

CHAPTER 3

METHODOLOGY

Introduction to Chapter

This chapter describes the plan for answering the research question. It explains why the vote-counting procedures of the meta-analytic methodology were used, and divides the research process into three major steps: 1) conducting an exhaustive search of the sources on the cost effectiveness and service quality of outsourcing; 2) from these sources, extracting and coding the findings across these two measurable variables; and 3) accumulating and summarizing the findings. This is then followed by a conclusion.

Vote Counting Procedure

In order to understand the vote counting methodology used for this dissertation, it is necessary to explain the meta-analytical statistical technique, of which vote counting is a subset. The purpose of meta-analysis as described by Hunter et al (1982, p137) is to either confirm or disconfirm the findings of individual studies through quantitative accumulation and analysis of descriptive statistics across studies. A meta-analysis is a survey in which the results of at least two primary studies are similar enough statistically (i.e., homogeneous) that the results can be combined and analyzed as if they were one study. The idea is that the bigger the sample size studied, the more statistically significant the results. The combined results of numerous research studies make it possible to avoid “problems inherent in individual studies, such as inadequate sample size and problems with statistical power (Davis and Steel, 1988, pp174-177).”

Meta-analyses are used quite often in the medical arena and while some people swear by them, others are not so confident. Professor Hans Eysenck (Greenhalgh, 1997, pp672-675) is not convinced that combining studies done on different populations at different times for different reasons produces viable results. He cites an infamously discredited meta-analysis of many small studies that incorrectly showed that giving intravenous magnesium to people who had had heart attacks was beneficial, that there was cause and effect. A later trial on 58,000 patients found no benefit (Eysenck, 1995, pp 64-74).

The reason this is mentioned here is to point out that this dissertation will not be using this type of meta-analysis. This is because the literature review reveals a lack of sufficient empirical research of outsourcing effectiveness to support such an approach, an inadequacy of which many researchers are aware. Most of the selected individual sources do not meet all of the criteria required for the meta-analytic accumulation of correlations procedures. They often give varying types of inadequate data in incompatible forms that make it difficult to do quantitative computation. Thus, these findings cannot be statistically combined in any meaningful way.

However, despite these shortcomings, this is not to say that the meta-analytic method is not an appropriate approach for integrating and analyzing findings across these sources, because within that method is a procedure called “vote-counting” which will be very useful. According to Hunter, Schmidt, and Jackson vote-counting came about as a

result of the negative aspects of the traditional narrative procedure of integrating results across studies. Specifically, they maintain that

studies will almost never be comparable in design, measures, and so forth, and findings will typically vary across studies in bizarre ways. As a result, the information-processing task becomes too taxing for the human mind. The result is usually one of three outcomes. First, the result may be pedestrian reviewing where verbal synopses of studies are strung out in dizzying lists. That is, the reviewer may not even attempt to integrate findings across studies. Second, the reviewer may simplify the integration task by basing his or her conclusions on only a small subset of the studies. Reviewers often reject all but a few of the studies as deficient in design or analysis, and then advance the one or two acceptable studies as the truth of the matter. This approach unjustifiably wastes much information, and, in addition, may focus on unrepresentative studies. Third, the reviewer may actually attempt the task of mentally integrating findings across all studies – and fail to do an adequate job. (Researchers) have shown that even when the number of studies reviewed is as small as seven, reviewers who use narrative-discursive methods and reviewers who use quantitative methods reach different conclusions (Hunter et al, 1990, p129-130).

As a result of the information-processing burden placed on the reviewer by the traditional narrative procedure of integrating results across studies, the traditional vote counting method was developed. The vote counting procedure according to Dr. Harris Cooper (2003, p5) involves counting up “the positive, negative, and null results, and the pile with the most studies in it is the winner.” The Educational Research Information Center (ERIC) defines vote-counting as “categorizing findings as significantly positive, significantly negative, or nonsignificant. The category with the most entries is considered the best representation of research in the area (Bangert-Drowns, 1991).” Frederic Wolf gives a similar definition (1986, p13) as do Hedges and Olkin (1985, p48) although they also use the phrase “box score methodology” as a synonym for vote counting. Hunter and Schmidt (1982, p130) state, "in its simplest form, the Traditional Voting Method consists merely of a tabulation of significant and non-significant findings." Light and Smith (1984, p74) describe the approach as follows:

All studies which have data on a dependent variable (e.g., cost savings or service quality) and a specific independent variable (e.g., outsourcing) are examined. Three possible outcomes are defined. The relationship between the independent variable and the dependent variable is either significantly positive, significantly negative, or there is no significant relationship in either direction. The number of studies falling into each of these three categories is then simply tallied. If a plurality of studies falls into any of these three categories, with fewer falling into the other two, the model category is declared the winner. This modal categorization is then assumed to give the best estimate of the direction of the true relationship between the independent and the dependent variable.

Besides a categorization, qualitative information and case narratives can provide a richness of description difficult to capture in a more coherent quantitative summary. Further, the vote-counting procedure was also chosen because of feasibility. It would cost too much money and take too much time to gather primary data large enough to generate significant results that would help to answer the research question. It seems more productive to try to make sense out of existing data than to add even more to the pile (Hunter and Schmidt, 1990, p497). Like meta-analysis, the results of different, localized, individual research will be aggregated and analyzed to determine the commonality of findings and to give a macro sense of whether or not outsourcing has in terms of saving money and/or improving service quality, been effective, ineffective, or had no discernible impact. The results of the vote count group of sources are shown in Table 4.1.

Vote counting in general is subject to three primary limitations. First, because journals are prone to publish reports that show statistically significant findings, this results in a dearth of reports showing no significant findings. However, that is not a problem for this dissertation because sources showing outsourcing effectiveness (or lack thereof) are both given equal attention in the published literature, regardless of how

significant the findings. The second area of concern has to do with the quality of the sources, that is, that well done reports will be included with poorly done reports. To minimize this problem, sources were eliminated that were lacking in at least a modicum of statistical rigor. As much as possible, only documents that presented methodologically defensible positions in peer reviewed sources were used. No documents that simply stated the author's opinion without any supporting data were used. A third limitation of some sources is the transferability of their results. Some focus on the impact of outsourcing of a particular service at a particular level of government, and it is possible that the findings from this situation may not apply to other functions, or at a different governmental level (Kozlowski et al., 1993, p3). However, this "mixing of apples and oranges," or reviewing dissimilar sources is not a problem here since the goal is simply to paint a broad picture of a research literature and from it to provide an objective general assessment of the effectiveness of outsourcing. In order to avoid the "mixing" problem as much as possible, sources will be analyzed in groups by level of government, and type of service.

The vote-counting procedures of the meta-analytic methodology divide the research process into three major steps: 1) conducting an exhaustive search of sources about the effectiveness of outsourcing; 2) from these sources, extracting and coding the findings across measurable variables such as savings and/or service quality; and 3) accumulating and summarizing the findings.

Collection of Sources

The initial step in the vote counting procedure involved identifying relevant outsourcing studies.” The initial expectation was that there would be hundreds of “studies” from which to choose, and that instead of studying outsourcing of many services at all levels of government, there would be so much data that this vote counting analysis would have to be limited to just a particular level of government or a few select services. Those concerns were completely unfounded as the outsourcing literature did not prove to be the gold mine of “studies” it was hoped to be.

The limited supply of useful “studies” was caused by the caveats previously mentioned (i.e., no international or private sector studies; no discussion of health care, social security, selling of assets, vouchers, grants/subsidies, volunteerism, donations, user fees, underground economy, deregulation, or tax reductions), as well as four additional self-imposed ideal criteria.

The first ideal criterion was that sources would be unbiased. Unfortunately, the concept of outsourcing carries tremendous political baggage with it, as discussed in Chapter 2. As a result, most of the sources found were less than objective. As Pratt and Maahs (1999, p1) state, “some researchers have appeared more interested in supporting their ideological position and consequently have conducted “simple cost comparisons” that lack analytical and methodological rigor.”

The second ideal criterion restricting the number of useful sources was that they come from vetted, peer-reviewed journals or otherwise reliable documents. This was a quality control issue, meant to filter out “opinion pieces” that had little in the way of reliable data. In a prescient e-mail conversation with E. S. Savas, a founding father of privatization, he indicated that finding these types of sources was going to be hard to do. According to Savas, “I’m not aware of any studies (that) are vetted by ‘panels of experts’ in this or and rarely in any other field. These are not studies to determine the efficacy of cancer cures.” Savas was correct in that while there is plenty of verbiage available about outsourcing, most of it not suitable for a vote counting analysis such as this.

Despite the dampening effect that this criterion in particular placed on the amount of literature that was ultimately useful, it was absolutely necessary. The alternative was to count every newspaper and magazine article, every press release from various non-objective sources, and every editorial and opinion piece that made claims about the effectiveness of outsourcing, albeit unsubstantiated. Not only would this have been a huge undertaking, it would have been meaningless, because in the end, whether the results proved or disproved the effectiveness of outsourcing, such a vote counting analysis could be readily dismissed as just being the result of the winners having written more about the subject. While this is a fundamental flaw with any vote counting analysis, it is not fatal. The way to correct for it is to ensure that the material used in the vote counting analysis contains some methodologically defensible rigor and comes from as many objective sources as possible.

A third ideal criterion restricting the number of useable sources was that they had to be from the original author, that is, primary data. In other words, one author commenting on another does not a study make. This circular nature of one study citing another study became especially obvious while reading the prison outsourcing literature. What at first seemed to be new material, was often just rehashes of a few other prominent studies, twisted to fit the new author's perspective. Further, many authors were fond of quoting themselves, thus being able to regurgitate old material which to the uninitiated seemed like new data.

A fourth and final ideal criterion was that the research documents used had to have been conducted after 1981 because this is when the United States, starting with the Reagan Administration, began to seriously undertake outsourcing.

How Documents Were Selected

The exhaustive literature search for relevant studies began by getting on the Internet and typing in the key words "privatization," "outsourcing," and "contracting out." Internet search engines (e.g., AltaVista, Google, Yahoo) were heavily utilized as was the website Ask Jeeves. Advanced searches attaching the words "effectiveness studies" were also conducted. This search also included Virginia Tech's electronic library; the University of Alabama (Huntsville) library; the City of Huntsville public library; book purchases; and e-mail and telephone contact with authors of various studies.

In addition to searching electronic dissertations at Virginia Tech, the University of Alabama (Huntsville), and Proquest, and using the web search engines to look for items containing the general synonyms of privatization, the following list gives examples of specific web sites that were contacted and searched individually for documents:

Academy of Management Executive
Academy of Management Journal
Academy of Management Review
Administrative Science Quarterly
Agriculture and Resource Economic Review
Alliance for Redesigning Government
American Review of Public Administration
Buckeye Institute
Business Executives for National Security
Carl Vinson Institute of Government (University of Georgia)
Center for Defense Information
Center for Naval Analysis Corporation (CNAC)
Council of State Governments
Department of Housing and Urban Development (HUD)
Econometrica
Employment Relations Today
EnterpriseEconomy.com
General Accounting Office (GAO)
General Services Administration (GSA)
Government Executive
Grandfather Economic Report Series
Heritage Foundation
Housing Policy Debate
Human Resource Development Quarterly
Human Resource Focus
Human Resource Management
Human Resource Management Journal
Institute for Defense Analysis (IDA)
Journal of Agriculture and Applied Economics
Journal of Agriculture and Resource Economics
Journal of Economic Behavior and Organization
Journal of Food Distribution Research
Journal of Government Information
LexisNexis “privatization” query
LexisNexis “outsourcing” query
Mackinac Center for Public Policy
Management Decision
National Tax Journal

North Eastern Journal of Agriculture and Resource Economics
OCLC FirstSearch of EconLit
OCLC FirstSearch Alternative Press
OCLC FirstSearch Article
OCLC FirstSearch Contemporary Women's Issues
OCLC FirstSearch PAIS
Office of Management and Budget (OMB)
Organizations Dynamics
Performance Improvement Quarterly
Public Administration Quarterly
Public Administration Review
Public Budgeting and Finance
Public Citizen
Public Manager
Public Personnel Management
Public Welfare
Rand Corporation
Research in Urban Policy
Review of Public Personnel Administration
Social Science Quarterly
Southern Economic Journal
Southern Journal of Agricultural Economics
Welfare Information Network
Western Journal of Agricultural Economics
Western Political Quarterly

As the time-consuming, and morale-destroying research for this dissertation got underway, it became quickly and painfully obvious that true “studies” about the effectiveness of outsourcing that met all of the caveats and criteria listed above, are rare. In order to overcome the self-imposed limiting criteria and caveats mentioned above, and in order to gather a reasonable number of sources with which to conduct a vote counting analysis, a looser definition of the word “study” had to be adopted.

To be perfectly clear, unfortunately some of the sources that became part of this vote counting analysis are not really “studies” per se. For example, GAO reports are not

technically studies in the traditional academic sense where a research question is asked, research is conducted and analyzed, conclusions are reached, references are cited, the study is vetted with a panel of experts, and then the article is finally published in an academic journal. A GAO report typically only references other GAO materials, is not peer reviewed, and isn't published in any academic journal. Another example is testimony before legislative committees which technically does not rise to the level of being a "study" but is still useful for this vote counting analysis.

Lowering the "study" standard resulted in the finding of additional sources that, while not meeting all of the ideal criteria listed above, were still relevant as defined by Hunter and Schmidt (1990, p497) because they focused on the relationship of interest and:

1. Presented data in measurable terms that demonstrated either savings or additional costs, and/or discussed service quality as a result of outsourcing;
2. Gave a conclusion one way or the other about the success or failure of the outsourcing effort being written about;
3. Involved an American government organization that had undergone outsourcing (no private sector, no foreign governments); and

4. Used unique data. For example, to include two or more studies that demonstrated that savings had resulted from outsourcing the investigative service office of the Office of Personnel Management would be counting at least one vote too many in favor of outsourcing. Only one vote per outsourcing attempt was counted.

The sources used for this vote counting analysis also capture invaluable information on outsourcing effectiveness through their qualitative descriptions. On the other hand, all of the sources had some data inadequacy that rendered them unusable for the accumulation of correlations procedures for meta-analysis such as not giving all the data required (e.g., sample size, reliabilities, and correlations).

This search of databases led in all sorts of directions, sometimes resulting in hundreds of possible “hits” that had to be whittled down to be workable. This was done by going through the following three step process.

First, by looking at the source of a document, one could tell whether it was just an opinion piece or article from a newspaper or magazine as opposed to being a report from a government organization, university, commission, academic journal, or testimony before legislative bodies. The former items were ignored while the latter types reached the second level of review, namely an appraisal of the title. Here some judgment had to be used. If a title read, “Former Soviet States Sell Off Assets in Bid to Privatize Economy,” it was rejected before even being read, both because it was not about the

United States and secondly, because it apparently dealt with the sale of state owned enterprises, a subject not of interest to this dissertation (see page 33 of this dissertation). On the other hand, a title such as “Outsourcing TA-4J Maintenance: Cost and Quality Experience” would be intriguing enough to make it to the third level of review. This level consisted of actually retrieving the document to determine if it contained at a minimum, the answer to the yes/no question of whether or not the outsourcing attempt resulted in lower costs and/or better service quality.

After this winnowing process, 99% of the documents reviewed at this level were still not useable, as evidenced by the fact that ultimately only forty research documents from thirty different sources were found that could be included in the vote-counting analysis. Some potential sources that made it through the three step review process, turned out to be press releases or articles from private organizations with a keen, but biased, interest in outsourcing, and therefore had to be viewed with caution. Despite quite a volume of material that was available from these groups (e.g., the Reason Public Policy Institute (RPPI), the Goldwater Institute, the Heritage Foundation), and despite the fact that at first glance the material appeared useful for this vote-counting analysis, and despite their protestations that they were independent observers of the outsourcing phenomena, only one report from each such organization was used. For example, it would have been easy to include twenty five reports from RPPI, but that would have skewed the vote counting to look very positive. RPPI contends in their literature that they are a “non-partisan public policy think tank promoting choice, competition, and a dynamic market economy as the foundation for human dignity and progress. RPPI

produces rigorous, peer-reviewed research.” In an e-mail conversation with Ed Balaker of RPPI, he maintains that all of their reports are reviewed by one outsourcing proponent, one outsourcing opponent, and two neutral observers before being published. However, not one RPPI study was found that was anti-outsourcing, thus bringing into question their neutrality on the issue. The reverse would have happened if twenty five reports were used from Public Citizen, Ralph Nader’s anti-outsourcing organization. Thus, to counteract for potential bias or for questions of methodologies that are not as easily defended, of the forty sources that were ultimately included in this vote counting analysis, there is only one per any organization, except for ten from the General Accounting Office which is very explicit about its methodology.

The rejected articles were did contain references, which were then put through the same three step winnowing process just detailed. However, these references also had their problems. Oftentimes they tended to come full-circle, leading right back to the seemingly most prolific writer on the subject of outsourcing, namely RPPI. RPPI was quoted over and over again without as much concern for methodological critiques as is typical in more academic sources.

**Extracting and Coding the Findings Across Comparable, Measurable Variables
(Cost Savings and/or Service Quality)**

Once a source was selected for meeting the above criteria, it was evaluated against the dependent variables, cost savings and service quality. Each source had to

define the independent variable, outsourcing, the same way in order for the measurement scale to be consistent across studies (see page 33 of this dissertation, Narrowing the Focus of this Discussion). This step controlled the possibility of misinterpretation, caused by variables disguised under another name or aggregated with secondary meanings.

The outsourcing literature presents several common dependent variables that are useful in answering the research question. “Common” is the key word here, because while some studies may have evaluated a variety of dependent variables (e.g., dollars saved, changes in service quality, evidence of cherry picking, evidence of corruption, changes in administrative costs, number of jobs losses) there were only two variables that were consistently assessed (not always, but most of the time) throughout all of the studies and they were “savings” and “service quality.” This finding is consistent with surveys by Savas (2000, p119) and Chi (1997, p1) which demonstrate that “saving money” is the main reason given by government agencies for considering outsourcing in the first place, and thus the main variable measured. While improving service quality was well down the list of reasons governments outsource, it, along with savings, were the only two variables that reasonably could be assumed to have been measured in a similar fashion (e.g., before and after measures of costs and service quality) between source documents.

It should be noted that simply saving money does not necessarily mean that outsourcing has “worked” or “been effective” because this result could have come at the expense of service quality (as shown in many instances in this vote counting analysis).

Conversely, neither does improved quality of service mean outsourcing has “worked” or “been effective” because this result could have come with additional dollar costs, as shown in a few instances in this vote counting analysis. But dollars saved and service quality improvements are two measurements that do not involve any value judgments by the researcher. While using this reduced set of measurements may not be ideal, it does conform with the purpose of this dissertation which again is to paint a methodologically defensible, broad picture about the overall effectiveness of outsourcing.

In order to be included in this dissertation, each source had to address at least dollar savings or service quality, and preferably both as dependent variables. While some sources did not explicitly define what they meant by “savings” and “service quality,” in all cases where they were addressed, it was easily discernible whether savings had been achieved or service quality had improved. Since this dissertation is only interested in the yes or no question of whether money was saved and service quality improved, issues about what time period was covered by the study, or how much was saved, or how much service quality improved, became irrelevant. If funds were saved, whether it was a million dollars or one dollar, it was recorded as a positive event. If service quality improved, even marginally, it was recorded as a positive event. If there was no change, it was recorded as a negative event, since the outsourcing had not achieved its intended result. If an effort ended up costing more or service quality suffered, it was also recorded as a negative event. The selected sources are summarized in Table 4-1 displaying whether the results were positive, negative, neutral, or were not available.

Data was tabulated by establishing a working Excel spreadsheet table (see Table 4.1), referred to as the Order List. In order to remain in an easily printable and viewable format, this spreadsheet, while providing substantial amounts of data, has been necessarily truncated. Column A provides a number which corresponds to a particular source document. Column B denotes whether the government service being rendered is at the federal, state or local (county, municipality) level. Column C provides the date of the study, and Column D lists the government agency involved (e.g., City of Phoenix) while Column E provides the department providing the service (e.g., fire department).

The next 26 columns (referred to collectively as Column F) are headed with a single letter of the alphabet. Each letter corresponds to a particular grouping of services that were determined to be similar enough so as to be able to make valid comparisons. The one exception was the letter “O” (other) which is a catchall for services that didn’t fit neatly in the other 25 columns. The number of entries in this column was purposely kept to a minimum.

The data in the twenty six columns of synonymous services shown below is more easily manipulated than if the data had been presented in 222 columns, one for each outsourced service in this vote-counting study. While each of the 222 services may be unique, each one wasn’t substantially different enough to warrant its own separate distinction in its own column. On the other hand, neither could the data be reasonably combined in four or five columns without greatly increasing the chances of mixing apples

and oranges and losing the richness of the data. Twenty six columns was a reasonable compromise between too many and too few categories

A = Administrative tasks (bookkeeping, forms, payroll, postal/mail duties)

B = Bookstores

C = Computers (information technology)

D = Data Entry (transcribing)

E = Employment services (welfare services, social services)

F = Food service (nutrition, dining services)

G = Groundskeeping (landscaping)

H = Hounding (bill collecting, child support payment enforcement)

I = Infrastructure (streets, irrigation pipes, housing)

J = Janitorial (custodial, housekeeping)

K = Keeping children (child care)

L = Laundry (dry cleaning)

M = Maintenance

N = Nurturance (base operating support)

O = Other

P = Prisons (corrections)

Q = Quality assurance (precision measurement)

R = Refuse (trash collection, landfill management)

S = Security (firefighting, police, guards)

T = Transportation (buses)

U = Utilities (water, waste water, telecommunications)

V = Vehicles (maintenance of, rentals, motor pools, inspections)

W = Warehousing (supply)

X = Xeroxing (copying, printing, duplication)

Y = Youth Services

Z = Zoo keeping (veterinary services, animal caretaking)

Within the collective Column F, a “Y(es)” denotes that savings were achieved by the outsourcing initiative, while an “N(o)” denotes an increase in costs. An “N/C” denotes no change in costs, and an “N/A” means it was not assessed by the source document. Column G records whether the quality of the service provided by the outsourcing initiative improved (+), declined (-), did not change (N/C) or was not assessed (N/A). Column H provides a few notes about each initiative. And finally, column I is a simple assignment of a number to each service outsourced. This simple numbering tool allows one to determine how many government services were analyzed, a completely different number than that provided by looking at the number of studies reviewed.

Once the data was tabulated, a brief description of each study is given in Chapter 4, as well as an overall assessment as to whether or not the study’s findings were positive about the effectiveness of outsourcing, negative, neutral, or unknowable.

Accumulating and Summarizing the Findings

The next step in the vote-counting procedure involves analyzing the inputs and sorting them to arrive at an overall assessment of the effectiveness of outsourcing. The data was sorted along Column B (level of government), and the votes were counted for cost savings and service quality changes to determine if successful outsourcing was limited to a particular level of government or whether it was pervasive throughout all levels. Next the data was sorted by type of service and the positive and negative votes for cost savings and service quality changes were counted to determine if some services were more prone to successful outsourcing than others.

Conclusion

The vote counting procedures and the appropriateness of this methodology for combining the analyses from independent sources to answer the research question have been described. Now it is necessary to provide a brief description of each of the source documents used for the vote counting analysis so that future researchers may verify and replicate the findings of this study.