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

Geon Lee
Survey Research Associate
Doctoral Student in Public Administration
glee29@uic.edu

and

Jennifer Benoit-Bryan
Survey Research Associate
Doctoral Student in Public Administration

and

Timothy P. Johnson
Director of Survey Research Laboratory
Professor in Public Administration

Survey Research Laboratory
Program in Public Administration
University of Illinois at Chicago
6th Floor CUPPA Hall
412. S. Peoria, Suite 615 M/C336
Chicago, IL 60607

This manuscript is in draft form. Please do not cite without written permission from the author

Paper presented at the ST&E Policy Research Symposium, College of Urban Planning and Public Affairs, University of Illinois, Chicago, March 17-18, 2009

ABSTRACT

Survey research is a common tool for collecting peoples' opinions, attitudes, and beliefs for analyses in other social sciences as well as public administration, yet little research has been conducted to examine how well survey techniques are used among researchers in the public administration domain. This article examines four mainstream public administration journals over an 8-year period regarding data collection modes, sampling techniques, sample sizes, response rates, and data analytical techniques in survey research. The findings show that the quality of data generated from survey research in the field of public administration is most commonly threatened by nonresponse bias, coverage errors, and selection bias. 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: Research Methodology, Survey Research, Public Administration Research

Introduction

Since McCurdy and Cleary (1984) first brought up quality issues on doctoral dissertations in Public Administration (PA) in 1984, a number of scholars here discussed the quality of research methodologies in the field of PA (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).

In general, research methodologies in the social sciences can be classified into two categories: quantitative vs. qualitative methods. Typically, the former involves experimental, quasi-experimental, and survey research designs; the latter includes case studies, grounded theory, and action research (Punch 2005). Previous studies have examined some features of quantitative research –measurement practices or analytical techniques – to judge the quality of research (Wright, Manigault, and Black 2004); little has unpacked quantitative methods.

Uniquely, this analysis focuses on a single quantitative method, survey research. The survey research method has been commonly used in the field of PA in order to collect individuals' perceptions, attitudes, and opinions through a small sample of the relevant population. But little research has explored the quality of survey data collection procedures and survey data across these studies. The primary purpose of this investigation is to review the scope and quality of the survey research methods now employed by researchers in public administration.

Previous Reviews of Public Administration Research

McCurdy and Cleary (1984) 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 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 in the field. Both studies expressed serious concerns about the lack of rigorous research methodology in the field. Stallings (1986) argued that research should shift its focus from prescriptive orientations for practitioners to address the intellectual foundations of PA's core questions. He found that the field of PA lacks the analytical standards which are present in the mainstream social sciences such as sociology and political science.

Issues regarding the research quality of doctoral dissertations, first initiated by McCurdy and Cleary (1984), triggered 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 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 journal.

These previous studies focused solely on articles in PAR for their sample. Houston and Delevan (1990) examined a variety of PA journals. The authors argued that research in public administration was underfunded, non-empirical, and not oriented towards theory-testing. They went on to argue that research designs used in published journal articles were monotonous and lacked rigor, and recommended increased utilization of various research designs, including quasi-experiments and experiments. The authors concluded that low quality research in PA might be due to poor research training in PA programs (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. As 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 correct positions within the PA literature. These 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.

Why Survey Research Common and Recognizable?

Survey research is a data collection method in which samples are drawn, interviewed, and analyzed in order to extrapolate to a population of interest. It is widely used in many disciplines such as sociology, political science, public administration, marketing research, public health and communication studies. A survey is an instrument designed to assess respondent's attitudes, perceptions, and opinions on particular social issues, as well as people's factual knowledge (Swidorski 1980). Similar to mainstream social sciences, survey research has been commonly utilized in PA research (Folz 1996). In addition to academic domains, practitioners at all levels of government have increasingly turned to these techniques to obtain citizens' demand and feedback (Daneke and Klobus-Edwards 1979).

With its growing popularity, the demand for quality survey research has been increasing in professional fields as well as academic ones. No large considered merely be a method, survey research is now also recognized as an independent academic discipline[1] in the U.S. Largely an 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.

The quality of survey data is influenced by both sampling and nonsampling errors. Sampling errors can be easily reduced by increasing the size of random samples. However, nonsampling errors are broader in scope and can include inability or unwillingness on the part of the respondents to provide correct information, inability to obtain information about all cases in the sample, and differences in interpretations of question items. Additional errors may take place in recording, coding, processing, or analyzing survey data, involving undercoverage error due to the mismatch between a target population and the sample frame (Groves 1989; Groves et al. 2004).

Nonsampling errors can seriously damage the quality of a survey (Assael and Keon 1982) and undermine both its credibility and impact. Even more importantly, these errors can also lead to false conclusions regarding the phenomenon being investigated (Fisher 1993; Moorman and Podsakoff 1992; Zerbe and Paulhus 1987). Consequently, understanding the direction and magnitude of nonsampling errors is crucial to the accurate application of survey research methods to both applied and theoretical problems.

Data

To investigate the quality of survey research methods currently being applied in the field of public administration, we selected 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] found to utilize either primary survey or secondary survey data. One coder reviewed the total number of articles, coding all the contents in which we were interested here. Then, we selected 24 articles with systematic random sampling, and the second coder coded the articles with same coding schemes as the first coder did. We used Cohen's Kappa tests to examine the consistency between both coders' works for the 24 articles; ultimately this is to judge the reliability of coding works done by the first coder. Table 1 presents that all of the coefficients are found to be over 0.80; attaining the Kappa value of 0.60 or higher has been considered a substantial level of mutual agreement (Landis and Koch 1977). Having tested with only small samples of articles, we calculated confidence intervals to predict the upper and lower levels of Kappa coefficients of a population of the total articles, showing that it is 95 percent confident that the Kappa coefficients of population will lie between 0.59 and 1.00 among coded contents. Accordingly, we claim that all coded works for total articles can be considered substantially reliable.

[Table 1 here]

In analyzing, we differentiated primary survey research and secondary research with survey data in this study, though there were some unidentified articles where the distinction was hard 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 2,135 articles (55 percent of total) with primary survey research were published in the four journals in this eight year period, while 92 articles (or 38 percent) with secondary survey analysis appeared in the four journals. There were 18 articles which did not specify who collected the survey data.

[Table 2 here]

Findings

Figure 1 depicts the number of articles using survey research found in the four PA journals over the eight years. Primary survey research was reported more often than was secondary data analysis in the periods. The number of publications with primary survey research also appears to be increasing gradually over time.

Table 3 shows that 95 percent of the primary studies reviewed were collected using a cross-sectional survey design; five percent collected data at multiple points in time, known as longitudinal surveys. Overall, 87 percent of the survey data reported was cross-sectional and only 13 percent was longitudinal. Over 20 years ago, Perry and Kraemer (1986) found a similar pattern of heavy reliance on cross-sectional data in PA research.

[Figure 1 here]

[Table 3 here]

Data Collection Modes

Table 4 examines the types of main data collection modes employed in PA survey research including mail, web, in-person, telephone, and mixed mode[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), we found that computer-assisted technologies were rarely reported in PA primary survey research.

Surveys are either self-administered or interviewer-administered way. The self-administered mode includes 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) of them were completed via a mail survey, followed by mixed mode surveys (8 percent), telephone interviews (6 percent), the web (4 percent) and in-person interviews (4 percent). In secondary survey research studies, about 33 percent of the total cases used mail surveys; in-person interviews (10 percent), telephone interviews (9 percent), and the web (2 percent). 41 percent of all studies did not specify the method of survey administration, which is much higher than that (8 percent) of primary surveys. Overall, 55 percent of all studies reported that surveys were conducted via mail. Hence, the most utilized survey mode is a mail survey in the field.

[Table 4 here]

In terms of nonsampling survey error, one of the potential problems 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). However, 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 in 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 lass 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 referring to as convenience sampling, does not rely on random selection. Instead, samples are collected based on non-random mechanism which renders 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 5 exhibits the general sampling techniques that were reported in the articles reviewed. About 29 percent (primary: 30 percent, secondary: 32 percent) of all articles reported that surveys were undertaken with probability sampling methods.

Approximately five percent of articles indicated that data were collected using nonprobability methods. It is surprising to find that approximately 65 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: 21 percent of secondary survey studies (see table 6). 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 random sampling without specifying the specific type of sample design.

[Table 5 here]

[Table 6 here]

Sample Size

Table 7 indicates that of the total of 245 journal articles reviewed, 84 percent (or 205 studies) of the total studies reported the sample size[5] of the survey examined. Reporting rates were greater 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 can occur 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 controlled.

[Table 7 here]

Figure 2 depicts the distribution of initial sample size by survey type. Nearly 60 percent of primary survey studies initially set the sample as more than 1,000, whereas 32 percent of secondary data was set as over 1,000. On the other hand, 42 percent of primary surveys reported that the initial sample size was less than 500; ten percent of secondary was as such. While the mean of initial sample sizes of primary surveys is 1,376 (standard deviation: 2,861), that of secondary data is 4,338 (standard deviation: 9,332).

Figure 3 illustrates how distributions of completed sample size differ between the two survey types. Analyses with more than 500 samples were even more made in secondary survey research than in primary survey research: 72 percent vs. 35 percent. On the other hand, 65 percent of primary survey research employed less than 500 samples for analyses. These results suggest that surveys employed in PA research tend to rely on small sample sizes, which may be a result of funding constraints and/or a lack of institutional support (Perry and Kraemer 1986).

[Figure 2 here]

[Figure 3 here]

Response Rates

There are multiple methods for calculating survey response rates. The American Association for Public Opinion Research (AAPOR 2008) provides six ways[6] to calculate 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 sample frame is required.

As shown in table 8, 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 did not specify whether or not they used a sample frame. A typical type of doing mail surveys with no frame would be unspecified mail-out to agency or department; for instance, self-selected persons are free to fill out the questionnaires. In order to accurately compute response rates recommended by AAPOR, it should be known whether or not specified sampled persons are eligible, contacted, or refused. There is no way of knowing such information in self-selection.

[Table 8 here]

Table 9 presents information on response rate reporting. In total, 69 percent (170 studies) of the total studies reported response rates for their surveys: 79 percent for primary and 57 percent for secondary. As shown in figure 4, the distribution of response rate categories is approximately normal in shape for both primary and secondary surveys, which can be demonstrated by the fact that the values of mean and median are quite similar in number: 55.9 vs. 56.3 for primary, and 53.4 vs. 53.4 for secondary. About 50 percent of 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 number of questionnaires sent out and the number returned. In secondary analyses of research, several large-scale data sets (e.g., GSS) reported using the AAPOR definitions for response rate calculations.

[Table 9 here]

[Figure 4 here]

Target Population

A target population is the group to which researchers are interested in generalizing their findings. Table 10 reveals that over 55 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, public managers working in local governments are 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: 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 construct their sample frame and how well it overlaps with the target population.

[Table 10 here]

Techniques Used to Analyze Survey Data

Table 11 illustrates the main statistical techniques that researchers employed in the studies reviewed. 46.7 percent (106 studies) of the total studies investigated was to test formal hypotheses to verify theories or previous findings; 53.3 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-test, 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%) was the dominant method used. Nearly 30 percent used only simple methods including descriptive statistics and bivariate correlation and 15 percent used more advanced techniques –WLS, 2SLS, SEM, HLM, longitudinal 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 11 here]

Discussion and Suggestions

As noted previously, researchers in the field have increasingly used survey techniques, but the use of data collection instruments is not diverse, relying heavily on self-administered instruments, especially mail surveys in the field. The mail survey has its advantages and disadvantages. The advantages of self-administered surveys are many: less costly, relatively easy to conduct, lack in interviewer-related bias, and less susceptible to socially desirable responses (Bridge 1974; Fowler 2002; Kanuk and Berenson 1975). But one of the most concerns of the mail survey is a high level 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 of 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 conduct of more sophisticated survey methods among PA researchers as well as the accomplishment of the quality of research (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).

This is not to say that the mail survey is a problematic data collection strategy. Rather, we regard it as an efficient 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 incentive (Armstrong 1975; Brown and Coney 1977; Goodstadt et al. 1977; Kanuk and Berenson 1975), mixed modes, or the employment of the special design for the questionnaire such as the Tailored Design Method[7] (Dillman

1999) are some recommended approaches to maximize response rates in mail surveys for PA research.

In addition, efforts at reducing measurement errors that may result from complex questionnaires or poor question wording in survey items must be made. In order to reduce these error sources, focus groups, expert review, and pretesting are necessary for the refinement of the questionnaires, yet few studies in our review specified such activities.

The common goal of sample surveys is to accurately estimate a population of interest with small amount of sample. In order to do so, the sample must be representative of the study population. The respresentativeness, 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, about 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 gives rise to coverage problems, obstructing the representativeness of the sample.

Another integral piece of judging representativeness is whether probability random sampling is employed in the survey; only 29 percent used probability random sampling in the sample (see table 5). Approximately 65 percent of surveys reviewed did not specify how respondents were drawn. 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 unbiased. 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. This reflects the recognition that random assignment is an integral element of experimental research as a criterion of research quality. Similarly, whether random sampling is used or not is a critical issue in judging the quality of survey data. Of course, there are multiple types of random sampling commonly employed in survey research.

The probability random sample is a sample technique by which every sampled person has a known probability of selection in the sample frame: this technique is not always used because of difficulties with obtaining rosters of public employees in public institutions. Convenience sampling appears to be efficient when it is impossible to obtain the sample frame (e.g., rare population such as homeless persons or people using illegal drugs). Survey results with convenience sampling, however, are hard to be generalized because of representative problems.

Moreover, survey results with nonrandom self-selection stemming from convenience sampling cannot obtain generalizability of the sample because there is no way of estimating the degree of precision for particular survey variables (Laverakas 1987). At present, AAPOR recommends against the reporting of margin of errors for survey data with self-selected samples, which indicates the survey data with selection bias do not have the capacity to predict the parameters of population. In addition, a fundamental assumption of many statistical models such as OLS regression is that samples are drawn at random (Wooldridge 2002). This reflects that randomness with an adequate sample frame is an integral element for survey data analysis.

It may be wise that the scope of a target population is as small as possible in order to produce quality survey data, which is easier to construct sample frame. The use of secondary data sets of high quality that address key questions of PA also seems advisable for quality research. Lynn, Heinrich, and Hill (2001) suggest such large-scale data sources as the International City Management Association survey of personnel, and the Bureau of Census Survey of Local Government Finances. The Merit Principles Survey (MPS) through the U.S. Merit Protection Board and Federal Human Capital Survey (FHCS) by the U.S. Office of Personnel Management would be a large-scale survey data sources appropriate for studies on public organization behaviors at the federal level.

Survey data are analyzed using quantitative techniques after collection. Numerous studies have identified that empirical research in the field 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 total published studies in our study over an eight year period.

In analyzing survey data, we found some researchers ignored the underlying assumptions of statistical techniques, leading to inaccurate results. The most popular statistical technique using survey data in the studies reviewed was the OLS method (see table 10). There are some critical assumptions which must hold for the OLS technique to be used: independent observation, multicollinearity, and homoskedasicity. These assumptions must be satisfied in order to get valid results using OLS. One assumption of OLS, for instance, is that every observation is independent; however, this assumption may be easily violated when each observation is nested with upper contexts such as agencies, schools, neighborhoods and government units because observations within the same context are more homogeneous (or heterogeneous between contexts) in behaviors and attitudes. In this particular case, instead of OLS, another method such as HLM should be used to address the problem.

For example, Generalized Least Square (GLS), Weighted Least Square (WLS), and 2-Stage Least Square (2SLS) are more complicated variants of OLS; those techniques are usually used if some assumptions of OLS are violated. Few studies detailed a rationale for choosing a statistical technique or revealed a consideration of underlying assumptions before analyzing their data.

What extent of methodological sophistication is sufficient to test PA research questions? Although there are no obvious answers to this question, some scholars suggest promising techniques as statistical tools for PA empirical research. Gill and Meier (2000) provide specific techniques as an analytical tool of PA: time series, maximum likelihood estimation, Bayesian estimation, substantively weighted analytical techniques (SWAT), and generalized linear model (GLM). In addition, DeLorenzo (2000) recommended visualization techniques in which statistics and visualization are combined to visually display the geographic statistics.

We recommend that Hierarchical Linear Modeling (HLM) could also a powerful tool in analyzing survey data in the study of public administration and management (see also Lynn, Heinrich, and Hill 2001; O'Toole 2000). This technique has been widely used in educational research and other social sciences. As we mentioned earlier, clustering effects within the data sets do violate the critical OLS assumption of independence of observation.

The LISREL method, known as Structural Equation Modeling (SEM) is also a sound technique as an alternative to OLS in the measurement of individuals' work attitudes, job satisfaction, work commitment, and public service motivation in public and nonprofit sectors. These concepts are not easy to capture with one survey question because of multidimensionality. In this study, we observed that some researchers used multiple survey items (after testing internal consistency reliability) to construct a latent variable (e.g., work attitude) in the regression model. But in this case, the error variances of the latent variable are neglected, assuming no measurement errors; other researchers employed but a single survey item to measure the latent variable in the model, which is hard to capture multidimensional features of a concept in the measurement. Unlike the traditional regression model, LISREL enables us to estimate error variance parameters, look at a relationship between latent and observed variables (or between latent variables), and deal with complex causal relationships simultaneously involving multiple outcome variables (Byrne 2001; Hoyle 1995). This technique can overcome model simplicity of a unidirectional relationship in PA research. Of course, it must be acknowledged that the appropriate analytic strategies must be guided by the specific research objectives of a given study.

The finale of survey research is to report procedures of survey data collection as well as analytical findings through journal publications, allowing the audience to judge how reliable and accurate the results are. But our finding shows that many aspects of survey research including sampling procedures, the construction of the sample frame, response rates, and sample sizes are not fully documented in PA mainstream journals. This renders the journal audience unable to judge the quality of survey data. Wright and his colleagues (2004) also raised the issue of scarcity of information reporting on method procedures and measures in journal articles. The potential reasons for this scarcity are twofold: (1) the professional journals have no explicit guidelines or policies on the full disclosure of data collection procedures (Johnson and Owens 2003); (2) the authors omit those pieces because reviewers of journals demand condensed papers due to space limitations (Luton 2007; Lynn, Heinrich, and Hill 2001). We are arguing that journal editors and reviewers should not only pay attention to 'how-to-analyze,' but also to 'how-to-collect' in empirical research with survey methods, and insure that the audience has available sufficient information in order to be able to judge the quality of survey data that underlie analytical results.

Conclusions

The evaluation of research is too focused on measurement and analytical techniques, and often assumes that data is perfect in quality. Even in the journal review process, data quality in our experience seems to often be a non-issue. This study highlighted survey data collection methodologies to 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 selection procedures and suspect sample frame quality. This study suggests three take-home messages. First, survey data quality can be improved by individual researchers' efforts at reducing nonsampling errors when conducting research as we suggested. Second, the data quality can be advanced by institutional efforts including funding or research consortiums. Finally, the data quality can be increased by the efforts of journal gatekeepers to strictly review the data collection procedures of research manuscripts, keeping in mind that achieving perfect measurement and analytical results in empirical research will be less than perfect unless the quality of underlying data is reliable.

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; graduate certificate programs in survey methodology are in the George Washington University, Ohio State University, University of Cincinnati, University of Illinois-Chicago, and University of North Carolina-Chapel Hill.

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

3. Mixed mode includes surveys in which two or more different survey modes are used. Typically this mode is employed in order to increase response rates.

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

5. We coded initial and completed sample sizes. The former refers to planned sample sizes before data collection; the latter is one in which unit nonresponse cases are excluded. Some articles indicate both sample sizes, other articles reported either initial or completed sizes. We calculated samples size that was not reported using reported response rates and reported sample size (initial or completed sample). When either initial or completed sample size was reported, we coded it as reporting the sample size. When multiple data sources used, we coded the sample size with the arithmetic average of them for both initial and completed sample sizes.

6. AAPOR provides six ways of calculating response rates: RR1 through RR6. In order to apply the formulas, a survey must be undertaken using probability sampling methods and clearly defined sample frame. The components of the formulas include complete interview, partial interview, refusal and break-off, non-contact, unknown etc. Most journal articles using a mail survey indicated that response rates were calculated as a ratio of the number of returned questionnaires to 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' completion, and partiality. In order to compute response rates for mail surveys, a sample frame containing a list of the target population should exist.

7. The Tailored Design Method (TDM) is a technique for questionnaire designs 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 the mail and web survey.

Many empirical findings show this method improves the understanding of respondents for survey questions as well as survey cooperation in surveys

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.

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.

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.

Byrne, Barbara M. 2001. *Structural Equation Modeling with AMOS*. Mahwah: Sage.

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.

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. 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.

Hoyle, Rick H. 1995. *Structural Equation Modeling*. Thousand Oaks: Sage.

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.

Laverakas, Paul J. 1987. *Telephone Survey Methods*. Newbury Park: SAGE.

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.

O'Toole, Laurence J. 2000. Research on Policy Implementation: Assessment and Prospects. *Journal of Public Administration Research and Theory* 10 (2):263-288.

Perry, James L., and Kenneth L. Kraemer. 1986. Research methodology in the Public Administration Review, 1975-1984. *Public Administration Review* 46:215-226.
Punch, Keith F. 2005. *Introduction to Social Research*. Thousand Oaks, CA: SAGE.
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 reconception. *Academy of Management Review* 12:250-264.

Table 1. Interrater Reliability Tests for Contents of 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; SE: standard error
Confidence Interval: range from 95% lower confidence limit and upper confidence limit

Table 2. Number (%) of Journal Articles Employing Survey Research, By journals and Survey Type

Journal	Primary	Secondary	Did not specify	Row total
ARPA	27	18	4	49 (20)
AS	23	14	1	38 (15)
JPART	30	39	6	75 (31)
PAR	55	21	7	83 (34)
Column total (%)	135 (55)	92 (38)	18 (7)	245 (100)

Percentage in parenthesis
ARPA: American Journal of Public Administration, AS: Administration & Society
JPART: Journal of Public Administration Research and Theory, PAR: Public Administration Review

Table 3. Number (%) of Journal Articles employing Survey Data, By Survey Design and Type

Survey Design	Primary	Secondary	Did not specify	Total
Cross-sectional	128 (95)	79 (86)	14 (78)	211 (87)
Longitudinal	7 (5)	13 (14)	4 (22)	34 (13)
Total (%)	135 (100)	92 (100)	18 (100)	245 (100)

Percentage in parenthesis

Table 4. Number (%) of Journal Articles Employing Survey Data, By Data Collection Mode and Survey Type

	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)

Percentage in parenthesis

Table 5. Number (%) of Journal Articles Employing Survey Data, By General Sampling Method and Survey Type

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 (63)	61 (66)	15 (83)	159 (65)
Total	135 (100)	92 (100)	18 (100)	245 (100)

Percentage in parenthesis
Other includes census

Table 6. Number of Journal Articles Employing Probability Sampling Methods, By Type of Probability Sample Design and Survey Type

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)

Percentage in parenthesis

Table 7. Number (%) of Journal Articles Reporting of Sample Size, By Survey Type

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)

Percentage in parenthesis

Table 8. Number (%) of Journal Articles Reporting How Sample Frame was Constructed, By Survey Type

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)

Percentage in parenthesis

Table 9. Number (%) of Journal Articles Reporting a Response Rate (RR), By Survey Type

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)

Percentage in parenthesis

Table 10.Number (%) of Journal Articles Employing Target Population, By Survey Type

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

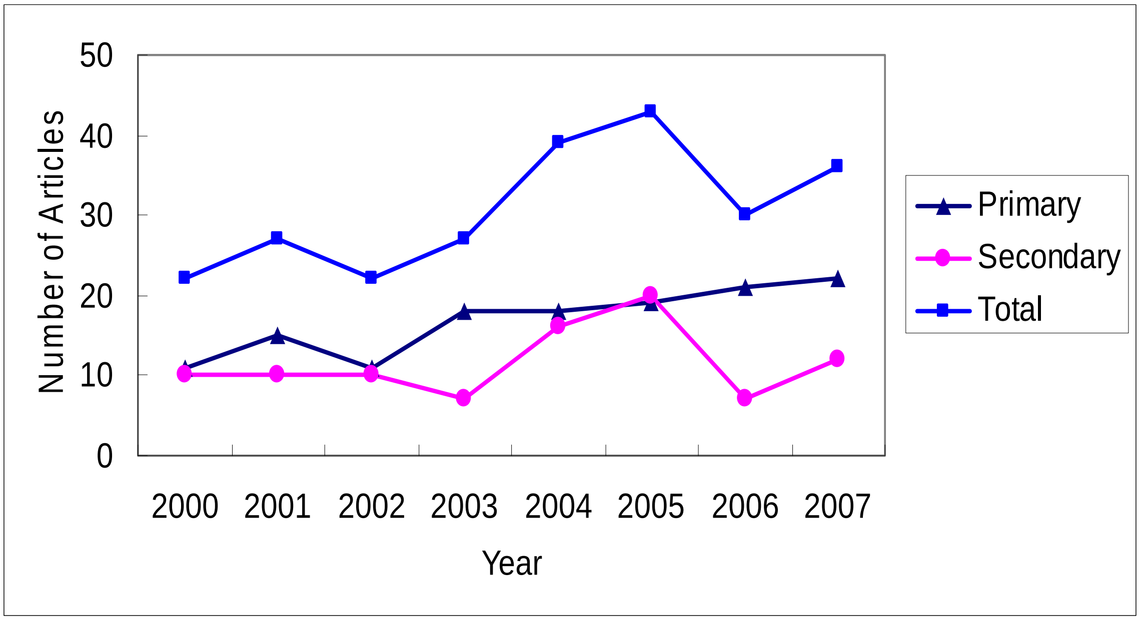
Other: any combinations of sectors, private employees, or specialists (e.g., police, firefighter etc)

Table 11. Number (%) of Journal Articles Employing Survey Data Analysis Techniques, By Survey Type and Hypothesis testing

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)
ttest/chi square/ANOVA	6 (10)	11 (14)	17 (13)	4 (9)	5 (10)	9 (10)	26 (11)
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)	81 (36)
Multinomial/logistic regression	13 (22)	6 (8)	19 (14)	10 (22)	15 (31)	25 (27)	44 (19)
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)	20 (9)
Hierarchical Linear Modeling (HLM)	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)	5 (2)
Total	62 (100)	73 (100)	135 (100)	44 (100)	48 (100)	92 (100)	227 (100)

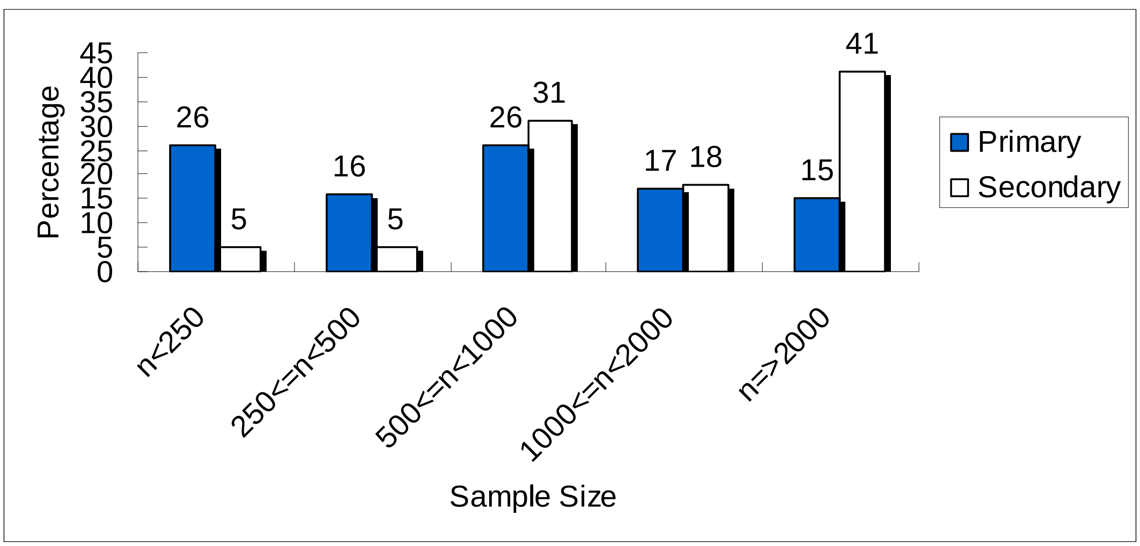
Unspecified cases (n=18) are excluded
Percentage in parenthesis

Figure 1. Number of Articles with Survey Research in Four Journals By Year (N = 245)



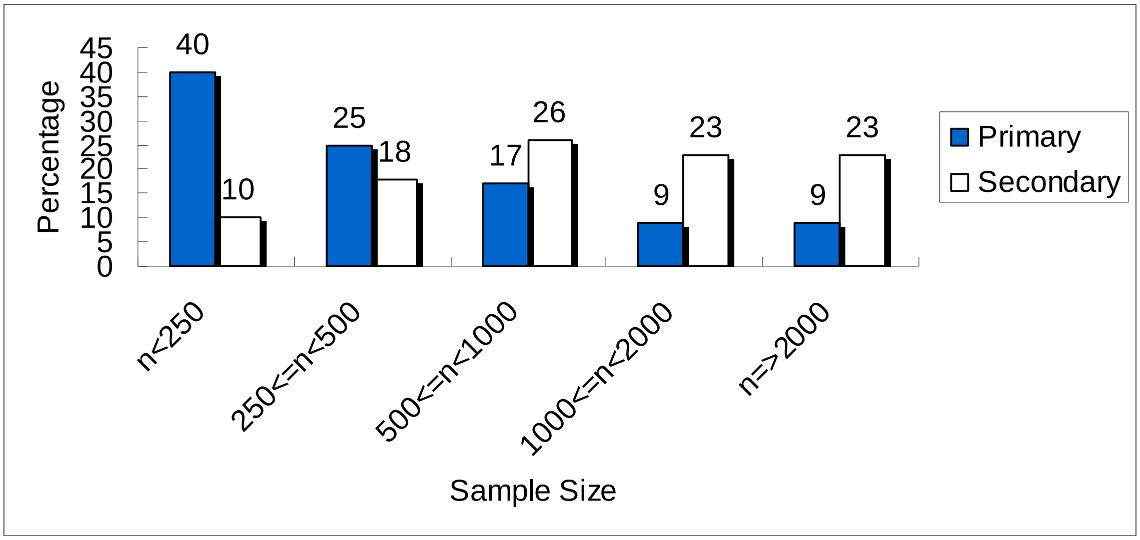
Note: Total includes primary, secondary, and unspecified case

Figure 2. Percentage of Journal Articles By Initial Sample Size and Survey Type



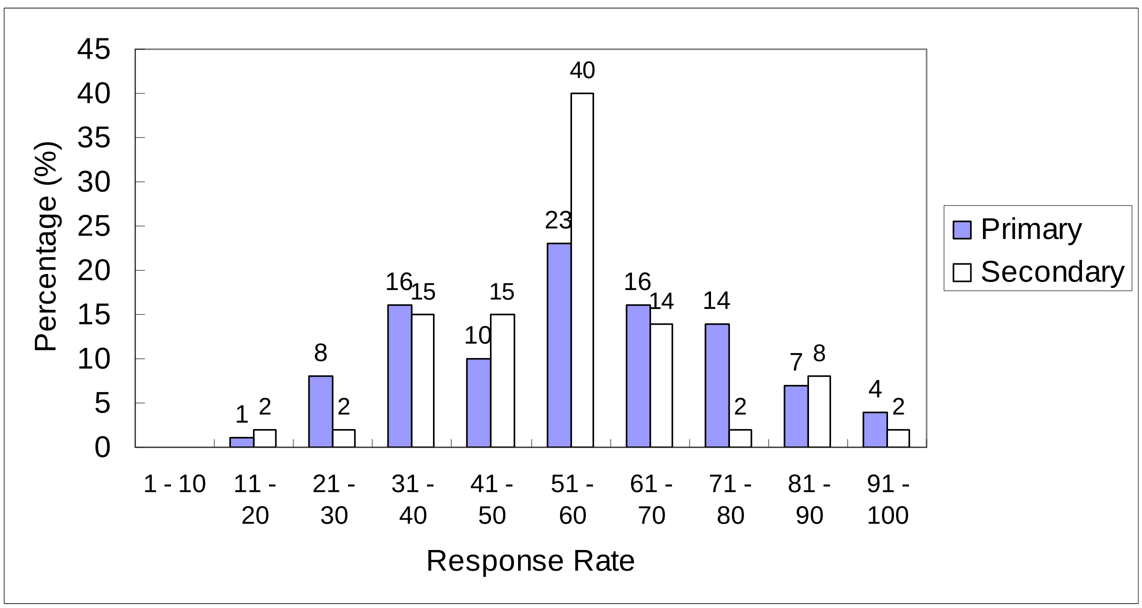
Primary: mean= 1,376, standard deviation = 2,861, median = 551
Secondary: mean = 4,338, standard deviation = 9,332, median = 1,328

Figure 3. Percentage of Journal Articles By Completed Sample Size and Survey Data Type



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

Figure 4. Percentage of Journal Articles with Primary and Secondary Survey Data, By Response Rates



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