

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Sociology Department, Faculty Publications

Sociology, Department of

2020

Survey Costs: Where Are We and What Is the Way Forward?

Kristen M. Olson

James Wagner

Raeda Anderson

Follow this and additional works at: <https://digitalcommons.unl.edu/sociologyfacpub>



Part of the [Family, Life Course, and Society Commons](#), and the [Social Psychology and Interaction Commons](#)

This Article is brought to you for free and open access by the Sociology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Sociology Department, Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Survey Costs: Where Are We and What Is the Way Forward?

Kristen Olson,¹ James Wagner,² and Raeda Anderson³

1 Leland J. and Dorothy H. Olson Professor of Sociology, Department of Sociology,
University of Nebraska-Lincoln, 703 Oldfather Hall, Lincoln, NE 68588-0324,
USA

2 Research Associate Professor, Institute for Social Research, University of Michi-
gan, 4053 ISR, 426 Thompson St, Ann Arbor, MI 48106-1248, USA

3 Research Scientist, Virginia C. Crawford Research Institute, Shepherd Center,
Atlanta, GA 30309, USA

Correspondence — Kristen Olson, Department of Sociology, University of Nebraska-Lincoln,
703 Oldfather Hall, Lincoln, NE 68588-0324, USA; *email* kolson5@unl.edu

Abstract

Survey costs are a major driver of survey design decisions and thought to be related to survey errors. Despite their importance, no common language exists for discussing survey costs, nor are there established criteria for identifying which cost metrics are useful for which purposes. Past efforts to study survey costs may have been hampered by the notion that more reporting is better reporting. This article starts by introducing a typology for survey cost metrics defined by the type of cost (estimated, observed in records, and actually incurred), currency versus non-currency measures, and level of aggregation (total, by components, per unit, relative). We also suggest a set of criteria – errors in costs, generalizability, and the degree to which a cost measure is informative about survey error sources – for evaluating the utility of cost metrics. We illustrate the evaluative criteria with the cost metrics. We argue that clearly articulating types of survey costs and resetting these baseline evaluative criteria for the utility of different types of costs will help us expand research in this critical area. We conclude with recommendations for future research in costs within and across organizations.

Keywords: Cost models, Paradata, Survey costs.

Published in *Journal of Survey Statistics and Methodology* (2020)

doi: 10.1093/jssam/smaa014

Copyright © 2020 Kristen Olson, James Wagner, and Raeda Anderson. Published by Oxford University Press on behalf of the American Association for Public Opinion Research. Used by permission.

Published 26 May 2020.

1. Introduction

In 2017, the combined budget for the thirteen principal US federal statistical agencies was approximately \$3.25 billion (COPAFS 2017). In 2004, the combined budget for the ten largest US federal statistical agencies was \$1.33 billion, a growth of 184 percent over budgets only 20 years before (Presser and McCulloch 2011). The cost of conducting large-scale government surveys clearly is increasing in the United States, driven partially by decreasing survey response rates (e.g., Brick and Williams 2013). Yet little systematic information is available about survey costs. Furthermore, survey costs are reported inconsistently, on different parts of studies, using different metrics. This paucity of cost information and variability in reporting has substantially limited researchers' understanding of survey costs overall and the potential relationship between survey costs and survey errors. Lack of information about costs also can lead to budget models for surveys that are removed from actual survey costs and, possibly, cost overruns.

Survey costs are important because they constrain the set of design, data collection, processing, and release decisions for any given survey. Therefore, survey costs are a major contributor to the risk of survey error (Groves 1989). The first step in identifying how survey costs relate to survey errors is having a common language for reporting survey costs, comparable to standard definitions that have been adopted for response rates (AAPOR 2016). This common language is important because organizations vary in the types of cost information that are recorded. This variation arises because of differences in systems that organizations have in place for tracking costs and organizations' desire to minimize the reporting of potentially proprietary information but hinders the understanding of costs more generally.

Lack of information about survey costs is well-established. In 2006, the National Institute of Statistical Sciences held a workshop to examine issues related to survey costs and cost modeling that identified a lack of quality cost data as a barrier to understanding survey costs (Karr and Last 2006). A National Academies panel examining redesign of the National Crime Victimization Survey stated that data collection costs are "difficult to discern" (Groves and Cork 2008, p. 67). Furthermore, Vehovar, Berzelak, and Lozar Manfreda (2010), in developing cost models for surveys of various modes, bemoan the difficulties of obtaining high quality, generalizable cost data.

The situation is exacerbated by a growing need for cost data. Changes in the survey environment have led to the elaboration of more complex designs, which, in turn, have led to a greater demand for more detailed cost information. Responsive and adaptive designs require detailed information about costs to optimize complex survey designs (Groves and Heeringa 2006; Schouten, Peytchev, and Wagner 2018). In addition, surveys moving to mixed modes (de Leeuw 2005; Olson, Smyth, Horwitz, Keeter, Lesser, et al. 2019) often use costs to justify these mode changes

In this article, we develop evaluative criteria for the utility of various types of cost reports. These evaluative criteria start from the general notion that data should be evaluated by their “fitness for use,” not “more cost information is better.” We propose a typology of cost reporting and apply the evaluative criteria to examples of each type. In doing so, we examine the strengths and weaknesses of alternative methods of reporting costs and provide examples from published reports of costs. We end with recommendations for internal research in survey organizations on costs and for more general work to move survey costs into an empirical domain of research.

2. Definitions and notation for cost reports

We begin by defining key terms and notation. Costs are traditionally divided into fixed (costs that are considered to be independent of sample size) and variable (costs that vary with sample size) costs (e.g., Kish 1965, p. 264). Each study has a number of design features that affect either fixed or variable costs, which we call “components.” These study components vary by mode and are organized around key design decisions. For example, in a mail survey, key components might include questionnaire development and pretesting, questionnaire layout and formatting, sample design, sample selection, printing, incentives, postage, data entry, and post-survey processing. In theory, the survey components are mutually exclusive and exhaustive and are known. The total cost for the survey is the sum of the cost of each of the components. Costs also may depend on *time*, including day, week, month, or year of a repeated cross-section survey, the wave of a longitudinal survey, or the phase of data collection in a responsive or adaptive design.

We introduce notation to aid our discussion of costs. In particular, let $i = 1, \dots, D$ be the mutually exclusive and exhaustive components of a study where costs may be incurred and $t = 1, \dots, T$ be time of a study. The costs that are observed or recorded through record-keeping systems in component i at time t are *observed costs*, with observed costs measured in currency denoted C_{it}^0 .

We denote the total observed cost in currency over all survey components for time t from cost records as $C_t^0 = \sum_{i=1}^D C_{it}^0$. The proportion of currency observed costs for each component is defined as:

$$\frac{C_{it}^0}{C_t^0} = \frac{C_{it}^0}{\sum_{i=1}^D C_{it}^0}$$

Relative total costs for a study administration at time t to time t_0 (e.g., phase 1 versus phase 2; wave 1 versus wave 2) is calculated as $C_t^0/C_{t_0}^0$; t and t_0 may also indicate alternative survey modes used at different times. Costs often are reported relative to the number of sample units, n , or the number of responding units, r . We define the observed cost of a particular component per sampled unit as C_{it}^0/n and the observed overall cost per sampled unit as C_t^0/n , for example.

Currency (e.g., dollars, euros) is an obvious choice for reporting costs; however, other units may be used for planning or reporting. For example, an organization may indicate the number of interviewers used in terms of “full-time equivalent” (FTE, where FTE calculations may vary across countries, e.g., Girasek, Kovács, Aszalós, Eke, Ragány, et al. 2016) staff. Similarly, recordkeeping systems for recruitment efforts may track the number and types of call attempts and total number of hours an interviewer worked, but not the currency expended on individual tasks (e.g., recruiting, interviewing).

We adapt the notation for costs that are recorded through nonmonetary methods (e.g., FTE, hours, call attempts). In particular, let J_{it}^0 denote observed costs in nonmonetary units for component i at time t . Total observed costs in nonmonetary units is defined as $J_t^0 = \sum_{i=1}^D J_{it}^0$ and relative costs in nonmonetary units at time t and t_0 as $J_t^0/J_{t_0}^0$, for example, and non-dollar observed costs per sampled unit as J_{it}^0/n .

We summarize our notation for these types of cost reporting in **Table 1**. The notation reflects three dimensions: level of aggregation (full study versus components), units used to measure costs (currency or not), and whether costs are reported as totals, cost per unit, or some relative measure.

Table 1. Types of Cost Reporting for Observed Costs

	<i>Full study</i>		<i>Study components</i>	
	<i>Currency</i>	<i>Not currency</i>	<i>Currency</i>	<i>Not currency</i>
Overall	$C_t^o = \sum_{i=1}^D C_{it}^o$	$J_t^o = \sum_{i=1}^D J_{it}^o$	C_{it}^o	J_{it}^o
Cost per unit	C_t^o/n	J_t^o/n	C_{it}^o/n	J_{it}^o/n
	C_t^o/r	J_t^o/r	C_{it}^o/r	J_{it}^o/r
Relative costs	$C_t^o/C_{t'}^o$	$J_t^o/J_{t'}^o$	$C_{it}^o/C_{it'}^o$	$J_{it}^o/J_{it'}^o$
			$C_{it}^o/C_{i't}^o$	$J_{it}^o/J_{i't}^o$
			$C_{it}^o/C_{i't'}^o$	$J_{it}^o/J_{i't'}^o$

3. Evaluative criteria

We now explore evaluative criteria for determining the utility of cost reports. First, we focus on errors in costs. Then, we examine two other criteria – generalizability and relationship to survey error sources.

3.1 Errors (Bias and Variance) in Cost Reports

Errors in costs (and more formally, the *mean squared error [MSE] in cost measures*) are an important *evaluative criterion* for the utility of cost reporting (Burger, Perryck, and Schouten 2017). To discuss these errors in cost measures, we introduce two additional types of costs and associated notation: (i) costs that are *estimated* before the survey occurs for component *i* at time *t*, denoted as C_{it}^E for currency-metric-estimated costs and J_{it}^E for non-currency-metric-estimated costs, and (ii) *actual* costs incurred through funds spent and/ or personnel time used for component *i* at time *t*, with currency-metric actual costs denoted as C_{Ait} and non-currency-metric actual costs denoted as J_{it}^A . Actual costs may not always be fully observed (**Table 2**).

Table 2. Cost Notation for Observed, Actual, and Estimated Costs per Component (i) and Time Period (t)

<i>Type of costs</i>	<i>Definition</i>	<i>Currency</i>	<i>Not currency</i>
Estimated (budgeted)	Costs that are estimated or budgeted prior to conducting the survey	C_{it}^E	J_{it}^E
Observed (recorded)	Costs that are available from record systems	C_{it}^O	J_{it}^O
Actual	Costs that are incurred while designing, conducting, and processing the survey	C_{it}^A	J_{it}^A

An analogy for these three types of costs comes from errors in records for survey variables. Records are typically assumed to be error-free “observed” true values for survey variables. Any discrepancy between the reported value and the records arises from reporting errors (e.g., Mathiowetz 1992; Davies and Fisher 2009). Yet administrative records contain errors, including errors in administrative earnings data (Abowd and Stinson 2007), poor quality voting records (Cassel 2004), and errors about employment in population register data (Pavlopoulos and Vermunt 2015). In each instance, the magnitude of the difference in survey reports and “truth” varies when comparing the reported values to the observed records versus the actual or “true” value.

We define three measures of bias in costs as the difference between any two of the types of costs. In statistics, bias is a property of an estimator that manifests over conceptual replications of an estimate and is nonzero when the expected value of an estimator differs from the true value (e.g., Kish 1965, p. 60). For instance, measurement bias manifests when reported values for a survey outcome differ systematically from the true value, in expectation, over replications (e.g., Cochran 1977, pp. 377–8). Unlike for survey outcome variables where there is an easily identified population value of interest to be estimated, for costs, there are different potential “true values” of interest. Therefore, there are different types of bias in the measurement of costs that are relevant. In practice, separating bias and variance components of costs may be difficult or impossible for a one-time survey (Groves 1989, Chapter 2) – an observed difference between two cost metrics could be due to bias (a constant difference in costs that would be observed over multiple replications) or variance (variation in the difference in costs that would be observed over multiple replications, but an average difference of zero), or both biasing and variable errors. We put aside this problem to focus on the conceptual differences between the three types of cost estimates in table 2, using the somewhat abused, but conceptually helpful, shorthand-term “bias” to represent these differences, as is often done in record-check studies for measurement error (Groves 1989, Chapter 7).

First, for any given component i at time t , *estimated costs deviate from observed costs* by $C_{it}^E - C_{it}^O$ (if estimated costs accurately predict observed costs, this bias is 0). This type of “prediction error” or “cost model bias” is important and relevant for survey organizations and is routinely monitored for certain survey components. “Cost model bias”

occurs, for example, when more effort is required to obtain a target number of interviews than anticipated. This is sometimes called an “over-run.” Clearly, cost model bias is consequential.

Aside from cost overruns in a completed study, systematic errors in judgment could lead to biases in cost estimates for future studies. For example, using an “anchor and adjust” strategy of estimation (Tversky and Kahneman 1992), an individual making a judgment starts with an estimate that is anchored on some existing value. Errors occur when the individual insufficiently adjusts the estimate, failing to account for known differences between the situation under consideration and the context of the anchoring value. For survey costs, when an existing survey is used as the baseline for cost prediction, budget models may be in error, especially if information on cost model bias on previous studies is not communicated back to the budgeting team.

Second, there may be “measurement error” in cost accounting systems. A “measurement bias” for costs occurs when *observed costs deviate from actual costs* by $C_{it}^A - C_{it}^O$. If all costs are assigned to the correct survey component in records, this bias is zero. Cost measurement bias may occur, for example, if the accounting system does not allow the organization to disentangle different types of tasks (e.g., hours spent on completing screening interviews versus noncontact attempts; Wagner 2019). When component-level costs are used to make decisions about design features, this type of bias can lead to inefficient designs. At a survey level, these errors might lead a survey organization to make incorrect conclusions about “profit” or “loss” on the balance sheet of a survey. We acknowledge that actual costs are difficult to measure outside of special studies. Nevertheless, defining an actual cost is helpful conceptually, even it is rarely observed.

The third type of bias occurs when estimated and actual costs differ. For any given survey component, *estimated costs deviate from actual costs* by $C_{it}^E - C_{it}^A$. If estimated costs accurately reflect actual costs, this bias is zero. This type of bias occurs, for example, if survey statisticians are *budgeted* for an inadequate number of hours to develop sample design compared to *actual* number of hours it takes to do so. If the same statisticians only report the total number of budgeted hours, rather than monitoring and reporting the hours actually spent developing the sample design (perhaps to avoid cost overruns), then it is unlikely that this bias will be detected without a special cost tracking study. Given that the

actual costs are rarely available, this bias is unlikely to be monitored. One consequence is that if this bias does exist, but actual and observed costs differ, the survey organization would be unaware of it.

In addition to bias, variance or uncertainty in costs occurs. First, as noted by Groves (1989), costs for the same survey are stochastic and vary if the survey is repeated multiple times (e.g., multiple cross-sections of an identical survey). We use this time aspect to define *variance in costs over time*, similar but not equivalent to the frequentist concept of repeated conceptual trials of a survey. The variance of observed survey costs over time—using inflation-adjusted currency—can be estimated by:

$$\text{Var}(C^o) = \frac{\sum_{i=1}^T (C^o_t - \bar{C}^o)^2}{(T - 1)}$$

For a one-time survey, this quantity is undefined, but grouping similar surveys may provide a reasonable estimate.

Variance in costs also can be observed over clustering units in a survey, such as over primary sampling units (PSUs), communities in a multiple-community survey, interviewers, or supervisors. As shown in **Figure 1**, behavior coders identified a set of interviewer and respondent behaviors in two telephone surveys, and each coder varied in time per interview. Average time to code an interview was 1.53 hours (SD = 0.42) in

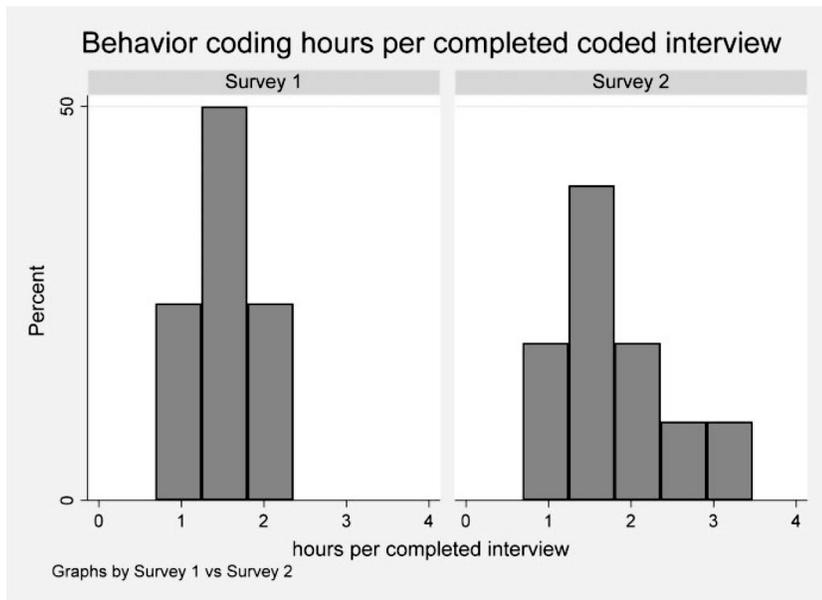


Figure 1. Distribution of Hours for Behavior Coding Two Telephone Interviews.

survey 1 and was 1.75 hours (SD = 0.69) in survey 2, shorter on average than the estimated time (estimated time for survey 1 = 2 hours; survey 2 = 2.17 hours). This kind of variation among clustering units, that is, “random effects” for interviewers, coders, or supervisors, may contribute to cost variation.

The next type of cost variance is related to *uncertainty about how estimated costs relate to observed costs*. When developing budgets for survey clients, some organizations make probability statements about how likely estimated costs are to be observed (e.g., “there is a 10 percent chance of the observed costs exceeding the budget under this design”, or $p(C_{it}^O > C_{it}^E)$), drawing on expert insights (and perhaps data) of “cost model bias” distributions from previous studies (Burger et al. 2017). Estimates could be derived through Monte Carlo simulations with draws from distributions specified by experts for key parameters. This could be called a Bayesian credible interval for costs, drawing implicitly on knowledge of error in previous cost estimates. A credible interval for costs is available for a one-time survey but relies upon expert judgment and a specified model. Uncertainty in budgeted costs relative to observed costs across multiple surveys was not reported in our search of existing literature and is likely viewed as proprietary but could improve cost projections.

3.2 Generalizability

Our *second evaluative criterion* has to do with utility of cost information across multiple survey contexts. That is, cost reports are useful to the extent they are *generalizable*. By generalizable, we mean that costs can be transported from one survey to another similar survey, including between surveys implemented by different organizations (Groves [1989] calls this a “domain of applicability”).

Understanding what is included and excluded from cost reports is important. When cost components are well defined, meaningful comparisons across studies are more likely. Comparisons are difficult to make when components are not independent of each other; an increase in the funds allocated to one component (e.g., incentive) might lead to a reduction in costs in another component (e.g., follow-up efforts). Thus, when reporting costs, it is important to identify the major influences on cost drivers for particular survey components.

Generalizability in costs also can be difficult to evaluate to the extent that costs are nonlinear and discontinuous (Groves 1989). For instance, Wagner and Olson (2018) found that the number of neighborhoods that interviewers visited (a driver of face-to-face survey costs), on average, differed by the phase of data collection and week of field period. Thus, when reporting costs related to interviewer travel, important influences on generalizability of these measures include field period length, whether the survey had a responsive design phase component, and the phase of data collection for travel.

One relatively unacknowledged limitation of cost reporting has been differences in “fixed” costs between types of organizations and related differences in indirect rates. Larger versus smaller survey organizations may experience different “fixed” costs for any given survey. For example, Kaplowitz, Hadlock, and Levine (2004) report “off-campus/non-faculty” rates for programming and hosting costs for web surveys as these would be more comparable to costs incurred for smaller organizations. Generalizability in cost structures may also vary when different survey sponsors vary in required numbers of meetings, types of documentation, and other deliverables that affect or constrain costs. Explicit statements about survey organization infrastructure and survey sponsor requirements for data collection and non-data collection components will help understanding the generalizability of costs and cost drivers.

3.3 Linkage to Survey Error Sources

Our final evaluative criterion for the utility of cost information is whether it is *informative about an error source* of a survey statistic (e.g., y). Despite longstanding interest in survey costs (e.g., Sudman 1967), the relationship between survey costs and survey errors is not well understood (e.g., Singer, Hoewyk, Gebler, Raghunathan, and McGonagle 1999; Karr and Last 2006). This lack of understanding is likely due to the proprietary nature of survey cost information, potential errors that arise in measuring costs, and difficulty in quantifying survey errors. Notably, there is no universally agreed upon metric of “survey quality” that is easily measured across all survey error sources. We provide examples of cost models linked to survey errors as a launching point.

In the survey sampling literature, explicit cost models that link to a corresponding *sampling error* model are common. For instance,

Kalsbeek, Mendoza, and Budescu (1983) develop a cost model to simulate alternative sample designs for a face-to-face cluster sample that incorporates cost per cluster, cost per element within the cluster, and an elaborate model for costs of follow-up with elements in each cluster that accounts for the distance traveled (e.g., to and within the cluster, and the interviewer's workload in the cluster).

Biemer and Stokes (1985) develop a cost model to optimize the design for *interviewer variance* experiments in face-to-face surveys as a function of the number of interviewers, their workloads, and the cost of within-neighborhood travel, linked to errors in survey variables measured as interviewer variance components. Nonresponse error outcomes also have received cost modeling attention. Vehovar et al. (2010) use cost data (euros per $n = 1,000$ units) and *nonresponse error on estimated mean age* (comparing survey reports to official statistics) from two organizations conducting surveys in different modes.

The relationship between survey costs and survey errors features prominently in discussions of *responsive and adaptive designs*. For instance, Calinescu, Bhulai, and Schouten (2013) use costs per interview, noncontacts, and refusal call attempts in different modes at a given time during the data collection process to constrain decisions of which modes to use with the goal of maximizing a proxy indicator for nonresponse bias, the R-Indicator. Other cost studies have used indicators of sampling, nonresponse error, response rates, and measurement error across different modes as a constraint on costs (Calinescu and Schouten 2015; Burger et al. 2017; Kaminska and Lynn 2017).

Ideally, the cost model should be parameterized in terms of each of the key design decisions to be made, and how each of these design decisions affect each survey error source (sampling, coverage, nonresponse, measurement, processing). This model would be fully informative about how costs relate to design decisions and the resultant error structure of estimates from the survey variables themselves. In practice, these design decisions are complex and difficult to articulate exactly in a cost model (e.g., comparing costs of alternative modes). However, like the Total Survey Error perspective itself, a fully articulated cost model linked to each survey error source remains a useful conceptual goal (e.g., Groves 1989, Chapter 2).

4. Examples of cost reporting

In this section, we use the three evaluative criteria from section 2—mean square error of costs, generalizability, and being informative about survey error sources—to assess the utility of different types (total costs, cost per unit, and relative costs) of reporting. Within each of these three types, we will look separately at currency and non-currency reports.

4.1 Total Costs: Overall and for Survey Components

Currency-related total costs provide insights into the set of possible major design decisions (such as mode) for a given survey cost (e.g., Judkins, Waksberg, and Northrup 1990; Morrison, Wahlgren, Hovell, Zakarian, Burkham-Kreitner, et al. 1997; Cobanoglu, Warde, and Moreo 2001; Cotter, Burke, Stouthamer-Loeber, and Loeber 2005; Hardigan, Succar, and Fleisher 2012; Olson et al. 2019). Non-currency measures for total costs may help avoid the disclosure of proprietary information.

Total costs in a relevant currency generally are easily obtained for both estimated and observed costs from record systems. They may vary over survey administrations, but *uncertainty* about the total budgeted or charged costs themselves is low. Yet, in our view, total costs in currency are among the *least generalizable* metrics between studies. Although some variation in cost can be attributed to obvious factors such as mode, sample size, and contact attempts, transporting “total cost of a survey” to other surveys is difficult because the set of individual components that are included in reports of total costs (e.g., noninterviewing staff labor, field supervision, and programming) varies dramatically. Total costs are not necessarily indicative of survey error sources, as surveys with lower total cost may perform better on quality metrics. For example, Link et al. (2008) report that a lower data collection cost mail version of the Behavioral Risk Factor Surveillance System had better coverage properties and higher response rates than the more expensive telephone version of the same survey.

Total costs are not always reported, and many reports only detail costs for some components. For example, studies of the costs and error implications of using incentives might ignore the cost of questionnaire development. Some component costs may be considered proprietary (e.g., costs for senior personnel; overhead costs) or impossible to

capture accurately (e.g., interviewer travel costs between primary sampling units (PSUs)). Identifying what is included and excluded in each reported component is an important, if overlooked, aspect of reporting costs.

Costs for individual survey components vary, including costs for direct field efforts, computer programming, coding and data entry, printing, postage, statistician and survey methodologist time, and data processing, among others (e.g., Sudman 1967; Kristal, White, Davis, Corycell, Raghunathan, et al. 1993; Morrison et al. 1997; Cobanoglu et al. 2001; Shannon and Bradshaw 2002; Fries, James, Hammer, Shugarman, and Morris 2004; Cotter et al. 2005; Hardigan et al. 2012). Some costs for components may be well recorded and thus have *low uncertainty*; others may not be recorded well at all and thus have *high uncertainty*. As with total costs, specific design features that contribute to component-specific costs, such as “direct field costs,” vary across studies. Unlike total costs, this information *may be more informative about survey error sources and may be more generalizable* across organizations fielding similar studies.

Component-specific non-currency costs can be reported in varying metrics. One common metric is the number of hours spent to complete a task. As each organization has its own charge rates, non-currency component-specific totals (e.g., hours spent on telephone calling) may be more generalizable across studies. Generalizability of component-specific costs and whether these costs are informative about survey error sources is likely to vary depending on the level of detail. Larger components (e.g., data collection costs in general) will be less generalizable to other studies because of the multitude of decisions that are made and included as part of data collection. Smaller categories in interviewer-administered surveys (e.g., total costs for field supervision) may be more likely to be generalizable if information is known about the inputs into that cost (e.g., number of supervisors, number of interviewers per supervisor, extent of supervision task). Component-specific cost information may also be *tied to directional hypotheses relating the task to survey error sources* (e.g., more supervision reduces interviewer-related errors). Smaller categories in self-administered surveys (e.g., printing for a mail survey) are more likely to be generalizable with the important proviso that the inputs for costs are known (e.g., number of surveys, number of mailings sent).

For example, in a study of costs in the National Health Interview Survey, Judkins et al. (1990) analyzed payroll data from 1988 and 1989 to obtain total and component-level cost estimates for individual types of tasks related to listing and interviewing. They estimated labor costs for listing (\$89,000), interviewing (\$1,449,000), data entry (\$22,000), training (\$139,000), re-interview and observation (\$62,000), and other interviewer and field supervision (\$38,000). They also estimated costs for mileage, per diem, telephone expenses, and other expenses, such that the total costs for listing were \$134,000 and \$2,179,000 for interviewing.

As noted by Judkins et al. (1990), *errors in costs recorded for components* exist, leading to discrepancies between observed costs and actual costs. Cost categories that are not easily tracked or for which tasks are not readily disambiguated in the record system are less accurate. Costs by survey component often rely on individuals accurately reporting the distribution of their work across each of those component-related categories. As with all measurement tasks that rely on people, this is likely to lead to errors in observed component-specific costs.

4.2 Cost per Unit

A common method of reporting costs is to report costs per interview or per sampled unit (e.g., Kalsbeek et al. 1983; Siemiatycki and Campbell 1984; Walker and Restuccia 1984; Kristal et al. 1993; Morrison et al. 1997; Cobanoglu et al. 2001; Teitler, Reichman, and Sprachman 2003; Kaplowitz et al. 2004; Schreckhise 2005–2006; Teisl, Roe, and Vayda 2005; Hardigan et al. 2012). Studies vary in reporting costs per sample unit, per respondent, or by some other metric (e.g., contact attempt).

For example, Cotter et al. (2005) report cost per unit in a longitudinal study using two metrics. First, they identified how much was spent for the project in total (not just data collection) (\$1,260,238) and divided the total budget by the total number of completed interviews (966) to obtain a cost per interviewed case of \$1,304. They were also interested in the costs for locating hard-to-reach respondents longitudinally. They divided the total costs for the person responsible for the locating attempts (\$214,421) by total number of contact attempts (10,756) to yield a per-contact-attempt cost of \$19.93.

Non-currency measures can be used to report costs per complete. For instance, in a review of the literature on tracking and locating sampled

units, McKenzie, Tulsy, Long, Chesney, and Moss (1999) report that one FTE will be needed for every 85–134 cases to be tracked (p. 417), reflecting variability in costs in a way that may be helpful in planning for future surveys. Kirgis and Lepkowski (2013) divide the total number of interviewer hours worked on all parts of the 2006–2010 National Survey of Family Growth (e.g., screening, interviewing, administrative time) by the total number of completed interviews, yielding 9.0 hours per completed interview.

Costs per unit are subject to the *same reporting and misalignment errors* as overall costs and costs in individual categories. Furthermore, estimated costs per unit may make certain assumptions about the total number of completed surveys that do not actually manifest (e.g., survey achieves a lower response rate than assumed during budgeting). Thus, estimated cost per complete may differ from actual/observed costs per complete.

Despite uncertainty, costs per completed unit or costs per sampled unit *is one of the most generalizable methods* of reporting costs across other similar studies. Assuming similar essential survey conditions, other studies can use costs per unit to develop cost estimates for components of a new study. To the extent that costs per complete are recorded in informative categories, they may be *informative about survey errors*.

4.3 Relative Costs

Relative costs are a commonly used metric that do not reveal proprietary cost information. Relative costs are sometimes reported as the proportion of total costs for individual components, relative costs of alternative experimental treatments, or costs of one data collection wave versus another data collection wave. **Table 3** contains a variety of methods of reporting relative costs from published cost reports. For example, Biemer, Murphy, Zimmer, Berry, Deng, et al. (2018) benchmark costs for eight experimental treatments to overall average cost, showing the percentage change in costs for different combinations of modes and incentives.

If the errors observed in cost measures in the numerator and denominator of the relative costs are equal or proportionate, then *relative cost measures may have smaller errors* than other cost measures. It is unknown, however, whether this assumption holds. It may also be that

Table 3. Example Articles using Relative Cost Measurements

<i>Article</i>	<i>Summary</i>
Andresen, Machuga, Van Booven, Egel, Chibnall, et al. (2008)	19.8% of cost/interview spent on tracing
Baker, Bradburn, and Johnson (1995)	Computer-Assisted Personal Interviewing (CAPI) training costs 70% higher than Paper-and-Pencil Interviewing (PAPI); CAPI direct field interviewing costs same as PAPI; PAPI data processing costs 2.5 times higher than CAPI; CAPI total costs per case 6% higher than PAPI
Cobanoglu et al. (2001)	After accounting for web survey software access fee, cost cheaper for web surveys than mail and fax when $n > 200$
de Leeuw and Nicholls (1996)	Propose rule of thumb that 1,000 interviews are needed for cost reduction in Computer-Assisted Telephone Interviewing relative to CAPI
Fries et al. (2004)	Telephone screening rather than in person assessment saves an estimated 11% of cost
Hagget and Mitchell (1994)	Prenotification letters increased costs per return of mail surveys by 13–59% in three studies; three other studies showed decreased costs with prenotification by 11–83%
Kaminska and Lynn (2017)	Costs for components in a mixed mode survey specified relative to the cost per sample unit in a web survey
Karr and Last (2006)	Interviewer efforts make up 70% of variable costs
Kirgis and Lepkowski (2013)	National Survey of Family Growth interviewers spent 55% of their time on screening and interviewing in 2006–2010, an increase over the 40% of their time on these tasks in 2002–2003
Kristal et al. (1993)	Costs increased by 64% for cases that were interviewed after 12–22 call attempts versus 1–11 call attempts and increased by 50% for refusal conversion
Sebestik, Zelon, DeWitt, O'Reilly, and McGowan (1988)	CAPI training costs 18% higher than PAPI; CAPI direct field interviewing 1% higher than PAPI; CAPI transmitting time costs 3% more than PAPI; CAPI total costs per case 17% higher than PAPI
Sudman (1967)	Percent of total interviewer time spent on interviewing ranges from 21% to 41% across seven studies
Teisl et al. (2005)	\$5 phone card in two mailings is 61% more expensive than \$1 cash in two mailings
Teitler et al. (2003)	In hospital interviews = 65% of total costs; telephone interviews = 20% of total costs; field cases = 15% of total costs

Table 4. Rating of the Utility of Types of Cost Measures across Three Evaluative Criteria

<i>Cost measure</i>	<i>Evaluative criteria</i>		
	<i>Errors in (observed) costs</i>	<i>Generalizability</i>	<i>Informative about survey error sources</i>
Total costs	Low	Low	Low
Total costs for components	Moderate to high	Moderate	Low to moderate
Costs per unit	Low to moderate to high	Moderate to high	Moderate to high
Relative costs	Low to moderate	Moderate	Moderate

Higher-quality cost measures will be low for errors in observed costs and moderate to high in generalizability and degree to which it is informative about survey error sources.

relative costs, as ratio estimates, *are more variable* than total or average costs per case, given that there are uncertain values in the numerator and denominator. It may be that relative costs are *generalizable* to other organizations for the same set of survey tasks. Relative costs may be *informative about survey error structures*, but in the absence of known survey errors are more likely to be informative about field efficiency of design alternatives. For instance, Schouten et al. (2018) report that an adaptive survey design reduced the average number of call attempts per case from 8.2 to 7.4 attempts for landline telephone numbers, thereby reducing interviewer calling time (figure 7.2, p. 120).

Table 4 summarizes these various measures and their utility relative to the suggested evaluative criteria. In this table, a more useful cost measure is rated as having “low” errors in costs, but as having “moderate” to “high” generalizability to other surveys and “moderate” to “high” levels of being informative about survey error sources. For instance, we expect that the errors observed in total costs will be low, but that total costs is not generalizable (rated low) or informative about survey error sources (also rated low).

5. Recommendations

Survey researchers reading this typology of costs and evaluative criteria may feel overwhelmed. We agree. In this section, we identify two sets of recommendations aimed at moving research on survey costs forward. The first set of recommendations is for internal research at a survey

organization to evaluate their own costs. The second set of recommendations is for reporting on costs more generally and publishing research results.

5.1 Recommendations for Internal Research at Survey Organizations

One of the biggest barriers to understanding survey costs is the difficulty of measuring costs. Thus, the first recommendation to survey organizations is to identify how costs are currently measured and recorded. In particular, *we urge survey organizations to identify and evaluate a range of currency or noncurrency metrics that can be measured in the current record-keeping system of costs.* Some costs may be easy to measure, such as the number of hours that interviewers charge to payroll, the amount of money spent on postage and printing advance letters, or cost of access to web survey software. These smaller (measurable) survey components provide a starting point for an organization in reporting costs. We recognize that developing systems for obtaining this cost information may pose an infrastructure expense and expect that the benefits of a more facile cost system will outweigh the costs of its development.

Then, as with all measurement (and statistical) problems, survey organizations should identify potential sources of error in measuring these costs, including when the measured costs associated with a survey component are incomplete. For example, at some organizations, basic infrastructure for surveys (e.g., telephones, sample management systems) may be already in place and thus the costs have already been incurred. Sources of error in cost measures can vary and are related to human error. For example, interviewers working on multiple projects may not accurately report time on individual projects. Other constraints on costs, including how frequently cost information is updated, should be evaluated. It may be helpful for survey organizations to examine the experience of other disciplines (e.g., manufacturing, operations research) in measuring and estimating costs, especially as these fields have insights into variable costs, draw on “paradata” about the process itself, and recognize that employee timesheets contain errors (Niazi, Dai, Balabani, and Seneviratne 2006; Mislick and Nussbaum 2015; Wouters and Stecher 2017).

Survey shops are likely to vary in the types of cost information that can be tracked and reported. Some organizations may have extensive budgeting, tracking, and accounts payable teams in place; other survey shops may be using off-the-shelf accounting software, permitting little flexibility in cost reports. This is especially important for understanding potential differences in cost structures for large versus small survey organizations.

Second, *we recommend that survey organizations identify the types of cost information that can be identified in their record-keeping systems, and compare it to the types of inputs found in their survey budgets.* Survey organizations may already do this kind of evaluation (budget bias: $C_i^E - C_i^O$) for certain survey components. If survey budget inputs cannot be matched to the costs kept in the record-keeping system, then it will be impossible for organizations to evaluate the places at which budgets go wrong. Anecdotally, the set of staff who work on budgeting are often different from the survey implementation team, who may be different from the survey management team watching costs at a high level (see also Wouters and Stecher [2017] for a similar situation that led to errors in assigned tooling costs to particular products). In addition, as discrepancies between estimated and observed costs are measured over multiple similar studies, organizations can use this information to develop probability estimates of how likely the study design can be achieved as budgeted, and measures of variability of costs over time.

Third, *we recommend that survey researchers and organizations identify how available cost information may link to available survey error indicators.* These survey error indicators are likely to come from paradata, may provide insights into coverage, measurement, nonresponse, and processing errors (e.g., Kreuter 2013), and should reflect outcomes that the survey organization is interested in changing or monitoring, especially if conducting a responsive or adaptive design (Groves and Heeringa 2006; Schouten et al. 2018). These cost-error models are unlikely to be as cleanly specified as cost models for sample size optimization (e.g., Biemer and Stokes 1985; Wagner, West, Guyer, Burton, Kelley et al. 2017). In addition, if the relationship between different cost metrics and different survey error indicators is examined over multiple surveys (or other units such as interviewers, supervisors, sampled strata), then information about the variability of costs will be obtainable.

5.2 Recommendations for Future Research on Costs

We now turn from recommendations for internal research within survey organizations to recommendations for more general future research on costs. This kind of research is needed before one can make more general recommendations about which cost measures and components are useful for understanding different types of survey error sources or for monitoring during data collection.

First, *we recommend that organizations share information publicly about what is recorded or not recorded on survey costs, and potential causes of errors in their record-keeping systems.* Note that this is not a recommendation to share proprietary cost information, but simply a catalog of what metrics are available and/or being used to monitor costs across organizations. This will allow the survey research field to evaluate what types of costs are not well-captured but viewed as critically important. For instance, many sampling models identify interviewer travel within PSUs as an important input into cost models constraining sampling error (Kalsbeek et al. 1983; Biemer and Stokes 1985). Yet, it is very difficult to separate within-PSU travel from between-PSU travel in record systems where interviewers simply report mileage traveled. Simply sharing what type of cost information is captured at different types of organizations is a useful place to start. Survey sponsors may also play a role by requiring the publication of cost information as a “deliverable.”

Second, *when survey organizations conduct internal research linking survey cost parameters to survey error parameters, we recommend sharing this research publicly through publications and conference presentations.* One advantage to linked cost-error research is that it is unlikely to directly reveal proprietary cost information, especially when the primary information reported is regression coefficients or other measures of association. Other organizations could then use this information for making inference about how costs may relate to survey errors in other studies. One may assume that more expensive studies yield smaller survey errors. Granquist and Kovar (1997) provide examples of how limiting editing procedures yields substantial cost savings in staff time (the cost metric) devoted to post-survey processing with little effects on final published estimates (the survey error metric).

Third, *we recommend more research on existing cost models assumed for existing survey error models.* An important first step in understanding

the utility of existing cost models would be to identify costs that are commonly specified, for example, in sampling optimization problems (e.g., cost per cluster, per element in a cluster) and assumptions that are made about these costs (e.g., fixed, knowable). A next step would be querying survey organizations for difficulty in obtaining records on cost parameters, and whether recorded information matches assumptions of inputs of the cost model. We suspect that in many organizations, survey statisticians developing sample designs do not fully evaluate whether estimated costs used in sampling error models are observed in practice, or that they even see observed cost data from the sample they are designing.

Finally, a great benefit to the field would be to *systematically evaluate biases and variance in estimated and observed costs for some survey components*. We realize that this recommendation is likely the most difficult. To the extent, however, that survey organizations absorb cost overruns to fixed price contracts with internal funds, then information about deviations from budgeted or observed costs can help survey funders build better budgets. Insights about cost uncertainty can occur without revealing proprietary information. For instance, a repeated cross-sectional survey may accumulate a database of costs in a variety of metrics for each administration of the survey (surveys using a continuous measurement approach also may have this information). Variation in costs over time could be analyzed using a time-series model with predictors of known factors about the season of data collection, aggregate information about survey climate, and other similar information, as well as information about the study design and implementation at that time period (similar to the analyses of aggregate refusal rates by Harris-Kojetin and Tucker 1999). To minimize the disclosure of proprietary information, costs for each administration could be benchmarked to an arbitrary starting point (e.g., first time period observed).

6. Discussion

In this article, we introduced evaluative criteria for the utility of cost information and applied those criteria to a typology for cost reports. We also recommend additional internal research at survey organizations and research on costs more generally. We end with a few observations.

Just as the lack of a consistent definition of response rates hindered an understanding of contributors to survey nonresponse in general, research on survey costs is also limited by variation in what is reported on costs. From our review, the vast majority of the studies that report costs across metrics are small academic and epidemiological studies comparing modes or incentives, not reports from larger private or governmental survey organizations about observed survey costs. Although these smaller studies provide useful insights into some costs of data collection, they may provide a distorted view of the costs of survey research.

We cannot evaluate how survey costs vary by the population of the study from our review of the existing literature. For instance, although we suspect that per-unit costs are higher for a “hard-to-reach” population survey than a general population survey, we have no way to evaluate that assumption from data here. More information on costs for surveys of different populations from the same organization and/or using identical metrics is needed.

Although Groves (1989) provided evidence that survey costs are stochastic, we know little about what empirically predicts variability in survey costs, nor how to incorporate uncertainty about costs into survey design information. This is especially important when certain major cost drivers, like response rates, cannot be known prior to fielding a survey. An empirical evaluation of a common cost metric across multiple surveys conducted by the same organization, with study design factors coded as independent variables and the common cost metric as the dependent variable, will provide useful information into how surveys vary in costs and why.

An understanding of survey costs is critical to making optimal decisions about cost-error trade-offs. An appeal of “big data” is that it can be obtained more cheaply than survey data. Therefore, it is imperative that survey research develops a common language for clearly communicating survey cost information. To remain a viable data collection method, survey organizations must become more transparent about why surveys can cost so much and what is being “bought” with that money. It is well worth the hard work to do this.

Funding This work was partially supported by the National Science Foundation [SES-1132015]. Any opinions, findings and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- AAPOR (2016), *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (9th ed.), Washington, DC: The American Association for Public Opinion Research.
- Abowd, J. M., and M. H. Stinson (2007), "Estimating Measurement Error in SIPP Annual Job Earnings: A Comparison of Census Survey and SSA Administrative Data," paper presented at the Federal Committee on Statistical Methodology Research Conference, Washington, DC.
- Andresen, E. M., C. R. Machuga, M. E. Van Booven, J. Egel, J. T. Chibnall, and R. C. Tait (2008), "Effects and Costs of Tracing Strategies on Nonresponse Bias in a Survey of Workers with Low-Back Injury," *Public Opinion Quarterly*, 72, 40–54.
- Baker, R. P., N. M. Bradburn, and R. A. Johnson (1995), "Computer-Assisted Personal Interviewing: An Experimental Evaluation of Data Quality and Cost," *Journal of Official Statistics*, 11, 413–431.
- Biemer, P. P., J. Murphy, S. Zimmer, C. Berry, G. Deng, and K. Lewis (2018), "Using Bonus Monetary Incentives to Encourage Web Response in Mixed-Mode Household Surveys," *Journal of Survey Statistics and Methodology*, 6, 240–261.
- Biemer, P. P., and S. L. Stokes (1985), "Optimal Design of Interviewer Variance Experiments in Complex Surveys," *Journal of the American Statistical Association*, 80, 158–166.
- Brick, J. M., and D. Williams (2013), "Explaining Rising Nonresponse Rates in Cross-Sectional Surveys," *The Annals of the American Academy of Political and Social Science*, 645, 36–59.
- Burger, J., K. Perryck, and B. Schouten (2017), "Robustness of Adaptive Survey Designs to Inaccuracy of Design Parameters," *Journal of Official Statistics*, 33, 687–708.
- Calinescu, M., S. Bhulai, and B. Schouten (2013), "Optimal Resource Allocation in Survey Designs," *European Journal of Operational Research*, 226, 115–121.
- Calinescu, M., and B. Schouten (2015), "Adaptive Survey Designs to Minimize Survey Mode Effects – A Case Study on the Dutch Labor Force Survey," *Survey Methodology*, 41, 403–425.
- Cassel, C. A. (2004), "Voting Records and Validated Voting Studies," *Public Opinion Quarterly*, 68, 102–108.
- Cobanoglu, C., B. Warde, and P. J. Moreo (2001), "A Comparison of Mail, Fax and Web-Based Survey Methods," *Proceedings of the American Statistical Association, Survey Research Methods Section*, 43, 1–410.
- Cochran, W. G. (1977), *Sampling Techniques* (3rd ed.), New York: John Wiley and Sons.
- COPAFS (2017), "2016-2017 Budget Authority and Proposals for FY 2018 Budget for Principal Statistical Agencies (in Millions of Dollars)," available at <http://www.copafs.org/UserFiles/file/2016-2017Proposed2018BudgetAuthorityforStatAgencies.pdf> Last accessed February 18, 2020.

- Cotter, R. B., J. D. Burke, M. Stouthamer-Loeber, and R. Loeber (2005), "Contacting Participants for Follow-Up: How Much Effort Is Required to Retain Participants in Longitudinal Studies?," *Evaluation and Program Planning*, 28, 15–21.
- Davies, P. S., and T. Fisher (2009), "Measurement Issues Associated with Using Survey Data Matched with Administrative Data from the Social Security Administration," *Social Security Bulletin*, 69, 1–12.
- de Leeuw, E. (2005), "To Mix or Not to Mix Data Collection Modes in Surveys," *Journal of Official Statistics*, 21, 233–255.
- de Leeuw, E., and W. L. Nicholls II, (1996), "Technological Innovations in Data Collection: Acceptance, Data Quality and Costs," *Sociological Research Online*, 1, 23–37.
- Fries, B. E., M. James, S. S. Hammer, L. R. Shugarman, and J. N. Morris (2004), "Is Telephone Screening Feasible? Accuracy and Cost-Effectiveness of Identifying People Medically Eligible for Home- and Community-Based Services," *The Gerontologist*, 44, 680–688.
- Girasek, E., E. Kovács, Z. Aszalós, E. Eke, K. Ragány, R. Kovács, Z. Cserháti, and M. Szócska (2016), "Headcount and FTE Data in the European Health Workforce Monitoring and Planning Process," *Human Resources for Health*, 14, 42.
- Granquist, L., and J. K. Kovar (1997), "Editing of Survey Data: How Much Is Enough?," in *Survey Measurement and Process Quality*, eds. L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin, pp. 415–435, New York: John Wiley and Sons, Inc.
- Groves, R. M. (1989), *Survey Errors and Survey Costs*, New York: John Wiley and Sons, Inc.
- Groves, R. M., and D. L. Cork (eds.) (2008), *Surveying Victims: Options for Conducting the National Crime Victimization Survey*, Washington, DC: The National Academies Press.
- Groves, R. M., and S. G. Heeringa (2006), "Responsive Design for Household Surveys: Tools for Actively Controlling Survey Nonresponse and Costs," *Journal of the Royal Statistical Society*, 169, 439–457.
- Hagget, S., and V.-W. Mitchell (1994), "Effect of Industrial Prenotification on Response Rate, Speed, Quality, Bias, and Cost," *Industrial Marketing Management*, 23, 101–110.
- Hardigan, P. C., C. T. Succar, and J. M. Fleisher (2012), "An Analysis of Response Rate and Economic Costs between Mail and Web-Based Surveys among Practicing Dentists: A Randomized Trial," *Journal of Community Health*, 37, 383–394.
- Harris-Kojetin, B., and C. Tucker (1999), "Exploring the Relation of Economic and Political Conditions with Refusal Rates to a Government Survey," *Journal of Official Statistics*, 15, 167–184.
- Judkins, D., J. Waksberg, and D. Northrup (1990), "Cost Functions for NHIS and Implications for Survey Design," *Proceedings of the American Statistical Association, Survey Research Methods Section*, 34–43.
- Kalsbeek, W. D., O. M. Mendoza, and D. V. Budescu (1983), "Cost Models for Optimum Allocation in Multi-Stage Sampling," *Survey Methodology*, 9, 154–177.

- Kaminska, O., and P. Lynn (2017), "The Implications of Alternative Allocation Criteria in Adaptive Design for Panel Surveys," *Journal of Official Statistics*, 33, 781–799.
- Kaplowitz, M. D., T. D. Hadlock, and R. Levine (2004), "A Comparison of Web and Mail Survey Response Rates," *Public Opinion Quarterly*, 68, 94–101.
- Karr, A. F., and M. Last (2006), *Survey Costs: Workshop Report and White Paper*, Research Triangle Park, NC: National Institute of Statistical Sciences.
- Kirgis, N. G., and J. M. Lepkowski (2013), "Design and Management Strategies for Paradata- Driven Responsive Design: Illustrations from the 2006–2010 National Survey of Family Growth," in *Improving Surveys with Paradata: Analytic Uses of Process Information*, ed. F. Kreuter, pp. 121–144, Hoboken, NJ: John Wiley and Sons.
- Kish, L. (1965), *Survey Sampling*, New York: John Wiley and Sons, Inc.
- Kreuter, F. (2013), *Improving Surveys with Paradata: Analytic Uses of Process Information*, Hoboken, NJ: Wiley.
- Kristal, A. R., E. White, J. R. Davis, G. Corycell, T. Raghunathan, S. Kinne, and T.-K. Lin (1993), "Effects of Enhanced Calling Efforts on Response Rates, Estimates of Health Behavior, and Costs in a Telephone Health Survey Using Random-Digit Dialing," *Public Health Reports*, 108, 372–379.
- Mathiowetz, N. A. (1992), "Errors in Reports of Occupation," *The Public Opinion Quarterly*, 56, 352–355.
- McKenzie, M., J. P. Tulsy, H. L. Long, M. Chesney, and A. Moss (1999), "Tracking and Follow- Up of Marginalized Populations: A Review," *Journal of Health Care for the Poor and Underserved*, 10, 409–429.
- Mislick, G. K., and D. A. Nussbaum (2015), *Cost Estimation Methods and Tools*, Hoboken, NJ: Wiley.
- Morrison, T. C., D. R. Wahlgren, M. F. Hovell, J. Zakarian, S. Burkham-Kreitner, C. R. Hofstetter, D. J. Slymen, K. Keating, S. Russos, and J. A. Jones (1997), "Tracking and Follow-up of 16,915 Adolescents: Minimizing Attrition Bias," *Controlled Clinical Trials*, 18, 383–396.
- Niazi, A., J. S. Dai, S. Balabani, and L. Seneviratne (2006), "Product Cost Estimation: Technique Classification and Methodology Review," *Journal of Manufacturing Science and Engineering*, 128, 563–575.
- Olson, K., J. D. Smyth, R. Horwitz, S. Keeter, V. Lesser, S. Marken, N. A. Mathiowetz, J. McCarthy, E. M. O'Brien, J. Opsomer, D. M. Steiger, D. Sterrett, J. Su, Z. T. Suzer-Gurtekin, C. Turakhia, and J. Wagner (2019), "Transitions from Telephone Surveys to Self-Administered and Mixed-Mode Surveys," *Journal of Survey Statistics and Methodology*, Advance Access. doi:10.1093/jssam/smz062.
- Pavlopoulos, D., and J. K. Vermunt (2015), "Measuring Temporary Employment. Do Survey or Register Data Tell the Truth?," *Survey Methodology*, 41, 197–214.
- Presser, S., and S. McCulloch (2011), "The Growth of Survey Research in the United States: Government-Sponsored Surveys, 1984–2004," *Social Science Research*, 40, 1019–1024.
- Schouten, B., A. Peytchev, and J. Wagner (2018), *Adaptive Survey Design*, Boca Raton, FL: CRC Press.

- Schreckhise, W. D. (2005–2006), “The Costs and Benefits of Mail, Fax, and Mixed-Mode Surveys in Organizational Research,” *Public Administration Quarterly*, 29, 492–510.
- Sebestik, J., H. Zelon, D. DeWitt, J. M. O’Reilly, and K. McGowan (1988), “Initial Experiences with CAPI,” paper presented at the Bureau of the Census Fourth Annual Research Conference, Arlington, VA, March 20–23.
- Shannon, D. M., and C. C. Bradshaw (2002), “A Comparison of Response Rate, Response Time, and Costs of Mail and Electronic Surveys,” *The Journal of Experimental Education*, 70, 179–192.
- Siemiatycki, J., and S. Campbell (1984), “Nonresponse Bias and Early versus All Responders in Mail and Telephone Surveys,” *American Journal of Epidemiology*, 120, 291–301.
- Singer, E., J. V. Hoewyk, N. Gebler, T. Raghunathan, and K. McGonagle (1999), “The Effect of Incentives on Response Rates in Interviewer-Mediated Surveys,” *Journal of Official Statistics*, 15, 217–230.
- Sudman, S. (1967), *Reducing the Cost of Surveys*, Chicago: Aldine.
- Teisl, M. F., B. Roe, and M. Vayda (2005), “Incentive Effects on Response Rates, Data Quality, and Survey Administration Costs,” *International Journal of Public Opinion Research*, 18, 364–373.
- Teitler, J. O., N. E. Reichman, and S. Sprachman (2003), “Costs and Benefits of Improving Response Rates for a Hard-to-Reach Population,” *Public Opinion Quarterly*, 67, 126–138.
- Tversky, A., and D. Kahneman (1992), “Advances in Prospect Theory: Cumulative Representation of Uncertainty,” *Journal of Risk and Uncertainty*, 5, 297–323.
- Vehovar, V., N. Berzelak, and K. Lozar Manfreda (2010), “Mobile Phones in an Environment of Competing Survey Modes: Applying Metric for Evaluation of Costs and Errors,” *Social Science Computer Review*, 28, 303–318.
- (2019), “Estimation of Survey Cost Parameters Using Paradata,” *Survey Practice*, 12, 1–10.
- Wagner, J., and K. Olson (2018), “An Analysis of Interviewer Travel and Field Outcomes in Two Field Surveys,” *Journal of Official Statistics*, 34, 211–237.
- Wagner, J., B. T. West, H. Guyer, P. Burton, J. Kelley, M. P. Couper, and W. D. Mosher (2017), “The Effects of a Mid-Data Collection Change in Financial Incentives on Total Survey Error in the National Survey of Family Growth,” in *Total Survey Error in Practice*, eds. P. P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, N. C. Tucker and B.T. West, pp. 155–177, New York: John Wiley and Sons, Inc.
- Walker, A. H., and J. D. Restuccia (1984), “Obtaining Information on Patient Satisfaction with Hospital Care: Mail versus Telephone,” *Health Services Research*, 19, 291–306.
- Wouters, M., and J. Stecher (2017), “Development of Real-Time Product Cost Measurement: A Case Study in a Medium-Sized Manufacturing Company,” *International Journal of Production Economics*, 183, 235–244.