Survey Methods in Public Administration Research: A Content Analysis of Journal Publications

Geon Lee

Jennifer Benoit-Bryan

Timothy P. Johnson

Department of Public Administration
University of Illinois at Chicago

Correspondence regarding this paper should be directed to:

Timothy P. Johnson, PhD
Survey Research Laboratory
University of Illinois at Chicago
412 S. Peoria St, Suite 615, M/C336
Chicago, IL 60607
Phone: 312-996-5310
Fax: 312-996-3358
E-mail: timj@uic.edu

Prepared for the 10th National Public Management Research Conference, Ohio State University, Columbus, OH, October 1-3, 2009

ABSTRACT

Survey research is a common tool for assessing public opinions, perceptions, attitudes, and behaviors for analyses in many social science disciplines. Yet, there is little knowledge regarding how specific elements of survey research methodology are applied in practice. This article examines four mainstream public administration journals over an 8-year period regarding survey data collection modes, sampling techniques, sample sizes, response rates, and data analytical techniques. The findings show that survey research in the field of public administration features mainly small scale studies, heavy reliance on a single specific mode, questionable sample selection procedures and suspect sample frame quality. Survey data are largely analyzed with less sophisticated statistical techniques. An informed evaluation of the quality of survey data is made more difficult by the fact that many journal articles do not detail data collection procedures. This study will offer suggestions for improving the quality of survey research in the field.

Key words: Survey Research, Research Methodology, Public Administration Research

Introduction

Since McCurdy and Cleary (1984) first discussed quality issues in doctoral dissertations in Public Administration (PA) in 1984, a number of the scholars have also explored the quality of research methodologies used in the field (for example, Box 1992; Cleary 1992, 2000; Forrester and Watson 1994; Houston and Delevan 1990; McCurdy and Cleary 1984; Perry and Kraemer 1986; Stallings 1986; Stallings and Ferris 1988; White 1986; White, Adams, and Forrester 1996). The majority of these reviews have examined quantitative research. Among the quantitative methods currently available, survey research is perhaps the most widely recognized and applied in PA. (Enticott, Boyne, and Walker 2009).

Survey research is a data collection methodology in which samples are drawn, respondents are interviewed, and data is analyzed in order to extrapolate to a population of interest. The survey instrument allows researchers to assess, with a small sample, people's attitudes, perceptions, and opinions on particular social issues, as well as factual knowledge, in a target population (Swidorski 1980). For this efficiency, it has gained popularity in many academic disciplines such as sociology, political science, marketing research, public health and communication studies, as well as public administration (Folz 1996). In addition to academic domains, practitioners at all levels of government have increasingly turned to these techniques to measure citizens' needs and feedbacks (Daneke and Klobus-Edwards 1979).

With its growing popularity, the demand for quality survey research has been increasing in professional fields as well as academia. No longer considered merely a

method, survey research is now also recognized as an independent academic discipline [1] in the U.S. Largely in acknowledgement of the importance of continually developing the rigor of this methodology, most university programs in survey research methodology emphasize a reduction in survey errors in order to maximize the congruency between sample estimates and population parameters.

Although survey research is an important tool, there are no reviews available that examine the conduct and reporting of survey research in the empirical literature. The purpose of this article is to review the survey research methods now employed by researchers in public administration. Specifically, we aim to assess data collection procedures, analytic issues, and reporting practices, and offer suggestions for quality survey data.

Previous Reviews of Public Administration Research

Research reviews in public administration first began with investigations of the quality of doctoral dissertations, and were subsequently extended to debates over the quality of scholarly journal articles. Through these review studies on PA research over the past 25 years, PA scholars have diagnosed the state and rigor of contemporary research and suggested constructive ideas for improving research quality.

McCurdy and Cleary (1984) first evaluated the quality of 142 PA doctoral dissertations written in 1981 using six criteria: research purpose, validity, testing of theory, causal relationships, importance of the topic, and whether or not the topic was considered cutting edge. They found that the majority of dissertations failed to meet these basic criteria, and claimed that the quality of PA research lagged behind that of

mainstream social science research. These findings were confirmed by White (1986), who evaluated 305 dissertations using similar criteria. In addition, White found that a large amount of dissertation research was never published in journals, limiting the contribution of dissertations to knowledge development. Both studies expressed serious concerns about the lack of rigorous research methodology in the field.

Issues regarding the research quality of doctoral dissertations subsequently directed scholarly attention to the quality of research published in professional scholarly journals. Perry and Kraemer (1986), and Stallings and Ferris (1988) investigated research articles in Public Administration Review (PAR) to examine the quality of research methodology for professional scholarly works. Both studies identified a need for more advanced statistical techniques and improvement of methodology in the work published in that journal.

Subsequently, Houston and Delevan (1990) examined a broader variety of PA journals. These authors argued that research in public administration was underfunded, non-empirical, and not oriented towards theory-testing. They went on to argue that the research designs employed in published journal articles were monotonous and lacked rigor. They recommended increased utilization of more rigorous research designs, including quasi-experiments and experiments. The authors concluded that low quality research in PA might be due in part to poor research training in PA programs (see also Adams and White 1995; Cozzetto 1994; DeLorenzo 2000; McCurdy and Cleary 1984; Stallings 1986; White 1986).

Cozzetto (1994) narrowed his focus to address the quality of quantitative research articles in order to assess their level of statistical sophistication. He found that 40 percent

of sampled articles incorrectly used statistical techniques, and 83 percent of the articles showed a lack of methodological sophistication. In another assessment of quantitative research, Wright and his colleagues (2004) also emphasized quantitative research, raising issues of potential biases in the process of data collection and inadequate information on measurement in journal articles.

While invaluable, these prior reviews have of necessity been very broad in their scope and consequently unable to examine specific details of the methodologies examined. In contrast, this study will focus more narrowly on one approach, survey research methodology, in order to provide a more detailed assessment of its applications within the PA literature. This will include an assessment of current practices and potential sources of survey related error. Some suggestions for the improvement of quality survey research in the field will also be offered. We select sample surveys for this assessment as it is perhaps most commonly employed data collection methodology in public administration.

Sources of Error in Survey Research

Researchers recognize multiple sources of survey-related errors. Those investigating error sources have been classified into a typology by Groves (1987), who identifies '*data collectors*,' '*analysts*,' '*describers*,' '*modelers*,' '*measurers*,' and '*reducers*.' (p156). Data collectors are interested in the data collection process and analysts focus on using survey data for a particular issue; describers are mainly interested in population estimation with survey data and modelers focus on model specification for testing causal theories; measurers tend to estimate survey errors and reducers are primarily concerned

with reducing survey errors in order to produce quality survey data (Groves 1987). Academic researchers who employ survey research methods are primarily modelers who are interested in issues of model specification for purpose of theory testing.

The Total Survey Error (TSE) framework addresses all possible sources of error that can bias survey findings, including sampling error, coverage error, nonresponse error, and measurement error. (Groves 1987, 1989). The last three errors are classified as nonsampling errors. Sampling error arises when the entire target population is not taken as a sample for surveys, and decreases as the sample size increases. The best way of minimizing sampling error is thus to increase the size of random samples (while also controlling for other sources of survey error). Coverage error takes place when a sample frame does not fully represent the total population sampled, which results in selection bias, leading to generalizability issues. Nonresponse error occurs when there are systematic differences in responses between respondents and total sampled persons. Generally, enhancing response rates is considered a good strategy to minimize this source of error (Fowler 2002). The sources of measurement error are many, including interviewer, respondent, and questionnaire. In face-to-face and telephone surveys, an interviewer can influence respondents to answer in a certain direction, generating interviewer variance in which average responses to a particular variable varies across interviewers. A typical example of measurement error due to respondents would involve inaccurate answers to retrospective questions because of recall or retrieval problems. Likewise, social desirability demands may systematically bias responses to some survey questions. Poorly designed survey questionnaires that are ambiguous or overly complicated make it difficult for respondents to comprehend and answer adequately.

Quality survey data can be obtained when TSE is minimized. More importantly, increased bias and variance due to survey errors may negatively affect analytical results–particularly, overestimation (or underestimation) of descriptive statistics and biased estimation or type II errors in regression analyses (Biemer and Lyberg 2003). For instance, survey measurement error due to social desirability may establish spurious relationships between variables in regression analysis (Fisher 1993; Moorman and Podsakoff 1992; Zerbe and Paulhus 1987). In addition, selection bias can influence analytical results in survey research. Enticott, Boyne, and Walker (2009), for example, show that a sample which contains multiple informants provides more accurate measures of organizational properties than can be obtained with only a single informant. While important as a research methodology, it is clear that survey research can generate unreliable data which, in turn, can generate biased analytical results unless the variety of errors that commonly surround the research process are properly controlled. Hence, a survey researcher needs to be a 'data collector,' 'analyst,' 'describer,' 'modeler,' 'measurer,' and 'reducer' all at once.

Data and Methods

To investigate the survey research methods currently being applied in the field of public administration, we examined four peer-reviewed journals listed in the Social Science Citation Index (SSCI): American Review of Public Administration (ARPA), Administration & Society (AS), Journal of Public Administration Research and Theory (JPART), and Public Administration Review (PAR). These four journals were chosen because they have been frequently employed to represent the mainstream PA research

literature (see Brower, Abolafia, and Carr 2000; Forrester and Watson 1994; Lan and Anders 2000; Wright, Manigault, and Black 2004).

We reviewed these four journals for the eight-year period between 2000 and 2007, selecting a total of 245 articles [2] reporting either primary survey or secondary survey data. One coder systematically reviewed all articles, coding each study characteristic of interest for this review. Then, we selected 24 articles (about 10 percent of the total) via systematic random sampling, and a second coder reviewed and recoded these. We used Cohen's kappa to examine the consistency between both coders' work for these 24 articles in order to judge coding reliability. Findings of this analysis , reported in the Appendix, indicate that all coefficients for each study design feature examined were 0.8 or greater; attaining a kappa value of 0.6 or higher is considered a substantial level of mutual agreement (Landis and Koch 1977). Having examined kappa with only a small sample of articles, we calculated confidence intervals to predict the upper and lower levels of kappa coefficients for the total population of articles. Based on this analysis, we are 95 percent confident that the kappa coefficient from the full sample of articles lies between 0.59 and 1.00 for all coded contents. Accordingly, we concluded that our coding of these articles was substantially reliable.

Findings

Our coding scheme differentiated primary and secondary research, though there were some unidentified articles where the distinction was difficult to judge. Primary survey research includes both data collection and analysis, while secondary survey research focuses on the application of analytical techniques to existing survey data

collected by others (Kiecolt and Nathan 1985). As shown in Table 1, 135 articles (55 percent of the total) reporting primary survey research were published in the four journals during this eight year period, while 92 articles (or 38 percent) reported secondary survey analysis. There were 18 articles which did not specify whether the survey data being presented were primary or secondary survey data.

Table 1 shows that 95 percent of the primary studies reviewed were collected using a cross-sectional survey design; five percent collected longitudinal data at multiple time points. Overall, 90 percent (221 articles) of the survey data sets employed in the articles explained were cross-sectional and 10 percent (24 articles) were longitudinal. Over 20 years ago, Perry and Kraemer (1986) found a similar pattern of heavy reliance on cross-sectional data in PA research.

[Table 1 here]

Data Collection Modes

Table 2 examines the types of survey modes employed in PA survey research, including mail, web, in-person, telephone, and mixed mode methodologies.[3] Though computer-assisted instruments [4] have been widely utilized in large scale surveys such as the General Social Survey (GSS) or the American National Election Survey (ANES), in practice we found that computer-assisted technologies were rarely reported in primary survey research in PA.

Surveys are either self-administered or interviewer-administered. Self-administered modes include mail and web surveys, whereas interviewer-administered surveys are generally completed via telephone or in-person. Of 135 articles reporting primary surveys, 70 percent (94 articles) involved mail surveys, followed by mixed mode surveys (8 percent), telephone interviews (6 percent), web surveys (4 percent) and in-person interviews (4 percent). In secondary survey research studies, about 36 percent of all studies used mail surveys; followed by in-person interviews (10 percent), telephone interviews (9 percent), and the web (2 percent). A somewhat surprising 41 percent of all secondary studies did not specify the method of survey administration, which is much higher than among primary surveys (8 percent). Overall, 55 percent of all studies reported that surveys were conducted via mail. Hence, the most utilized survey mode in the field is a mail survey.

[Table 2 here]

In terms of nonsampling error, one potential problem for mail surveys is nonresponse bias (Bridge 1974; Ellis, Endo, and Armer 1970; Filion 1975; Fowler 2002; Kanuk and Berenson 1975; Wright, Manigault, and Black 2004). In contrast, measurement error stemming from social desirability bias is generally believed to be reduced in self-administered modes such as mail surveys (Fowler 2002). Social desirability bias arises when survey questions are viewed by respondents as being sensitive. Sensitive topics tend to be more intrusive and include the threat of disclosure

such as criminal activities, sexual behavior, and voting behavior (Tourangeau and Yan 2007).

Given that the predominant topic of PA research employing survey research methods is organizational or administrative behavior, in which respondents' motivations, attitudes, and perceptions of work and organizations are examined (rather than more personal and/or private matters), social desirability bias would seem to be less of a concern in PA, compared to other behavioral sciences. Accordingly, loss due to nonresponse bias may outweigh the gain due to reducing socially desirable bias via mail surveys in PA research.

Sampling Methods

In general, two types of sampling techniques are used in social surveys: probability and nonprobability sampling. Probability sampling techniques involve the selection of samples from a defined target population using a random mechanism such that every sample unit in the target population has a known probability of selection. In contrast, nonprobability sampling, also referred to as convenience sampling, does not rely on random selection. Instead, samples are collected based on non-random mechanisms which render it impossible to know the probability of selection for each sample unit (Folz 1996; Fowler 2002; Henry 1990). It is widely acknowledged that nonprobability samples suffer from selection bias threatening the representativeness of a sample survey (Henry 1990; Kalton 1983). Obviously, probability sampling is much preferred over nonprobability methods in terms of the quality of survey data in academic research.

Table 3 exhibits the general sampling techniques that were reported in the articles reviewed. About 29 percent of all articles reported that surveys were undertaken with probability sampling methods. Approximately five percent of all articles indicated that data were collected using nonprobability methods. It is surprising to find that approximately 55 percent of articles did not specify how sampled persons were selected, making it impossible for readers to judge the quality of the sampling plan.

Of the probability sample designs reported, the simple random sampling method was most common; 27 percent overall; 31 percent of primary survey studies; and 21 percent of secondary survey studies (see table 4). In total, 24 percent of the probability samples employed stratification, 13 percent used multi-stage sampling, and one percent employed cluster sampling. About 35 percent of these surveys used probability sampling without specifying the specific type of sample design.

[Table 3 here]

[Table 4 here]

Sample Size

Reporting sample size is a fundamental requirement when presenting survey research findings (Johnson and Owens 2003). Table 5 indicates that, of the total of 245 journal articles reviewed, 84 percent (N=205 studies) reported survey sample size.[5] Reporting this information was more common for studies using primary data (93 percent) than for those using secondary data (72 percent). Few researchers, of course, are successful in

collecting data from all units sampled for a survey. Commonly referred to as unit nonresponse, this occurs for many reasons, most typically because sampled respondents are unwilling or unable to participate.

Under random sampling, sample size is negatively related to sampling errors: the smaller the sample size, the higher the sampling errors. Higher sampling errors due to small sample sizes lead to larger confidence intervals which, in turn, produce less accurate predictions. A large sample size has more precision and usually better statistical properties in analytical models. Large sample sizes, however, do not reduce nonsampling errors, which in many cases are more influential in determining the quality of survey data. In other words, a large sample size does not guarantee quality survey data unless nonsampling errors are also controlled.

[Table 5 here]

Figure 1 depicts the distribution of initial sample size by survey type. About 32 percent of primary survey studies initially selected samples of more than 1,000, whereas nearly 60 percent of secondary data did so. On the other hand, 42 percent of primary surveys reported that the initial sample size was less than 500 compared to 10 percent of the secondary studies reviewed. While the mean initial sample size for primary surveys was 1,376 (standard deviation: 2,861), that of secondary data was 4,338 (standard deviation: 9,332).

Figure 2 illustrates how distributions of completed sample size differ between the two survey types. Analyses with more than 500 respondents were more common in studies that examined secondary survey data than in primary survey research: 72 percent vs. 35 percent, respectively. About 65 percent of the articles that used primary survey data analyzed samples of less than 500 cases.

[Figure 1 here]

[Figure 2 here]

Response Rates

There are multiple methods for calculating survey response rates. The American Association for Public Opinion Research (AAPOR 2008) provides six formulas [6] for calculating response rates for surveys with probability sampling, but does not offer any methods for surveys with nonprobability sampling. AAPOR standards are used to calculate response rates for telephone, in-person, mail, and web surveys. In order to accurately compute response rates, the construction of a clearly defined sample frame is required.

As shown in table 6, only about 28 percent (primary: 33 percent; secondary: 25 percent) of the articles examined reported constructing a formal sample frame for sample selection. A majority of the articles (70 percent) did not specify the type of sample frame used. A typical example of a mail survey that did not employ a frame would be an instance in which the researcher reported an unspecified mail-out to a governmental agency or department in which self-selected persons were free to complete questionnaires.

In order to accurately compute response rates recommended by AAPOR, it is essential to have a final case status (or disposition) for each sampled case (i.e., whether each cooperated, refused, was ineligible or not available, etc). There is no way of knowing such information in cases where respondent's self-select.

[Table 6 here]

Table 7 presents information on response rate reporting. In total, 69 percent (170 studies) of the studies examined response rates for their surveys: 79 percent for primary and 57 percent for secondary. As shown in figure 3, the distribution of response rate categories is approximately normal in shape for both primary and secondary surveys. The values of mean and median are quite similar: 55.9 vs. 56.3 for primary, and 53.4 vs. 53.4 for secondary. About 50 percent of all primary surveys lie between 40 and 60 percent in response rates, whereas about 70 percent of secondary surveys fall into the same range. The mean response rate for primary survey research was 55.9 percent, which was remarkably close to the mean (53.4 percent) for secondary research.

It is important to note that none of the studies reported following the AAPOR standards to compute response rates in primary surveys, instead simply reporting a ratio between the original number of questionnaires and the number completed. In secondary analyses of survey data, several large-scale data sets (e.g., GSS) reported using the AAPOR definitions for response rate calculations.

[Table 7 here]

[Figure 3 here]

Target Population

A target population is the group to which researchers are interested in generalizing their findings. Table 8 reveals that about 56 percent of the studies examined were focused on the public sector, including federal, state, and local government employees in each survey category. Overall, excluding the "other" category, employees working in local governments most commonly served as study subjects (34 percent), followed by those in state governments (12 percent). Employees in nonprofit organizations (10 percent) and citizens (11 percent) were also common target populations in these studies. It is clear that issues of local government, the behaviors of public managers in local government, and local governance are central themes in empirical studies in our field. Research on public perceptions of administration (or policy), civil participation, and public attitudes have mainly used secondary data: primarily public polls and national survey data.

Coverage errors are a type of nonsampling error related to the target population, and occur when the target population and sample frame are mismatched. In other words, if the sample frame list does not cover the complete population to be studied, it suffers from undercoverage bias (Groves et al. 2004). An example of this problem in a PA survey might be if a target of public managers in a certain county was being sampled via a sample frame of public officials in several large cities within the county, while several small townships were not included. As previously noted, few studies specify how the

researchers actually constructed their sample frame and how well it overlaps with the target population.

[Table 8 here]

Statistical Techniques Used to Analyze Survey Data

Table 9 illustrates the main statistical techniques that researchers applied in the studies reviewed.[7] Nearly 47 percent (106 studies) of all studies investigated were designed to test formal hypotheses in order to verify theories or previous findings, and about 53 percent (or 121 studies) did not use hypothesis testing and were more descriptive in nature. Survey research within this context thus appears to have been somewhat a bit more focused on exploring administrative phenomenon than on theory verification.

The statistical techniques employed in 37 percent of the primary survey studies employed basic univariate or bivariate methods such as descriptive statistics, t-tests, χ^2-tests and Pearson correlations, while only 18 percent of the secondary survey research studies used these methods. Primary survey studies thus more commonly relied on simple techniques than did secondary survey research studies. Regression methods such as OLS, logistic, multinomial regression were used for 48 percent of the primary survey studies examined, while 65 percent of the secondary survey studies used these multivariate methods.

In total, OLS regression analysis (36 percent) was the dominant method used. Nearly 30 percent used only simple methods including descriptive statistics and bivariate correlation, while 15 percent used more advanced techniques such as weighted least

square (WLS), 2 stage least squares (2SLS), structural equation modeling (SEM), hierarchical linear modeling (HLM), and time series analysis. In summary, about 65 percent of all survey studies reviewed here were analyzed with less sophisticated methods such as linear regression analysis, descriptive statistics, and simple bivariate tests. This finding supports previous reviews claiming that PA empirical research relies on less advanced statistical techniques, relative to other social sciences (see Cozzetto 1994; DeLorenzo 2000; Gill and Meier 2000; Perry and Kraemer 1986).

[Table 9 here]

Discussion and Suggestions

Prior research evaluations have been largely focused on measurement issues and analytical techniques, and often assume that data quality is perfect. This study highlighted survey data collection methodologies in order to assess the quality of survey data being reported in PA journals. Our findings show that survey research in PA features small-scale studies, heavy reliance on a single specific mode (the mail survey), questionable sample frame quality, suspect sample selection procedures, and less sophisticated analytical tools.

Quality Survey Data

Our findings clearly show that survey research relies heavily on self-administered instruments in PA research. The advantages of self-administered surveys are many: they

are less costly, relatively easy to conduct, are not affected by interviewer-related bias, and are less susceptible to socially desirable responses (Bridge 1974; Fowler 2002; Kanuk and Berenson 1975). One important concern with this particular mode, however, is a potentially high degree of nonresponse bias due to low response rates (Bridge 1974; Ellis, Endo, and Armer 1970; Filion 1975; Kanuk and Berenson 1975; Wright, Manigault, and Black 2004).

The main reason for the preference for mail surveys in PA survey research is likely to be funding constraints: mail surveys are less costly to carry out. Lack of financial resources may be an impediment to the application of more sophisticated survey data collection methods among PA researchers (Gill and Meier 2000; Perry and Kraemer 1986). It is worth noting that there is a tradeoff between costs and data quality in surveys (Groves 1989).

We make no claim that the mail survey is a problematic data collection strategy. Rather, we regard it as an efficient and practical tool that can be undertaken within the constraints of research funding, as long as nonresponse bias and measurement errors are adequately controlled. The use of pre-/post-notification (Assael and Keon 1982; Kanuk and Berenson 1975), monetary incentives (Armstrong 1975; Brown and Coney 1977; Goodstadt et al. 1977; Kanuk and Berenson 1975), mixed modes, or the employment of special questionnaire design protocols such as the Tailored Design Method [8] (Dillman 1999) are some recommended approaches to maximize response rates in mail surveys for PA research.

In addition, efforts at reducing the measurement errors that may stem from complex questionnaires or poor question wording in survey items must be considered. In order to

reduce these error sources, focus groups, expert review, and pretesting are necessary for the refinement of questionnaires (for example, Leeuw, Borgers, and Smits 2004), yet we rarely observed the reporting of such activities while reviewing these articles.

The common goal of sample surveys is to accurately estimate a population of interest with a small sample. In order to do so, the sample must be representative of the study population. Sample representativeness, known as external validity, is a cornerstone of the sample survey. In this respect, the quality of a sample frame and sampling methods exerts a considerable influence on the representativeness of a survey. As shown in our findings, nearly 70 percent of the total studies did not detail whether a frame was used or not when samples were drawn. An imperfect or nonexistent sampling frame can give rise to coverage problems, throwing into doubt the representativeness of the sample.

Another important element for assessing representativeness is whether probability sampling methods are employed. Approximately 65 percent of the surveys reviewed did not specify how respondents were selected. As with all other social science methodologies, those employing surveys are obligated to inform readers as to the degree to which their data may be biased. Papers in psychology, for instance, are generally expected to report whether subjects are assigned at random to experimental and control groups when an experimental design is utilized in a study (APA 2008). This reflects the recognition that random assignment is an integral element of experimental research as a criterion of research quality. Similarly, whether probability sampling is used or not is a critical issue in judging the quality of survey data. Of course, there are multiple types of probability sampling commonly employed in survey research.

Probability sampling is a technique by which every sampled person has a known probability of selection into a sample. Convenience sampling appears to be efficient when it is impossible to construct sample frame (e.g., when sampling rare and/or highly dispersed populations). Survey results with convenience sampling, however, are difficult to generalize because it is unknown to what extent the sample represents the population of interest. Moreover, there is no way of estimating the degree of precision of estimates attained from convenience sampling (Lavrakas 1987). At present, AAPOR recommends against the reporting of a margin of errors for survey data from nonprobability samples. In addition, a fundamental assumption of many statistical models such as OLS regression is that samples are drawn at random (Wooldridge 2002). This serves as an important reminder that random selection from a well-defined sample frame is an integral assumption of survey data analysis. However, as we noted previously, few studies specified as to how their sample frames were constructed. The quality of the frame determines the degree of coverage error in survey research. If the frame list does not fully represent a target population, generalizability issues again emerge.

Survey Data Analytical Techniques

Numerous studies have documented that empirical research in PA lacks statistical sophistication in analyzing quantitative data (Cozzetto 1994; DeLorenzo 2000; Gill and Meier 2000; Perry and Kraemer 1986). Our findings confirm this conclusion. OLS regression and simple univariate/bivariate techniques, including descriptive statistics account for over 65 percent of the published studies that were reviewed here.

In analyzing survey data, we found that some researchers ignored the underlying assumptions of statistical techniques, leading to inaccurate results. The most popular statistical technique employed in the studies reviewed was OLS regression. There are some critical assumptions which must hold for OLS to be used, including independent observations, a lack of multicollinearity, and homoskedasicity of residuals. These assumptions must be satisfied in order to obtain valid results using OLS. The assumption that every observation is independent, however, may be easily violated when each observation is nested within upper contexts such as agencies, schools, neighborhoods and work units because observations within the same context are more homogeneous (or heterogeneous between contexts) in behaviors and attitudes. In this particular case, rather than OLS, another method such as hierarchical linear modeling (HLM) should be used to address the problem.

Also, Generalized Least Squares (GLS), Weighted Least Squares (WLS), and 2-Stage Least Squares (2SLS) are more complicated variants of OLS that can be used if some assumptions of OLS are violated. Few studies detailed a rationale for choosing a statistical technique or revealed a consideration of their underlying assumptions before analyzing their data.

Reporting Issues

It is important to also acknowledge the importance of careful reporting of survey data collection, as well as analytical findings. Our findings, however, suggest that many key elements of survey methodology including sampling procedures, sample frame construction, response rates, and sample sizes are often not fully documented in PA

mainstream journals. In particular, papers reporting secondary survey data analyses are more likely to fail to provide this basic information. This can leave readers unable to judge the quality of the data being reported.

One might think that there is no need to specify the description of data collection procedures in each research paper, when multiple papers are published using the same data. Each paper, though, should be independent in describing its methodology. Wright and his colleagues (2004) here previously raised the issue of the scarcity of information provided regarding methodological procedures and measures in the published literature. Two potential reasons for this scarcity include: (1) many professional journals have no explicit guidelines or policies regarding the full disclosure of data collection procedures (Johnson and Owens 2003); and (2) authors omit this information because reviewers of journals demand condensed papers due to space limitations (Luton 2007; Lynn, Heinrich, and Hill 2001).

We conclude that journal editors and reviewers should not only pay attention to *how-to-best-analyze*, but also to *how-to-best-collect* survey data, and insure that the audience has available sufficient information with which to judge the quality of the survey data underlying analytical results. Thus, to the extent that they do not currently exist, we encourage journals to establish standards for methodological transparency in the disclosure of survey data collection procedures.

NOTES

1. Currently, in the U.S., graduate degree programs in survey methodology are run at the University of Maryland-College Park, University of Michigan, and University of Nebraska-Lincoln; other graduate programs in survey methodology are at George Washington University, the Ohio State University, University of Cincinnati, University of Illinois-Chicago, and University of North Carolina-Chapel Hill.

2. We considered only quantitative survey methods, excluding in-depth interviews and focus groups as qualitative research. In coding, when authors published several articles using the same primary survey data, we counted them multiple times as primary research. Studies using multiple data sources -a combination of survey data and other types of data such as administrative records, - we classified it as survey research.

3. Mixed mode includes surveys in which two or more different survey modes are used.

4. Computer-Assisted Interviewing (CAI) is a data collection mode in which interviewers use computers, rather than paper and pencil questionnaires.

5. We coded initial and completed sample sizes. The former refers to a sample initially selected; the latter is one in which unit nonresponse cases are excluded. Some articles indicated both sample sizes, other articles reported either initial or completed sizes. For a case where either initial or complete sample size was reported, we calculated the sample size that was not provided using reported response rates and reported sample size (initial or completed sample), and coded the study as reporting sample size. When multiple data sources were used, we coded the sample size with the arithmetic average of each for both initial and completed sample sizes.

6. AAPOR provides six formulas for calculating response rates: RR1 through RR6. In order to apply these formulas, a survey must be undertaken using probability sampling methods and a clearly defined sample frame. The components of the formulas include complete interview, partial interview, refusal and break-off, non-contact, etc. Most journal articles focused on mail surveys indicated that response rates were calculated as a ratio of the number of returned questionnaires to the total number of mailed-out questionnaires. This, strictly speaking, is a complete rate or mail return rate, not a response rate. The AAPOR standard considers various factors such as the sampled persons' eligibility, unknown eligibility, and questionnaires' total vs. partial completion. In order to compute response rates for mail surveys, a sample frame containing a list of the target population should exist.

7. For studies using multiple statistical analyses, we chose only one technique that was most relevant to research questions or hypotheses.

8. The Tailored Design Method (TDM) is a technique for questionnaire design in which words, pictures, and coloring are mobilized to produce user-friendly questionnaires, differently from a traditional written format. This new method is designed to improve the

quality of survey data in self-administered surveys such as these completed via mail and web. Many empirical findings show that this method can improve the understanding of respondents for survey questions as well as survey cooperation in surveys (Dillman 1999).

REFERENCES

AAPOR. 2008. Standard Definitions: Final disposition of case codes and outcome rates for surveys.

Adams, Guy B., and Jay D. White. 1995. The Public Administration Doctorate: A Degree in Search of a Purpose. *Journal of Public Administration Education* 1 (1):67-76.

APA. 2008. Reporting Standards for Research in Psychology: Why do we need them? What might they be? . *American Psychologist* 63:839-851.

Armstrong, J.Scott. 1975. Monetary Incentives in Mail Surveys. *Public Opinion Quarterly* 39:111-116.

Assael, Henry, and John Keon. 1982. Nonsampling vs.Sampling Errors in Survey Research. *Journal of Marketing Research* 46:114-123.

Biemer, Paul P., and Lars E. Lyberg. 2003. *Introduction to Survey Quality*. Hoboken, NJ: Wiley.

Box, Richard C. 1992. An Examination of the Debate over Research in Public Administration. *Public Administration Review* 52:62-70.

Bridge, R. Gary. 1974. Nonresponse Bias in Mail Surveys: The Case of the Department of Defens Post-Service Survey. Santa Monica: Rand.

Brower, Ralph S., Mitchel Y. Abolafia, and Jered B. Carr. 2000. On Improving Qualitative Methods in Public Administration Research. *Administration & Society* 32:363-397.

Brown, Stephen W., and Kenneth A. Coney. 1977. Comments on "Mail Survey Premiums and Response Bias". *Journal of Marketing Research* 14:385-7.

Cleary, Robert E. 1992. Revisiting the Doctoral Dissertation in Public Administration: An Examination of the Dissertations of 1990. *Public Administration Review* 52:55-61.

———. 2000. The Public Administration Doctoral Dissertation Reexamined: An Evaluation of the Dissertation of 1998. *Public Administration Review* 60:446-455.

Cozzetto, Don A. 1994. Quantitative Research in Public Administration: A Need to Address Some Serious Methodological Problems. *Administration & Society* 26 (3):337-343.

Daneke, Gregory A., and Patricia Klobus-Edwards. 1979. Survey Research for Public Administrators. *Public Administration Review* 39 (5):421-426.

DeLorenzo, Lisa. 2000. Stars Aren't Stupid, But Our Methodological Training Is: A Commentary on Jeff Gill and Ken Meire' s Article "Public Administration Research and Practice: A Methodological Manifesto". *Journal of Public Administration Research and Theory* 11 (1):139-145.

Dillman, Don A. 1999. *Mail and Internet Surveys: The Tailored Design Method*. New York: John Wiley & Sons.

Ellis, Robert A., Calvin M. Endo, and J. Michael Armer. 1970. The Use of Potential Nonrespondents for Studying Nonresponse Bias. *The Pacific Sociological Review* 13:103-109.

Enticott, Gareth, George A. Boyne, and Richard M. Walker. 2009. The Use of Multiple Informants in Public Administration Research: Data Aggregation Using Organizational Echelons. *Journal of Public Administration Research and Theory* 19:229-253.

Filion, F.L. 1975. Estimating Bias Due to Nonresponse in Mail Surveys. *The Public Opinion Quarterly* 39:482-492.
Fisher, Robert J. 1993. Social Desirability Bias and the Validity of Indirect Questioning. *Journal of Consumer Research* 20:303-315.
Folz, David H. 1996. *Survey Research for Public Administration*. Thousand Oaks: SAGA Publications.
Forrester, John P., and Sheilah S. Watson. 1994. An Assessment of Public Administration Journals: The Perspective of Editors and Editoral Board Members. *Public Administration Review* 54 (5):474-485.
Fowler, Floyd J. 2002. *Survey Research Methods*. Thousand Oaks: Sage Publications.
Gill, Jeff, and Kenneth J. Meier. 2000. Public Administration Research and Practice: A Methodological Manifesto. *Journal of Public Administration Research and Theory* 10 (1):157-199.
Goodstadt, Michael S., Linda Chung, Reena Kronitz, and Gaynoll Cook. 1977. Mail Survey Response Rates: Their Manipulation and Impact. *Journal of Marketing Research* 14:391-5.
Groves, Robert M. 1987. Research on Survey Data Quality. *Public Opinion Quarterly* 51:S156-72.
———. 1989. *Survey Errors and Survey Costs*. New York: Wiley-Interscience.
Groves, Robert M., Jr. Floyd J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2004. *Survey Methodology*. Hoboken: Wiley Interscience.
Henry, Gary T. 1990. *Practical Sampling*. Newbury Park, CA: SAGE Publications, Inc.
Houston, David J., and Sybil M. Delevan. 1990. Public Administration Research: An Assessment of Journal Publications. *Public Administration Review* 50 (6):674-681.
Johnson, Timothy P., and Linda Owens. 2003. Survey Response Rate Reporting in the Professional Literature. In *American Association for Public Opinion Research*. Nashville, TN.
Kalton, Graham. 1983. *Introduction to Survey Sampling*. Newbury Park, CA: SAGE Publications.
Kanuk, Lelie, and Conrad Berenson. 1975. Mail Surveys and Response Rates: A Literature Review. *Journal of Marketing Research* 12:440-53.
Kiecolt, K. Jill, and Laura E. Nathan. 1985. *Secondary Analysis of Survey Data*. Newbury Park: SAGE Publications.
Lan, Zhiyong, and Kathleen K. Anders. 2000. A Paradigmatic View of Contemporary Public Administration Research: An Empirical Test. *Administration & Society* 32 (2):138-165.
Landis, J. R, and G. G Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33:159-174.
Lavrakas, Paul J. 1987. *Telephone Survey Methods*. Newbury Park: SAGE.
Leeuw, Edith de, Natacha Borgers, and Astrid Smits. 2004. Pretesting Questionnaires for Children and Adolescents. In *Methods for Testing and Evaluating Survey Questionnaires*. Hoboken, NJ: Wiley & Sons.
Luton, Larry S. 2007. Deconstructing Public Administration Empiricism. *Administration & Society* 39:527-544.

Lynn, Jr., Laurence E. , Carolyn J. Heinrich, and Carolyn J. Hill. 2001. *Improving Governance: A New Logic for Emprical Research*. Washington, D.C.: Georgetown University Press.

McCurdy, Howard E., and Robert E. Cleary. 1984. Why Can't Resolve the Research Issue in Public Administration? . *Public Administration Review* 44 (1):49-55.

Moorman, R.H., and P.M Podsakoff. 1992. A Meta-Analytic Review and Empirical Test of the Potential Confounding Effects of Social Desirability Response Sets in Organizational Behavior Research. *Journal of Occupational and Organizational Psychology* 65:131-149.

Perry, James L., and Kenneth L. Kraemer. 1986. Research methodology in the Public Administration Review, 1975-1984. *Public Administration Review* 46:215-226.

Stallings, Robert A. 1986. Doctoral Programs in Public Administration: An Outsider's Perspective. *Public Administration Review* 46 (3):235-240.

Stallings, Robert A., and James M. Ferris. 1988. Public Administration Research: Work in PAR, 1940-1984. *Public Administration Review* 48 (1):580-587.

Swidorski, Carl. 1980. Sample Surveys: Help for the "Out-of-House" Evaluator. *Public Administration Review* 40 (1):67-71.

Tourangeau, Roger, and Ting Yan. 2007. Sensitive Questions in Surveys. *Psychological Bulletin* 133 (5):859-883.

White, Jay D. 1986. Dissertations and Publications in Public Administration. *Public Administration Review* 46 (3):227-234.

White, Jay D., Guy B. Adams, and John P. Forrester. 1996. Knowledge and Theory Development in Public Administration: The Role of Doctoral Education and Research. *Public Administration Review* 56:441-452.

Wooldridge, Jeffrey. 2002. *Introductory Econometrics: A Modern Approach*. Chula Vista: South-Western College.

Wright, Bradley E., Lepora J. Manigault, and Tamika R. Black. 2004. Quantitative Research Measurement in Public Administration: An Assessment of Journal Publications. *Administration & Society* 35 (6):747-764.

Zerbe, W.J, and D. L Paulhus. 1987. Social Desirable Responding in Organizational Behavior: A Rreconception. *Academy of Management Review* 12:250-264.

Table 1. Number (%) of Journal Articles Employing Survey Research, By Journal, Survey Type, and Survey Design: 2000-2007

Journal	Survey Type	Survey Design		Total
		Cross-sectional	Longitudinal	
ARPA	Primary	26(96)	1(4)	27(100)
	Secondary	15(72)	3(28)	18(100)
	Did not specify	2(50)	2(50)	4(100)
		43(88)	6(12)	49(100)
AS	Primary	21(91)	2(9)	23(100)
	Secondary	12(86)	2(14)	14(100)
	Did not specify	1(100)	-	1(100)
		34(89)	4(11)	38(100)
JPART	Primary	29(97)	1(3)	30(100)
	Secondary	31(79)	8(21)	39(100)
	Did not specify	4(67)	2(33)	6(100)
		64(85)	11(15)	75(100)
PAR	Primary	52(91)	3(9)	55(100)
	Secondary	21(100)	-	21(100)
	Did not specify	7(100)	-	7(100)
		80(96)	3(4)	83(100)
Primary subtotal		128(95)	7(5)	135(55)†
Secondary subtotal		79(86)	16(14)	92(38)‡
Total (%)		221(90)	24(10)	245(100)

ARPA: American Journal of Public Administration, AS: Administration & Society
JPART: Journal of Public Administration Research and Theory, PAR: Public Administration Review
† Percentage is a ratio of primary subtotal to total
‡ Percentage is a ratio of secondary subtotal to total

Table 2. Number (%) of Journal Articles Employing Survey Data, By Data Collection Mode and Survey Type: 2000-2007

	Self-administered survey		Interviewer-administered survey		Mixed mode Survey	Did not specify	Total
	Mail	Web	In-person	Telephone			
Primary	94 (70)	5 (4)	5 (4)	9 (6)	11 (8)	11 (8)	135 (100)
Secondary	33 (36)	2 (2)	8 (10)	8 (9)	2 (2)	39 (41)	92 (100)
Did not specify	7 (39)	0 (0)	1 (6)	3 (11)	2 (11)	5 (33)	18 (100)
Total	134 (55)	7 (3)	14 (6)	20 (8)	15 (5)	55 (23)	245 (100)

Table 3. Number (%) of Journal Articles Employing Survey Data, By General Sampling Method and Survey Type: 2000-2007

Sampling type	Primary	Secondary	Did not specify	Total
Probability sampling	40 (30)	29 (32)	2 (11)	71 (29)
Nonprobability sampling	10 (8)	2 (2)	-	12 (5)
Other	2 (1)	-	1 (6)	3 (1)
Did not specify	83 (61)	61 (66)	15 (83)	159 (65)
Total	135 (100)	92 (100)	18 (100)	245 (100)

Other includes census

Table 4. Number (%) of Journal Articles Employing Probability Sampling Methods, By Type of Probability Sample Design and Survey Type: 2000-2007

Probability sampling type	Primary	Secondary	Did not specify	Total
Simple random sampling	12 (31)	6 (21)	1 (33)	19 (27)
Stratified sampling	10 (25)	7 (24)	0	17 (24)
Cluster sampling	1 (3)	0 (0)	0	1 (1)
Other Complex Design	3 (8)	6 (21)	0	9 (13)
Did not specify	13 (33)	10 (34)	2 (67)	25 (35)
Total	39 (100)	29 (100)	3 (100)	71 (100)

Table 5. Number (%) of Journal Articles Reporting of Sample Size, By Survey Type: 2000-2007

Type of Survey	Report	No report	Total
Primary	125 (93)	10 (7)	135 (100)
Secondary	66 (72)	26 (28)	92 (100)
Did not specify	14 (78)	4 (22)	18 (100)
Total	205 (84)	40 (16)	245 (100)

Table 6. Number (%) of Journal Articles Reporting How Sample Frame was Constructed, By Survey Type: 2000-2007

Construction of sample frame	Yes	No	Did not specify	Total
Primary	45 (33)	1 (1)	89 (66)	135 (100)
Secondary	23 (25)	3 (3)	66 (71)	92 (100)
Did not specify	1 (6)	-	17 (94)	18 (100)
Total	69 (28)	4 (2)	172 (70)	245 (100)

Table 7. Number (%) of Journal Articles Reporting a Response Rate (RR), By Survey Type: 2000-2007

Type of Survey	Report	No report	Total
Primary	107 (79)	28 (21)	135 (100)
Secondary	52 (57)	40 (43)	92 (100)
Did not specify	11 (61)	7 (39)	18 (100)
Total	170 (69)	75 (31)	245 (100)

Table 8. Number (%) of Journal Articles By Target Population and Survey Type: 2000-2007

Target population	Primary	Secondary	Did not specify	Total
Federal employees	10 (7)	11 (12)	1 (5)	22 (10)
State employees	12 (9)	18 (20)	3 (17)	33 (12)
Local employees	52 (39)	21 (23)	7 (39)	80 (34)
Nonprofit employees	21 (15)	7 (8)	-	28 (11)
Citizens	12 (9)	19 (20)	-	31 (11)
Other	28 (21)	16 (17)	7 (39)	51 (22)
Total	135 (100)	92 (100)	18 (100)	245 (100)

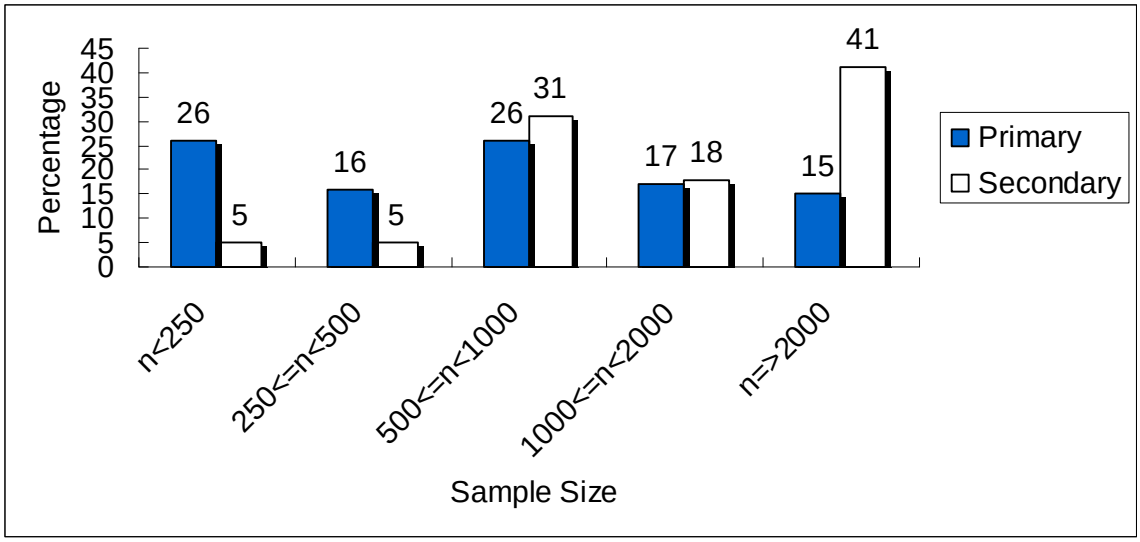
Note: the unit of the target population is employees; percentage in parenthesis
Other: any combinations of sectors, private employees, public (or private) schools or specialists (e.g., police, firefighter etc)

Table 9. Number (%) of Journal Articles Employing Survey Data Analysis Techniques, By Survey Type and Hypothesis testing: 2000-2007

Statistical techniques	Primary			Secondary			Total
	Hypothesis testing		Primary Total	Hypothesis testing		Secondary Total	
	Yes	No		Yes	No		
Descriptive statistics	4 (5)	24 (33)	28 (21)	-	6 (12)	6 (6)	34 (15)
t-test/chi square/ANOVA	6 (10)	11 (14)	17 (13)	4 (9)	5 (10)	9 (10)	26 (12)
Pearson correlation	2 (3)	2 (3)	4 (3)	1 (2)	1 (2)	2 (2)	6 (3)
OLS regression	21 (33)	24 (34)	45 (34)	21 (47)	14 (30)	35 (38)	80 (35)
Multinomial/logistic regression	13 (22)	6 (8)	19 (14)	10 (22)	15 (31)	25 (27)	44 (20)
WLS/GLS/2SLS	2 (3)	1 (1)	3 (2)	2 (5)	-	2 (2)	5 (2)
SEM/Factor/Path analysis	10 (17)	3 (4)	13 (10)	2 (5)	4 (8)	6 (7)	19 (8)
Hierarchical Linear Modeling	1 (2)	-	1 (1)	-	-	-	1 (0)
Time series/Longitudinal analysis	-	1 (1)	1 (1)	2 (5)	2 (4)	4 (4)	5 (2)
Other	3 (3)	1 (1)	4 (2)	2 (5)	1 (2)	3 (3)	7 (3)
Total	62 (100)	73 (100)	135 (100)	44 (100)	48 (100)	92 (100)	227 (100)

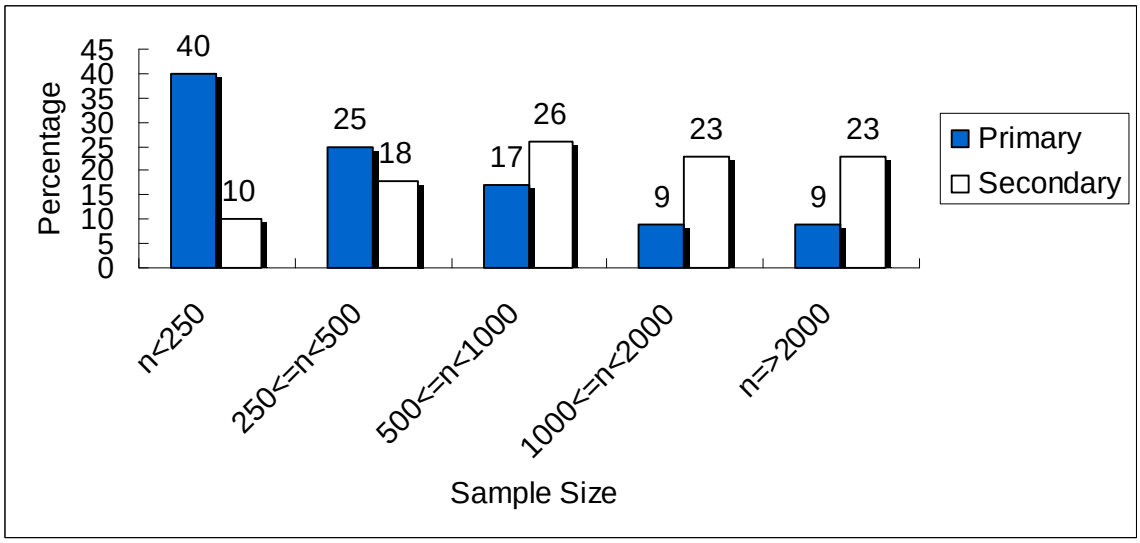
Unspecified cases (n=18) are excluded

Figure 1. Percentage of Journal Articles By Initial Sample Size and Survey Type: 2000-2007



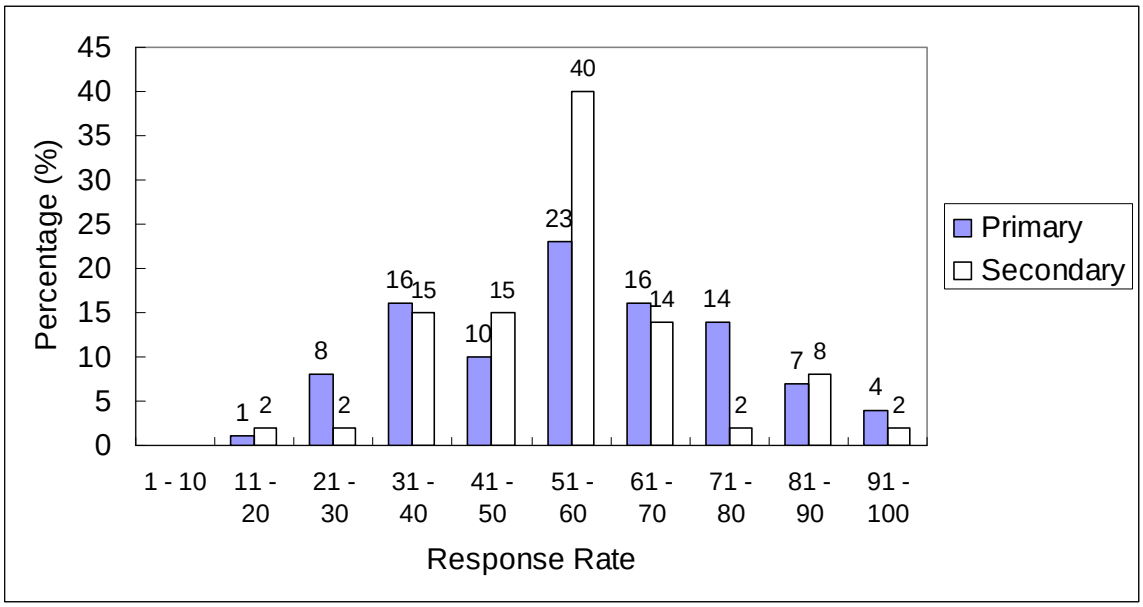
Primary: mean= 1,376, standard deviation = 2,861, median = 551
Secondary: mean = 4,338, standard deviation = 9,332, median = 1,328

Figure 2. Percentage of Journal Articles By Completed Sample Size and Survey Data Type: 2000-2007



Primary: mean = 804, standard deviation = 1,561, median = 296
Secondary: mean = 2,722, standard deviation = 5,610, median = 796

Figure 3. Percentage of Journal Articles with Primary and Secondary Survey Data, By Response Rates: 2000-2007



Primary: mean = 55.9, standard deviation = 18.9, median = 56.3, N = 107
Secondary: mean = 53.4, standard deviation = 15.4, median = 53.4, N = 52

Appendix

Interrater Reliability Tests for Data Abstracted from Sampled Articles (N=24)

Contents	Cohen's kappa (SE)	Confidence Interval
Survey type	0.84 (0.11)	0.63-1.00
Survey design	0.86 (0.13)	0.60-1.00
Data collection mode	0.85 (0.10)	0.65-1.00
Sampling method	0.80 (0.11)	0.59-1.00
Sample size	1.00 (0)	1.00
Response rate	0.90 (0.09)	0.72-1.00
Sample frame	0.83 (0.11)	0.62-1.00
Target population	0.89 (0.07)	0.75-1.00
Use of hypothesis testing	0.88 (0.11)	0.66-1.00
Survey data analysis technique	0.94 (0.06)	0.84-1.00

Note: All coefficients of Cohen's Kappa are significant at the level of 0.001
Confidence Interval: range from 95% lower confidence limit and upper confidence limit
SE: standard error