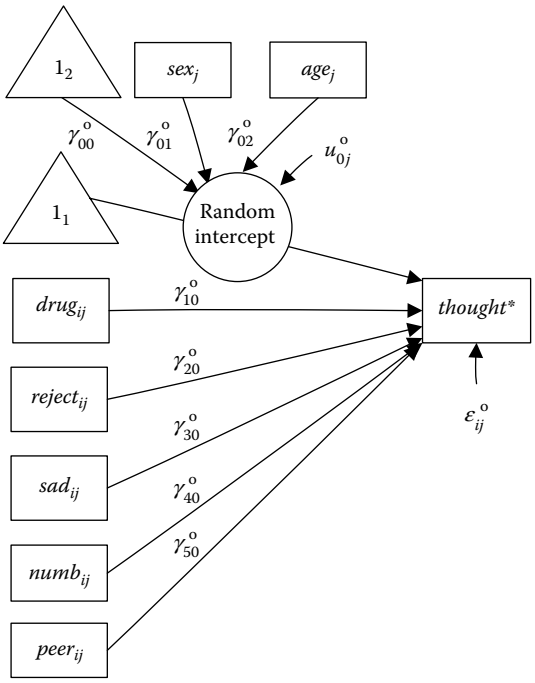


FIGURE 10.1
Path diagram for empirical case example: outcome model only, ignoring selection. Squares are measured variables. Circles are latent coefficients. Triangles are constants. Straight arrows are regression paths. Symbols are defined in the text equations. The multilevel model path diagram framework used here was introduced in Curran and Bauer (2007).

outcome-generating mechanisms are dependent, the effect of a predictor on the outcome is confounded with the effect of the predictor on the probability of selection. Therefore, the slope coefficients in Table 10.2, column 1, would simultaneously represent both effects. To investigate whether the potentially nonignorable selection is affecting estimates in Table 10.2, column 1, we need to specify not just an outcome-generating submodel, as per usual, but also a model for the sample selection mechanism (let us call it a selection model). Then we need to assess the extent to which these two models are interdependent. In the selection submodel, we are predicting the log odds of self-initiated response (*selection* = 1) versus a signal-initiated response (*selection* = 0). It was hypothesized that persons would be more likely to self-select if they were not with peers, were female, were feeling less numb, and were feeling more rejected.

$$\log\left[\frac{\Pr(selection_{ij}=1)}{1-\Pr(selection_{ij}=1)}\right] = \gamma^{s}_{00} + \gamma^{s}_{10}reject_{ij} + \gamma^{s}_{20}numb_{ij} + \gamma^{s}_{30}peer_{ij} + \gamma^{s}_{01}sex_{j} + u^{s}_{0j} \qquad (10.2)$$

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

TABLE 10.2

Empirical Case Example Results: Sensitivity Analysis for Nonignorable Selection in a Daily Diary Study

	Model 1: Ignoring Selection		Model 2: Accounting for Selection	
	Estimate (SE)	***p* Value**	**Estimate (SE)**	***p* Value**
Outcome (sub)model:				
Fixed effects				
Intercept	0.764 (2.720)	.779	1.080 (2.539)	.671
Using drugs	0.304 (0.200)	.129	0.328 (0.103)	.001
Feeling rejected	1.108 (0.482)	.022	1.055 (0.474)	.026
Feeling sad	0.665 (0.274)	.015	0.671 (0.266)	.012
Feeling numb	−0.561 (0.277)	.043	−0.548 (0.273)	.044
With peers	0.014 (0.375)	.970	0.015 (0.350)	.965
Age	−0.050 (0.112)	.658	−0.081 (0.144)	.571
Sex	0.529 (0.636)	.406	0.707 (0.699)	.312
Variance components				
τ^o	3.355 (1.270)	.008	2.726 (0.885)	.002
Selection submodel:				
Fixed effects				
Intercept			−0.169 (0.674)	.802
Feeling rejected			1.340 (0.336)	.000
Feeling numb			0.098 (0.416)	.813
With peers			−0.748 (0.339)	.027
Sex			0.814 (0.377)	.031
Variance components				
τ^s			0.125 (0.116)	.283
$\tau^{o,s}$			0.288 (0.122)	.018

Estimates are in the logit scale. τ^o, intercept variance for the outcome (sub)model; τ^s, intercept variance for the selection submodel; $\tau^{o,s}$, covariance between the random intercepts in both (sub)models.

Here the superscript *s* stands for the selection submodel and $u^s_{0j} \sim N(0, \tau^s)$. Note that the selection and outcome submodels can have the same or different covariates (Follmann & Wu, 1995). It was hypothesized that, controlling for these observed covariates, the probability that *selection* = 1 would still be dependent on the probability that *thought* = 1 because the self-initiated nature of the event-contingent responding is partially dependent on the presence of a thought. However, this dependency is now accounted for by simultaneously estimating the outcome and selection submodels, and by allowing the individual deviations on *thought* to covary with the

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

TAF-Y101790-10-0602-C010.indd 285 12/4/10 9:35:55 AM

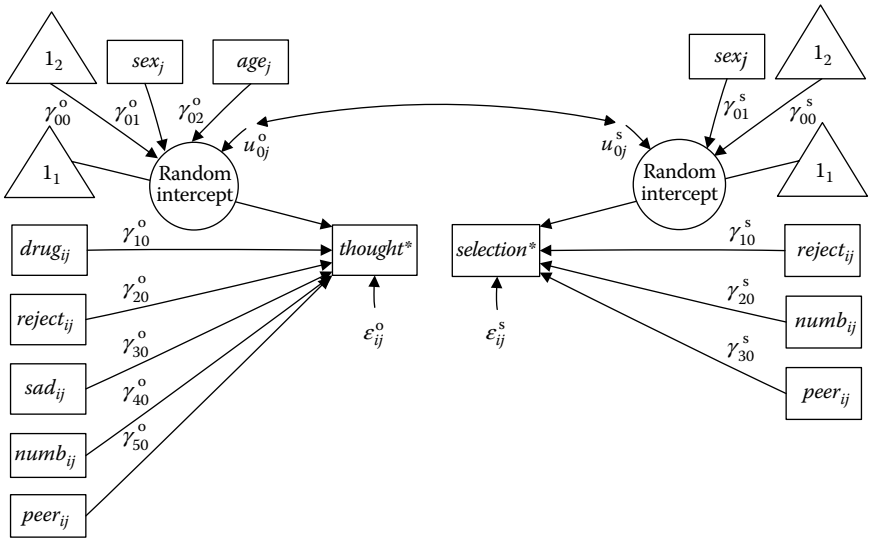


FIGURE 10.2
Path diagram for empirical case example: joint outcome and selection model. Curved arrows are covariances.

individual deviations on *selection*. That is, the intercept random effects for the outcome equation and the selection equation covary because of the term $\tau^{o,s}$:

$$\begin{matrix} u^{o}_{0j} \\ u^{s}_{0j} \end{matrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau^{o} & \\ \tau^{o,s} & \tau^{s} \end{pmatrix}\right]. \tag{10.3}$$

This joint model assumes that, *conditional on the random effect*, the outcome and selection processes are independent.[12] This assumption is more lenient than when we just estimated the outcome-generating model, in which selection and NSSI thoughts were assumed to be *unconditionally* independent. A path diagram for this joint outcome–selection model is given in Figure 10.2.

Sensitivity Analysis Results and Conclusions

Results of the joint outcome–selection model are shown in Table 10.2, column 2. Being alone, being female, and feeling rejected increased the probability of self-selection, as hypothesized. However, feeling numb did not

[12] This model is highly related to a random coefficient-dependent selection model. The latter model typically requires the correlation between the random intercept in the selection and outcome equations to be 1.0, whereas here we are freely estimating it.

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

decrease the probability of self-selection, as was hypothesized. In addition, the selection and outcome submodels are statistically dependent, controlling for observed covariates ($\tau^{o,s}$ = .288 [.122], p = .018). Even though our selection mechanism meets the technical definition of nonignorability, we are reassured to find that most of our substantive conclusions stay the same once we allow for the nonignorable selection. The only change occurred in the effect of drug use on the log odds of NSSI thoughts, which is now significant.

It is important to underscore that we called this approach a sensitivity analysis because we are not claiming that the Table 10.2, column 2, model is the *one true model* per se. The joint outcome–selection model rests on untestable assumptions about both the selection and outcome submodels and assumes that both submodels are properly specified—even though the researcher may be less confident about specifying the selection model (Little, 1995). We recommend specifying several theoretically compelling selection models (just as one would specify competing outcome models) and then investigating whether consistent results are found across perturbations in the selection model. Additional background and rationale for using selection models in sensitivity analyses can be found in Molenberghs and Verbeke (2005). In cases like this example, we recommend reporting (a) that there is evidence of a nonignorable selection mechanism; (b) that a sensitivity analysis was conducted; (c) whether and which parameter estimates differed when nonignorable selection was accounted for; and (d) whether these changes were found across alternative theoretically driven selection models.

Conclusion

In this chapter, we showed that publicized methodological and ethical guidelines have focused on the issue of reporting about sample selection more so than the issue of statistically accounting for sample selection in data analysis. Whereas this discrepancy may exist because the former issue is more straightforward to address, the former issue is certainly no more important than the latter. Further, we showed that a sizeable gap exists between methodological recommendations and applied practice for both issues. In response, we provided statistical rationales and ethical imperatives for why researchers ought to pay greater attention to both issues. Finally, we supplied two case examples illustrating certain ways this gap could be narrowed for particular nonprobability sampling designs.

Of course, the issues of incomplete reporting about sample selection and incomplete accounting for complex sampling features are just two

TAF-Y101790-10-0602-C010.indd 287 12/4/10 9:35:56 AM

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

of the methodologically and ethically important issues reviewed in the chapters of this book. However, we would argue that for psychologists these two issues are overlooked more often than some others considered in this book. For example, part of the culture of our discipline is to spend comparatively much less time dealing with sample selection issues in analysis and reporting than, say, measurement issues (whereas the reverse is true in other disciplines, like epidemiology; Sterba, 2009). Our recommendation for speeding the closure of the methods–practice gap is to make doing so a proximal priority, rather than a distal aspiration. Many ethics codes, including that of the APA, are primarily aspirational in nature. In contrast, medical journals have successfully elevated a number of methodological issues to proximal priorities by forming a cohesive International Committee of Medical Journal Editors and including these issues in their "Uniform Requirements for Manuscripts Submitted to Biomedical Journals" (see also Fidler, Chapter 17, this volume). The medical journal model is certainly worth further consideration by psychological journal editors as a stimulus for improving sample selection reporting and analysis practices. In this regard, it is worth emphasizing that *exclusively improving reporting practices*, as a first step, would likely spur subsequent improvements in data analysis practices as well. That is, simply identifying the complex sampling features that were used in the sampling design would alert readers and reviewers of the features that should have been accounted for in data analysis.

References

American Association for Public Opinion Research. (2005). *Code of professional ethics and practice*. Retrieved from http://www.aapor.org/aapor_code.htm

American Psychological Association. (2002). *Ethical principles of psychologists and code of conduct*. Retrieved from http://www.apa.org/ethics/code/index.aspx

American Statistical Association. (1999). *Ethical guidelines for statistical practice*. Retrieved from http://www.amstat.org/about/ethicalguidelines.cfm

Berk, R. A. (1983). An introduction to sample selection bias in sociological data. *American Sociological Review, 48,* 386–398.

Biemer, P., & Christ, S. (2008). Weighting survey data. In E. de Leeuw, J. Hox, & D. Dillman (Eds.), *International handbook of survey methodology* (pp. 317–341). New York: Erlbaum.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54,* 579–616.

Bowley, A. L. (1906). Address to the economic and statistics section of the British Association for the Advancement of Science, York, 1906. *Journal of the Royal Statistical Society, 69,* 540–558.

TAF-Y101790-10-0602-C010.indd 288 12/4/10 9:35:56 AM

Cassel, C., Sarndal, C., & Wretman, J. (1977). *Foundations of inference in survey sampling.* New York: Wiley.

Chambers, R. L., & Skinner, C. J. (2003). *Analysis of survey data.* Chichester, UK: Wiley.

Council of American Survey Research Organizations. (2006). *Code of standards and ethics for survey research.* Retrieved from http://www.casro.org/codeofstandards.cfm

Curran, P. J., & Bauer, D. J. (2007). A path diagramming framework for multilevel models. *Psychological Methods, 12,* 283–297.

du Toit, S. H. C., du Toit, M., Mels, G., & Cheng, Y., (2005). *Analysis of complex survey data with LISREL: Chapters 1–5.* Unpublished manual. Retrieved from http://www.ssicentral.com

Ebner-Priemer, U. W., Eid, M., Kleindienst, N., Stabenow, S., & Trull, T. J. (2009). Analytic strategies for understanding affective (in)stability and other dynamic processes in psychopathology. *Journal of Abnormal Psychology, 118,* 195–202.

Fienberg, S. E., & Tanur, J. M. (1987). Experimental and sampling structures: Parallels diverging and meeting. *International Statistical Review, 55,* 75–96.

Follmann, D., & Wu, M. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics, 51,* 151–168.

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

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica, 47,* 153–162.

International Statistical Institute. (1985–2009). *Declaration on professional ethics*. Retrieved from http://isi.cbs.nl

Jaffe, E. (2005). How random is that? *Association for Psychological Science Observer, 18,* 9.

Jensen, A. (1926). Report on the representative method in statistics. *Bulletin of* the International *Statistical Institute, 22,* 359–380. Extensive discussion on pp. 58–69, 185–186, and 212–213.

Kaier, A. N. (1895). Observations et expériences concernant des dénombrements représentatifs. *Bulletin of the International Statistical Institute, 9,* 176–183.

Kish, L. (1996). Developing samplers for developing countries. *International Statistical Review, 64,* 143–162.

Kish, L., & Frankel, M. R. (1974). Inference from complex samples (with discussion). *Journal of the Royal Statistical Society Series B, 36,* 1–37.

Kruskal, W., & Mosteller, F. (1979a). Representative sampling III: The current statistical literature. *International Statistical Review, 47,* 245–265.

Kruskal, W., & Mosteller, F. (1979b). Representative sampling II: Scientific literature, excluding statistics. *International Statistical Review, 47,* 111–127.

Little, R. J. A. (1982). Models for nonresponse in sample surveys. *Journal of the American Statistical Association, 77,* 237–250.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association, 90,* 1112–1121.

Lohr, S. L. (1999). *Sampling: Design and analysis*. Pacific Grove, CA: Brooks/Cole.

Mitcham, C. (2003). Co-responsibility for research integrity. *Science and Engineering Ethics, 9,* 273–290.

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

Molenberghs, G., & Verbeke, G. (2005). *Models for discrete longitudinal data*. New York: Springer.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society, 109,* 558–606.

Nock, M., Prinstein, M. J., & Sterba, S. K. (2009). Revealing the form and function of self-injurious thoughts and behaviors: A real-time ecological assessment study among adolescents and young adults. *Journal of Abnormal Psychology, 118,* 816–827.

Olsen, M. K., & Schafer, J. L. (2001). A two-part random-effects model for semi-continuous longitudinal data. *Journal of the American Statistical Association, 96,* 730–745.

Peterson, R. A. (2001). On the use of college students in social science research: Insights from a second order meta-analysis. *Journal of Consumer Research, 28,* 250–261.

Pfeffermann, D. (1993). The role of sampling weights when modeling survey data. *International Statistical Review, 61,* 317–337.

Pfeffermann, D. (1996). The use of sampling weights for survey data analysis. *Statistical Methods in Medical Research, 5,* 239–261.

Pickles, A., Dunn, G., & Vazquez-Barquero, J. L. (1995). Screening for stratification in two-phase epidemiological surveys. *Statistical Methods in Medical Research, 4,* 73–89.

Royall, R. M., & Herson, H. J. (1973). Robust estimation in finite populations I. *Journal of the American Statistical Association, 68,* 880–889.

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

Rubin, D. B. (1983). An evaluation of model-dependent and probability-sampling inferences in sample surveys: Comment. *Journal of the American Statistical Association, 78,* 803–805.

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology, 51,* 515–530.

Shiffman, S. (2007). Designing protocols for ecological momentary assessment. In A. A. Stone, S. Shiffman, A. A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 27–53). New York: Oxford University Press.

Skinner, C. J., Holt, D., & Smith, T. M. F. (1989). *Analysis of complex surveys*. New York: Wiley.

Smith, T. M. F. (1983). On the validity of inferences from non-random samples. *Journal of the Royal Statistical Society: Series A, 146,* 394–403.

Smith, T. M. F. (1994). Sample surveys 1975–1990: An age of reconciliation? *International Statistical Review, 62,* 5–19.

Stephan, F. (1948). History of the uses of modern sampling procedures. *Journal of the American Statistical Association, 43,* 12–39.

Sterba, S. K. (2006). Misconduct in the analysis and reporting of data: Bridging methodological and ethical agendas for change. *Ethics & Behavior, 16,* 305–318.

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

TAF-Y101790-10-0602-C010.indd 290 12/4/10 9:35:57 AM

Sterba, S. K., Egger, H. L., & Angold, A. (2007). Diagnostic specificity and nonspecificity in the dimensions of preschool psychopathology. *Journal of Child Psychology and Psychiatry, 48,* 1005–1013.

Sterba, S. K., Prinstein, M. J., & Nock, M. (2008). Beyond pretending complex nonrandom samples are simple and random. In A. T. Panter & S. K. Sterba (Co-chairs), *Quantitative methodology viewed through an ethical lens*. Boston: Division 5, American Psychological Association.

Stoop, I. A. (2004). Surveying nonrespondents. *Field Methods, 16,* 23–54.

Sugden, R. A., & Smith, T. M. F. (1984). Ignorable and informative designs in survey sampling inference. *Biometrika, 71,* 495–506.

United Nations. (1946). *Economical and social council official records: Report of the statistical commission, first year*. Lake Success, NY.

United Nations. (1947). *Economical and social council official records: Report of the statistical commission, second year*. Lake Success, NY.

United Nations. (1948). *Economical and social council official records: Report of the statistical commission, third year*. Lake Success, NY.

United Nations. (1949a). *Economical and social council official records: Report of the statistical commission, fourth year*. Lake Success, NY.

United Nations. (1949b). United Nations economic and social council sub-commission on statistical sampling: Report to the statistical commission on the second session of the sub-commission on statistical sampling I. *Sankhyā: The Indian Journal of Statistics, 9,* 377–391.

United Nations. (1949c). United Nations economic and social council sub-commission on statistical sampling: Report to the statistical commission on the second session of the sub-commission on statistical sampling II. *Sankhyā: The Indian Journal of Statistics, 9,* 392–398.

Wheeler, L., & Reis, H. T. (1991). Self-recording of everyday life events: Origins, types, and uses. *Journal of Personality, 59,* 339–354.

Wilkinson, L., & the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594–604.

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

TAF-Y101790-10-0602-C010.indd 291 12/4/10 9:35:57 AM

Downloaded By: 10.3.98.93 At: 12:59 23 Oct 2018; For: 9780203840023, chapter10, 10.4324/9780203840023.ch10

TAF-Y101790-10-0602-C010.indd 292 12/4/10 9:35:57 AM