

Evaluation of Prospective Benefits, Costs and Risks of Standards for Statistical Programs December 2007

John L. Eltinge

Office of Survey Methods Research, U.S. Bureau of Labor Statistics
PSB 1950, OSMR, BLS, 2 Massachusetts Avenue NE, Washington, DC 20212

Eltinge.John@bls.gov

Abstract

Standards have an important role in the development, adoption and diffusion of many types of technology. This paper considers statistical survey methodology as a form of technology, and uses the resulting conceptual framework to explore several ways in which to evaluate the prospective costs and benefits of standards for statistical programs. The conceptual framework places primary emphasis on types of standards; methods for calibration; methods for application and enforcement; and special issues for government-sponsored statistical programs. This framework leads to discussion of prospective benefits of standards, including improved data quality; reduction of quantifiable costs for stakeholders; and reduction of risks for stakeholders. In parallel with benefits, the paper also reviews potential costs and risks associated with survey standards, including direct costs, indirect costs and inefficient allocation of resources. The paper closes with comments on practical implications of these general ideas for development, implementation and enforcement of standards; mechanisms for enforcement; training and management of statistical program personnel; and communication with external stakeholders.

Key Words: Adoption and diffusion of technology; Approximation error; Barriers to entry; Certification; Commodity labor; Constraints; Gresham's Law; Local and global optimization; Operational risk; Rent seeking; Risk management; Satisficing; Survey cost structures; Total survey error; Training; Transparency.

1. Introduction

I would like to thank the organizers for the opportunity to speak today in this session on "The Role of Statistical Standards in Federal Surveys." Standards for statisticians and statistical organizations have received considerable attention in the literature for many years. For example, the reference list at the end of this paper includes over fifty papers on various aspects of standards, and the statistical community continues to have very active interest in this topic today.

The previous statistical literature on standards covers a wide range of issues, but has focused primarily in two areas. First, our profession benefits from a very substantial body of work related to statistical ethics and related standards of personal and institutional conduct. For example, *Ethical Guidelines for Statistical Practice* (American Statistical Association, 1999) presents guidelines that apply broadly across the entire statistical profession. More specifically for government statistical organizations, *Principles and Practices for a Federal Statistical Agency, Third Edition* (National Research Council, 2005) provides very valuable documentation of principles and practices that are essential to quality and integrity of work in the highly decentralized federal statistical system.

Second, the sample survey community has produced a large body of technical standards for specific statistical processes and products. Some prominent examples are the relatively early work by Hansen (1952); subsequent publications by Gonzalez et al. (1975) and Frankel (1975); and recent work by AAOPR on nonresponse standards (Smith, 1999), by the International Standards Organization (2006) on market, opinion and social research, and by the Federal Committee on Statistical Methodology and OMB on federal surveys (United States Office of Management and Budget, 2006).

Government statistical agencies in other countries have developed related sets of standards. See, e.g., Australian Bureau of Statistics (2007), Statistics Canada (2003), Statistics Finland (2006), Statistics New Zealand (2007), United Kingdom Office of National Statistics (2002), and references cited therein.

All of the abovementioned work is very important, but persistent anecdotal evidence indicates that the benefits of these efforts can be attenuated by issues related to resources, stakeholder expectations and managerial constraints. In an attempt to address some of these concerns, this paper will present a general framework for the development, implementation and management of standards for statistical programs, with emphasis on four areas. Section 2 suggests evaluation of survey methodology as a form of technology, and reviews some related ideas from the literature on the adoption and diffusion of innovations in technology. Section 3 discusses standards as tools for the efficient management of institutional investments in statistical technology, with special emphasis on the importance of clarity regarding

prospective benefits, costs and risks. This leads to a conceptual framework for standards in statistical programs, including the targets for standards; calibration of standards; mechanisms for application and enforcement of standards; and special issues encountered with statistical standards and government-sponsored surveys. Section 4 explores the prospective benefits of statistical standards. Section 5 considers a corresponding set of costs and risks. Finally, Section 6 reviews the main ideas in this paper and highlights the importance of transparency; technical training and qualifications of personnel; and integration of standards with institutional values and incentive structures.

2. Statistical Survey Methodology as a Form of Technology

The statistical community works at the fascinating interface between first-principles science and pragmatic technology. Several authors have noted the natural – and I believe largely creative – tensions that arise at this interface. See, for example, Mahalanobis (1965, citing earlier comments by R.A. Fisher) and Healy (1978). These creative tensions in turn can inform a wide range of approaches to standards for statistical programs. The current paper will view statistical methodology primarily as a form of technology, which we will define as the application of fundamental scientific results (largely mathematics, the behavioral sciences and relevant substantive areas like economics or medicine) to address the specific needs of specific groups of stakeholders, based on balanced consideration of tangible benefits, costs and risks. Moreover, any serious discussion of technology management (including standards for said technology) will involve nontrivial trade-offs among the abovementioned benefits, costs and risks.

In addition, it is worthwhile to consider statistical methodology within a framework defined by the general literature on adoption and diffusion of innovations in technology. See, e.g., Rogers (1995), Katz (2004) and references cited therein. For this current discussion of standards, two points are of special interest. First, the term “standard” implies a certain degree of standardization of the product or process of interest. Consequently, a specific technology (including specific forms of survey methodology) needs to reach a certain level of maturity and stability before one can reasonably develop meaningful standards. Second, carefully calibrated development and implementation of standards can accelerate the adoption and diffusion of a given technology. This is especially true for standards that are transparent and are carefully linked with the articulated needs of users. For example, standards for the quality, integrity and objectivity of published survey data, and for interpretation of those data, can help to enhance public

confidence in results reported by the federal statistical agencies. Similarly, standards for the quality of work with sample design, fieldwork, edit and imputation procedures, estimation systems, personnel training, and cost-accounting systems can help to assure non-specialists that statistical organizations are using available resources as efficiently as possible in production of these data.

Conversely, the lack of standards, or the presence of inappropriate standards, can seriously impede the broad use of a technology by non-specialists, and can lead to corresponding weaknesses in the market for that technology and for further investments in improvement of that technology. This latter point may be especially important for statistical survey programs, which have somewhat unusual market features, e.g., the fact that many of our products are ultimately handled as public goods; and the presence of complex funding streams, production processes and dissemination paths that involve a mixture of statistical agencies, other government agencies, universities, and the private sector.

3. Conceptual Framework: Standards as Tools for the Management of Investments in Statistical Technology

A large proportion of work with statistical methodology, including survey methodology, may be best characterized as a form of capital investment, and thus should be managed accordingly. For example, investments in survey methodology will differ from current expenditures on survey production processes in the time elapsed between the investment and the production of tangible benefits for stakeholders; expected milestones and management decision points; hurdle rates and risk profiles; and the leverage provided by highly specialized skill sets that often require very long learning curves.

If taken seriously, survey standards amount to an institutional commitment regarding investments in statistical technology, e.g., the development and implementation of a specific type of new methodology; or the training of personnel in a given methodological area. In this sense, statistical agencies should view standards as a fundamental tool in the management of our investments in survey methodology. Consequently, the efficient development, implementation and management of standards require solid consensus among the primary stakeholders regarding the framework for prospective standards; the prospective benefits of said investments; and the concomitant costs and risks. The remainder of this section provides a general conceptual framework for evaluation of these standards. Sections 4 and 5 use this framework to explore the resulting benefits, costs and risks.

3.1 Features of a Survey Program That May Warrant Standards

3.1.1 Standards for Statistical Products

In principle, one could consider development of standards for many different features of a survey program. Of greatest interest to some data users would be standards for specific *statistical products* disseminated by the survey organization. For example, many federal statistical agencies use the p-percent rule or disclosure-limitation criteria to determine which cells of a contingency table may be published. In addition, some agencies require that published estimates satisfy pre-specified criteria for response rates, relative standard errors or other indicators of data quality. In some cases, agencies also require that published survey reports incorporate information that can assist a data user in understanding the published data. Examples include documents covering definitions, collection instruments and the methodology used to produce the published estimates; tables of standard errors; and other numerical measures of data quality.

In considering these standards, two issues are of special interest. First, it is useful to distinguish between standards for a final product (e.g., relative standard errors for a specific reported index estimate) and standards for an intermediate product (e.g., response rates for the individual data elements that contribute to that index estimate). Second, one should consider standards for statistical products within the broader context defined by the information needs of data users and by the incentive structures for data providers. For example, the “number of published estimates” is one of several factors used to determine funding levels for some statistical programs. In such cases, the program management has a natural incentive to increase the number of published lines; and acceptance and implementation of a proposed standard for data quality is more likely if one integrates the standard with the underlying incentive structures, e.g., by allocating budgets on the basis of the number of published estimates *that meet a specific objective criterion* like maximum relative standard error.

3.1.2 Standards for the Properties of Statistical Processes, and for the Objective, Reproducible and Transparent Evaluation Thereof

Most methodological research and evaluation work tends to focus on statistical *processes* and the properties of these processes. Examples include sample design, fieldwork protocols, estimation methods, and dissemination methods. For each of these components of a statistical survey procedure, one could consider a corresponding set of *process standards* for, e.g., selection

of sample units (including probability-based sample reductions when necessary); rigorous development and testing of instruments; efficient use of sample data and available auxiliary information; procedures to reduce respondent burden; contact rules for sample units; the number and form of callbacks for nonresponding units; internal and external quality checks; and the approximate unbiasedness of estimators.

In parallel with these process standards, one generally expects to see standards for *objective, reproducible and transparent evaluation and reporting* of the properties of statistical programs. For example, an OMB package for a survey program should include (or provide appropriate references to) clear, detailed and explicit descriptions of the proposed sample design; formulas for the principal point estimation and inference methods used; and documentation of previous empirical results that support claims regarding anticipated response rates, standard errors and other quality measures.

3.1.3 Cost Structures for Statistical Programs

In practical survey operations, performance – and compliance with applicable standards – depends both on methodology and on the efficient allocation of a wide variety of resources. Efficient allocation in turn depends very heavily on available information on fixed and variable costs at a relatively fine level of aggregation. Some simple fieldwork examples include the relative costs of: listing and mapping of households in a sample area; collection of data through the internet, mail, telephone interview or personal-visit interview; an initiation interview and a re-interview in a panel survey; and initial contact attempts, re-contact attempts, and “reluctant unit” conversion attempts. Other examples include the relative costs of: obtaining and “cleaning” a new source of administrative records; development of a new component of a survey collection or processing system; maintenance of system components; and training of interviewers on new instruments.

OMB regulations require reporting of aggregate estimated costs (to the government and to respondents) of a proposed federal survey. It appears, however, that for many government surveys, there is relatively little information available publicly regarding variable cost structures like those covered in the preceding paragraph. Development of standards in this area would require a large amount of work, but could contribute significantly to efficient management of federally sponsored surveys. Such standards may be of special interest in the current budgetary environment, because efficient management of cost-quality trade-offs in a constrained fiscal environment generally require a sophisticated empirical understanding of variable cost structures at a relatively fine level of aggregation.

3.1.4 Qualifications and Performance of Personnel and Organizations

Many of the abovementioned survey processes require a very high level of technical and managerial sophistication. The underlying individual and institutional capabilities often involve long learning curves and a high degree of specialization. These requirements can impact survey program standards in two related ways. First, a realistic set of technical standards will be effective only if they are developed, implemented and managed by personnel with extensive training and experience in applicable areas of survey work. Second, it can be appropriate to have a component of survey program standards focused specifically on the training, experience and track records of the individuals and institutions that are scheduled to carry out the proposed survey work.

3.2 Calibration of Standards: Width and Height of the Bar

For each of the areas considered in Section 3.1, development of appropriate standards will involve decisions on *calibration* of that standard, i.e., determination of the scope of work covered by the standard (sometimes colloquially called the “width of the bar”) and the stringency of the standard (or the “height of the bar”). For statistical programs, the appropriate scope of coverage is largely an empirical question driven by whether the proposed standard will contribute substantially to produce benefits, reduce costs or manage risks for some legitimate major stakeholders.

One may loosely classify the stringency of a standard into one of three groups. *Descriptive calibration* compares performance with current common practice. The resulting standards may be useful to maintain current institutional strengths in changing budgetary or management environments, or to codify reasonable and customary practice for the benefit of new entrants or organizations that lag behind standard practice. *Normative calibration* compares performance with current or historical best practice. Normatively calibrated standards will tend to “raise the bar” for the performance of many participants. *Aspirational calibration* compares performance with scientific principles or other idealized criteria. Aspirational calibration may be appropriate for surveys that differ fundamentally from those commonly conducted by high-quality survey research organizations. Examples might include surveys of populations that are markedly different from standard household or establishment populations, or surveys conducted under unusual constraints. However, such surveys may require extensive monitoring and mid-course adjustments, and may not be compatible with customary approaches to

standards. In such cases, it is especially important to distinguish carefully between *standards* and *nominal optima* developed under idealized conditions.

3.3 Mechanisms for Implementation, Enforcement and Revision of Standards

Four important issues in the implementation, enforcement and revision of standards are as follows. First, and arguably most important, is the training of personnel. Some of this training may center on the specifics of the standards, but of much greater concern is the quality and depth of training in the fundamentals (e.g., mathematical statistics, behavioral sciences and related substantive fields) that led to the standards. A second issue is the co-ordination of the implementation plan with externally driven features of the survey, e.g., revision cycles related to periodic census collections, changes in classification systems, system upgrades and budgetary constraints. Third, an agency will need to choose the forms in which to enforce the standard, e.g., through transparent external reporting, formal management monitoring and control, or third-party evaluation and enforcement. Fourth, in keeping with comments in Mitchell (1998), the monitoring of standards involves both the evaluation of compliance by a particular survey program; and follow-up evaluation of the true impact that the implemented standard had on the benefits, costs and risks for the survey program and its primary stakeholders.

With these four elements in mind, it is useful to consider three primary questions in the development, implementation and enforcement of standards for survey programs:

1. Are the standards and enforcement mechanisms coherent with the overall value system and incentive structure of the institution, and of the overall statistical system?
2. If not, do the proposed standards flow from a serious consensus – and a workable plan - to change the institutional culture and incentive structure?
3. Do key stakeholders have a reasonably clear understanding of benefits, costs and risks arising from these standards?

3.4 Special Issues for Government-Sponsored Statistical Programs

Many of the ideas in Sections 3.1-3.3 apply to a wide range of standards for general types of technology. Surveys and other areas of statistical methodology,

however, also have some relatively unusual characteristics that warrant special consideration in the development of standards. First, empirical feedback loops on statistical performance may be limited, slow, inefficient or not transparent to non-specialists. For example, a general data user can readily observe – and have concerns about – the temporal volatility of a published series of monthly estimates. However, without specialized analyses, that user may not be able to distinguish between volatility attributable to, respectively, monthly variability in the underlying true series and sampling error or other forms of noise. Similarly, the reduction of biases related to selection mechanisms, nonresponse or reporting error may require a substantial investment of agency resources, but it may be difficult or impossible to provide nonspecialists with a transparent assessment of the magnitude of the resulting reduction in bias for a particular set of published estimates. To some degree, external third-party peer review by responsible specialists can help to address these issues with limited feedback loops.

Second, customary statistical properties (e.g., mean squared error) may not have strong linkage with the utility functions of the primary stakeholders of a given survey program. Consequently, statistical agencies need to consider ways in which to develop standards that are consistent both with underlying statistical methodology and with the primary articulated needs of data users and other important stakeholders.

Third, some areas of technology develop and implement standards largely in an audit-type framework, with a correspondingly adversarial style of review and enforcement. This can be appropriate for certain areas that involve commodity work (in which there is relatively little additional value provided by products that go beyond minimum standards) or that require exclusion of a substantial number of low-quality participants. Most survey work by federal statistical agencies, however, fits much better with a continuous-improvement model that emphasizes a collaborative work environment and recognizes that management of a high-performing survey organization requires many nuanced judgment calls and work that consistently goes far beyond minimum standards, with a corresponding allocation of resources.

Fourth, there is increasing interest in the use of administrative record sources in conjunction with, or in place of, customary sample surveys. However, many statistical agencies have little or no direct control over important technical features of the administrative record systems, e.g., the populations covered; variables included in the records; definitions of these variables; the completeness and timeliness of records; and other measures of data quality. Consequently, the direct development and implementation of standards for statistical uses of administrative record systems may be subject to severe practical constraints, and may center on

the extent, if any, that one may use the administrative record source for, respectively, direct production of statistical reports; for auxiliary data integrated into both sample design and estimation work; or for frame information only.

Finally, statistical standards processes for government agencies are also complicated by the presence of multiple stakeholders with potentially disparate utility functions; variability of agency structure (large centralized statistical agencies, decentralized statistical systems or statistical contract work sponsored by non-statistical agencies); and changing expectations of stakeholders regarding, e.g., burden, data quality, confidentiality, disintermediation of information, and degree of data access.

3.5 Communication of Standards to External Stakeholders

As noted in Section 2, standards can be important in ensuring the efficient operation of markets for many types of technology, including survey methodology. However, that efficient operation requires extensive and ongoing communication with a wide range of stakeholders, many of whom are outside of the statistical program of interest. Three external groups are of special interest. First, data users need to understand the extent to which statistical standards provide assurance regarding the quality of specific data releases, and the impact of those quality characteristics on specific uses of the data. Second, standards can also impose additional burden or other constraints on some stakeholders. For example, precision requirements in publication standards may limit the level of aggregation at which we can publish estimates. In addition, disclosure-limitation standards may lead to suppression of cells in tabular reporting or limitations on design information and extreme values included in microdata releases. Third, in many statistical programs, a substantial amount of work is carried out by groups beyond a single statistical agency. Examples include programs managed jointly by multiple federal agencies; federal-state cooperative programs; and work performed by universities or the private sector.

4. Prospective Benefits of Statistical Standards

Within the general framework for statistical standards defined by Section 3, we can identify several classes of benefits that standards may convey to survey stakeholders. Some of these benefits are relatively tangible and have direct linkage with the utility functions of some stakeholders. For example, many standards are linked directly to the improvement in one or more dimensions of data quality, e.g., accuracy, timeliness,

accessibility, relevance, interpretability and coherence, per the framework for data quality outlined in Brackstone (1999).

In addition, some standards can lead directly to reduction of tangible costs for some stakeholders. For instance, compliance with standards for efficient sample design can reduce data collection costs for a statistical agency; burden constraints can reduce costs for respondents; and requirements for standard data formats, and methodology reports can reduce costs for data users.

Beyond these tangible benefits, and perhaps most important, is the fact that standards can contribute to the improved coherence and efficiency of markets for statistical information. Standards, and supporting technical material, provide a common language and set of shared expectations, which in turn reduce search and transaction costs. In addition, if managed appropriately, standards can enhance the external stature of the statistical agency and support the professionalism of internal agency staff and other stakeholders. This in turn can significantly reduce some forms of risk for the statistical agency and its data users.

For example, the classic Gresham's Law ("bad money drives out good money") has a close equivalent for statistical work, in which the presence of low-quality statistical information will tend to discourage the production of high-quality statistical information, unless non-specialist stakeholders have low-cost, transparent mechanisms by which they can distinguish between low- and high-quality data. Statistical standards, and related communication with non-specialists, can provide one possible mechanism for this.

Similarly, the methodological community often encounters variants on the "tragedy of the commons" related to under-investment in certain critical resources. For example, with very few exceptions statistical methodology has been treated as non-proprietary, and acceptance of a methodological innovation has generally required dissemination of that innovation in the public domain. In addition, high-level technical training of methodological personnel will often involve very long learning curves; long lead times between the initial training and its contribution to a specific statistical program; and the risk that personnel will leave the institution soon after completion of their in-depth training. Formal standards for the training, experience and performance of high-level technical personnel can help to ameliorate these costs and risks, and thus will convey benefits to individual statistical programs and the overall methodological community.

5. Prospective Costs and Risks of Standards

To use a rough analogy from time series, all standards impose a "filter" through which stakeholders

will view a statistical program as well as its products and processes. Furthermore, if such a filter is nontrivial, it will distort or eliminate some information about the survey program that may be important for one or more stakeholders. This in turn can impose significant costs and risks on the survey program and its stakeholders; and can lead to unintended changes in the operations of a survey program. Results may include direct costs to some survey stakeholders, general loss of efficiency, and several forms of indirect costs and increased risks.

5.1 Direct Costs

Some of the costs arising from standards allow for relatively direct evaluation. For example, careful development of standards generally requires a substantial amount of time from personnel who have both extensive knowledge of the applicable methodology and program, and the "soft skills" required for negotiation and development of consensus across a wide range of stakeholders. Additional costs arise from related work in implementation, training, compliance, reporting and external review or enforcement. In an environment with finite resources, this in turn can lead to significant opportunity costs.

5.2 Loss of Efficiency

In many cases, if a survey organization does not develop and apply standards appropriately, they can potentially lead to serious losses of efficiency. For example, meaningful standards invariably impose some barriers to entry into the market for statistical work. This may be a reasonable outcome if these barriers are consistent with applicable laws and if they do not exclude organizations that reasonably can be expected to produce high-quality work. On the other hand, artificial barriers that do not arise from objective, high-priority quality criteria can result in a substantial loss of productive competition.

In addition, the general literature on standards broadly recognizes that the process of standard-setting and enforcement can lead to various forms of "rent seeking" behavior, in which a participant seeks to manipulate a regulatory or standards-setting process to obtain comparative advantage over other potential market participants, without providing any additional value to the "customers" of the program. The general literature on rent-seeking behavior is large and controversial (see, e.g., Mitchell (1998), Yandle (1999) and references cited therein), and a general treatment is well beyond the scope of the present paper. Also, one may not always have a perfectly clear dividing line between rent-seeking and sincere advocacy of a given quality standard for statistical programs. For the current discussion, it suffices to say that participants in standards-setting work for statistical

programs should bear in mind the potential for rent-seeking behavior, and should attempt to minimize such behaviors through a careful and transparent examination of the value provided to legitimate survey stakeholders through the proposed standards.

Furthermore, inappropriately developed and implemented standards can degrade the performance of a survey organization. For instance, such standards can lead to internal confusion regarding institutional priorities, and can implicitly encourage satisficing or minimax behaviors that degrade overall performance. A simple version of this confusion arises in organizations dominated by constrained budgets or low-bid contracting procedures, wherein minimum standards can become *de facto* maximum standards, and thus discourage serious efforts to optimize survey processes. A more extreme version of this confusion can arise if an organization has critical functions that are difficult or impossible to standardize. For such an organization, establishment and rigid enforcement of readily formulated standards in many low-priority areas can be very counterproductive.

To address these issues, survey organizations should focus standards in selected high-priority areas; calibrate the standards with legitimate measures of stakeholder value; and integrate the resulting standards with incentive structures that encourage the allocation of resources toward serious process improvement.

5.3 Indirect Costs and Related Risks

In extreme forms, some of the phenomena covered in Section 5.2 can impose additional indirect costs and related risks on the statistical organization and other stakeholders. For example, many areas of survey methodology require a significant degree of refined judgment. Reductionist management of such work as a standardized commodity can lead to a number of problematic outcomes, including exacerbation of institutional and individual patterns of risk aversion; creation of unnecessary barriers to subsequent innovations to improve quality and efficiency; degradation of morale among high-performing personnel; and consequent risks and efficiency losses related to recruitment and retention.

In addition, application of standards to a survey program can incur forms of *operational risk*, i.e., the risk that a standard will not be used as intended, and will in turn lead to degraded performance of the survey program. This risk can arise in many settings, but may be especially likely when a standard is applied to unusual populations, data sources, survey procedures or inferential goals. For instance, application of simple standards on response rates can be problematic for surveys that include multiple stages of sampling or extensive use of auxiliary information; cf. the recommendation by Little and Robin (2002) that analysts use a “fraction of missing information” quality measure in place of customary

response-rate measures for some discussions of the impact of nonresponse on survey data quality.

Operational risk may be especially acute for the aspirational standards discussed in Section 3.2. Clear statements of the intended scope of a statistical standard and related cautionary remarks can help one to manage operational risk for statistical standards.

6. Conclusions

6.1 Summary

This paper has sought to supplement the previous literature on standards for statistical programs by exploring the ways in which standards can enhance (or impede) the development, adoption and diffusion of innovations in statistical technology (methodology and specific statistical products) for a wide range of stakeholders. That approach led to the interpretation of standards as institutional commitments on investments in specified areas of statistical technology; and to the corresponding requirements for evaluation of the prospective benefits, costs and risks arising from these standards. Moreover, when they are integrated fully into common agency practice, the standards themselves constitute an important part of the institutional capital for individual statistical organizations and the federal statistical system as a whole.

Within the wide range of ideas covered here, five points warrant special emphasis. First, standards generally will be the most effective if we develop them in ways that are coherent with the values, incentive structures and scope of feasible practice for individual agencies and the overall federal statistical system. Second, it is important to link standards to tangible benefits for stakeholders in high-priority areas. Third, it is also important to ensure that for applicable stakeholders, compliance with standards will involve costs and risks that are low relative to prospective benefits for the federal statistical system and its data users; and that appropriate contractual and funding mechanisms mediate the predominant cost-benefit trade-offs. Finally, for essentially all standards for statistical programs, there are two critical factors: transparency on methodology, cost structures and empirical results; and appropriate high-level technical training and qualifications of personnel.

6.2 Research Issues

In closing, the framework suggested in this paper leads to several open questions that might warrant in-depth methodological and empirical research.

1. What empirical information is currently available regarding stakeholders' perceptions of benefits, costs and risks arising from a given statistical survey program? Has there been any methodological work done on, e.g., development and testing of instruments to measure these perceptions? (See, Singer et al. (1995) for one partial example involving confidentiality assurances.)
2. In keeping with comments in Sections 3 through 5, to what extent do we have empirical evidence on the linkage between specific statistical standards and tangible value for the stakeholders (respondents, data users and statistical agencies) of statistical survey programs?
3. Based on results from questions (1) and (2), can one develop an informative mathematical characterization of trade-offs among benefits, costs and risks arising for a proposed set of survey standards?
4. Based on results from question (3), what are the marginal benefits, if any, of specific proposed technical standards, conditional on broad compliance with stringent standards regarding transparency of processes and technical training of personnel?

Bayless, D.L., Cahalan, M. and Stufflebeam, D.L. (1990). Written standards as a process to improve survey data quality: A review of selected national standards. Paper presented to the Section on Survey Research Methods, American Statistical Association at the 1990 Joint Statistical Meetings, Anaheim, California.

Beatty, Paul, Willis, Gordon, Hunter, Jennifer E. and Miller, Kristen (2005). Design of the q-bank: Determining concepts, content, and standards. *ASA Proceedings of the Joint Statistical Meetings*, 981-988. American Statistical Association (Alexandria, VA)

Biemer, Paul P. (1990). Comment on "Development of standards and quality procedures in federal surveys" *ASA Proceedings of the Section on Survey Research Methods*, 18-19 American Statistical Association (Alexandria, VA)

Bisgaard, Søren, (1991). Design of standards and regulations. *Journal of the Royal Statistical Society, Series A: Statistics in Society*, **154**, 93-96

Blyth, B. (2007). A new international standard for survey process quality control (ISO 20252). *ISI Newsletter* **30**, p. 25.

Boen, J. and Smith, H. (1975). Should statisticians be certified? *The American Statistician*, **29**, 113-114

Boen, James R. (1990). Studying certification for statisticians. *The American Statistician*, **44**, 184.

Boulanger, Michèle, Boulanger, Michele, Johnson, Mark, Perruchet, Christophe and Thyregod, Poul (1999). Evolution of international statistical standards via life cycle of products and services. *International Statistical Review*, **67**, 151-171

Brackstone, Gordon (1999). Managing data quality in a statistical agency. *Survey Methodology* **25**, 139-149.

Brackstone, Gordon, Johannis, Paul and Kovar, John (2004). Standards for surveys at Statistics Canada. *ASA Proceedings of the Joint Statistical Meetings*, 1454-1461 American Statistical Association (Alexandria, VA)

Braithwaite, William R. (2001). Data field standards and the Health Insurance Portability and Accountability Act. *Statistics in Medicine*, **20**, 1323-1330

Burgess, James F., Jr., Christiansen, Cindy L., Michalak, Sarah E. and Morris, Carl N. (2000). Medical profiling: Improving standards and risk adjustments using hierarchical models. *Journal of Health Economics*, **19**, 291-309

Acknowledgements and Disclaimer

The views expressed in this paper are those of the author and do not necessarily represent the policies of the U.S. Bureau of Labor Statistics. The author thanks Katharine Abraham, Bill Barron, David Binder, Ron Bosecker, Lynda Carlson, Larry Cox, Don Dillman, Bob Eddy, Bob Fay, Nancy Gordon, Bob Groves, Brian Harris-Kojetin, Phil Kott, Janice Lent, Rod Little, Tom Petska, Don Rubin, Fritz Scheuren and Clyde Tucker for many helpful discussions of this topic.

References

American Statistical Association (1999). Ethical Guidelines for Statistical Practice. Available at: <http://www.amstat.org/profession/index.cfm?fuseaction=ethicalstatistics>

Australian Bureau of Statistics (2007). *Methods, Classifications, Concepts and Standards*. Available through: <http://www.abs.gov.au/websitedbs/d3310114.nsf/Home/Methods,+Classifications,+Concepts+&+Standards>

Bushery, John M. and Pamela D. McGovern (2005). Implementing and improving quality standards at the U.S. Census Bureau. *Proceedings of the 2005 Research Conference of the Federal Committee on Statistical Methodology*. Available through: http://www.fcsm.gov/05papers/Bushery_McGovern_VIIC.pdf

Cecco, Kevin and Cohen, Stephen (2004). Standards for statistical surveys for producing and disseminating survey estimates in the U.S. statistical system. *ASA Proceedings of the Joint Statistical Meetings*, 1494-1502. *American Statistical Association (Alexandria, VA)*

Chiswick, Carmel U. (1987). Labor, leisure and home production: Implications of the new international standards for statistics. *ASA Proceedings of the Business and Economic Statistics Section*, 1-5. *American Statistical Association (Alexandria, VA)*

Colledge, Michael and March, Mary (1997). Quality policies, standards, guidelines, and recommended practices at national statistical agencies. *Survey Measurement and Process Quality*, 501-522. Lyberg, Lars (ed.), Biemer, Paul (ed.), Collins, Martin (ed.), de Leeuw, Edith (ed.), Dippo, Cathryn (ed.), Schwarz, Norbert (ed.) and Trewin, Dennis (ed.) *John Wiley & Sons (New York; Chichester)*

COPSS Committee on Symbols and Notation (1965). Recommended standards for statistical symbols and notation. *The American Statistician*, 19, 12-14

Court, Andrew T. (1952). Statistical standards. *The American Statistician*, 6, 6

Deming, W. Edwards (1958). Standards of probability sampling for legal evidence. *The American Statistician*, 12, 25-26

Dillman, D.A. (1996). Why innovation is difficult in government surveys (with discussion). *Journal of Official Statistics* 12, 113-197.

Eltinge, John L. (2004). Discussion: Standards and gold standards in the development, adoption and diffusion of statistical technology for government agencies. *ASA Proceedings of the Joint Statistical Meetings*, 1555-1557. *American Statistical Association (Alexandria, VA)*

Farlie, D. J. G. (1989). Standards and statistics. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences*, 327, 573-579

Fellegi, Ivan P. and Ryten, Jacob (2005). The effectiveness of a supranational statistical office: Pluses, minuses, and challenges viewed from the outside. *Journal of Official Statistics*, 21, 145-169

Fenwick, D. and G. Tippen (2003). Quality Management Using ISO 9000 for Price Indices in the UK. *Journal of Official Statistics* 19, 365-382.

Folsom, R. E., Horvitz, D. G. and LaVange, L. M. (1989). The design of surveys using measurement design standards. *ASA Proceedings of the Section on Survey Research Methods*, 80-87. *American Statistical Association (Alexandria, VA)*

Frankel, Martin R. (1975). The development of broadcast rating council standards for accuracy of approximate standard error estimates. *ASA Proceedings of the Business and Economic Statistics Section*, 106-109. *American Statistical Association (Alexandria, VA)*

Freedman, Stanley R. (1990). Quality in federal surveys: Do standards really matter? *ASA Proceedings of the Section on Survey Research Methods*, 11-17 *American Statistical Association (Alexandria, VA)*

Freeman, William W. K. (1956). Report of the Ad Hoc Committee to Explore Opinion on Standards. *The American Statistician*, 10, 13-13

Gillman, Daniel W., Appel, Martin V. and LaPlant, William P., Jr (1996). Statistical metadata management: A standards based approach. *ASA Proceedings of the Section on Survey Research Methods*, 55-64. *American Statistical Association (Alexandria, VA)*

Gonzalez, Maria E. (ed.), Ogus, Jack L. (ed.), Shapiro, Gary (ed.), Tepping, Benjamin J. (ed.), Ferber, Robert (ed.) and Barabba, Vincent P. (ed.) (1975). Standards for discussion and presentation of errors in survey and census data (Suppl. to No. 351). *Journal of the American Statistical Association*, 70, 5-23

Hansen, Morris H. (1952). Statistical standards and the census. *The American Statistician*, 6, 7-10

Healy, M.J.R. (1978). Is statistics a science? *Journal of the Royal Statistical Society, Series A* 141, 385-393.

Hibbert, Jack (1993). The need for research into the effectiveness of international statistical systems and standards. *Statistics and Computing*, 3, 194-196

Hogan, Howard R. and Mulry, Mary H. (1987). Operational standards for determining the accuracy of census results. *ASA Proceedings of the Social Statistics*

Section, 46-55 *American Statistical Association* (Alexandria, VA)

Horvitz, Daniel G., Folsom, Ralph E. and Lavange, L. M. (1987). The use of standards in survey estimation. *ASA Proceedings of the Section on Survey Research Methods*, 546-551. *American Statistical Association* (Alexandria, VA)

International Standards Organization (2000). *ISO 9000:2000 – Quality Management Systems – Fundamentals and Vocabulary*

International Standards Organization (2006). *ISO 20252: 2006 – Market, Opinion and Social Research – Vocabulary and Service Requirements*.

Johnson, Timothy and Owens, Linda (2003). Survey response rate reporting in the professional literature. *ASA Proceedings of the Joint Statistical Meetings*, 127-133 *American Statistical Association* (Alexandria, VA)

Kartha, C.P. (2001). ISO9000:2000 quality system standards and Baldrige criteria *ASA Proceedings of the Joint Statistical Meetings*, - *American Statistical Association* (Alexandria, VA)

Katz, R., ed. (2004). *The Human Side of Managing Technological Innovation, Second Edition*. New York: Oxford University Press.

Larson, L. Odell and Narain, Pratap (1998). International standards and definitions for agricultural statistics and requirements for World Census of Agriculture (Pkg: p7-55) *Agricultural Statistics 2000 Proceedings*, 42-50. Holland, Theresa E. (ed.) International Statistical Institute (Voorburg, The Netherlands)

Lehnen, Robert G. (1988). Statistical policy for state and local governments (Com: p16-22). *The American Statistician*, 42, 10-16

Lievesley, Denise (2001). Making a difference: A role for the responsible international statistician? *Journal of the Royal Statistical Society, Series D: The Statistician*, 50, 367-404.

Little, Roderick J.A. and Donald B. Rubin (2002). *Statistical Analysis of Missing Data, Second Edition*. New York: Wiley.

Long, Clarence D. (1952). Comments on "Statistical standards and the census." *The American Statistician*, 6, 10-14

Mahalanobis, P.C. (1965). Statistics as a key technology. *The American Statistician* 19, 43-46.

Martin, Elizabeth (2000). Survey standards and response rate disclosure. *ASA Proceedings of the Section on Survey Research Methods*, 75-78. *American Statistical Association* (Alexandria, VA)

Melnick, Dan (2002). The legislative process and the use of indicators in formula allocations. *Journal of Official Statistics*, 18, 353-369

Mindlin, Albert (1977). Comments: "Professional standards of statisticians in state and local government." *ASA Proceedings of the Social Statistics Section*, 200-202 *American Statistical Association* (Alexandria, VA)

Mitchell, Ronald B. (1998). Sources of transparency: Information systems in international regimes. *International Studies Quarterly* 42, 109-130.

Morton, J. E. (1952). Comments on "Statistical standards." *The American Statistician*, 6, 6-7.

Mulry-Liggan, Mary and Hogan, Howard (1986a). Research plan on census adjustment standards. *Proceedings of the Bureau of the Census Annual Research Conference, Volume 2*, 381-392. U.S. Bureau of the Census (Suitland, MD)

Mulry-Liggan, Mary and Hogan, Howard (1986b). Research plan on census adjustment standards. *ASA Proceedings of the Section on Survey Research Methods*, 566-570 *American Statistical Association* (Alexandria, VA)

National Center for Education Statistics (1991). Standards for Education Data Collection and Reporting (SEDCAR). Washington, DC: National Center for Education Statistics. Available at: <http://nces.ed.gov/pubs92/92022.pdf>

National Research Council (2005). *Principles and Practices for a Federal Statistical Agency, Third Edition*. Committee on National Statistics. Margaret E. Martin, Miron L. Straf and Constance F. Citro, Editors. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

O'Neill, Harry W. (1988). The need for and structure of standards of conduct for survey research practitioners. *ASA Proceedings of the Section on Survey Research Methods*, 6-8 *American Statistical Association* (Alexandria, VA)

- Osmint, Jeffrey and Absalom, Sandra (1993). Survey statistical standards: Development and role. *ASA Proceedings of the Section on Government Statistics*, 225-227. American Statistical Association (Alexandria, VA)
- Paulk, Mark C. (1994). A Comparison of ISO 9001 and the Capability Maturity Model for Software.
- Rogers, E.M. (1995). *Diffusion of Innovations, Fifth Edition*. New York: Free Press.
- Rosenberg, Harry M. and Lunde, Anders S. (1977). Professional standards of statisticians in state and local governments. *ASA Proceedings of the Social Statistics Section*, 192-199. American Statistical Association (Alexandria, VA)
- Salter, Liora, Levy, Edwin and Leiss, William (1988). Mandated science: science and scientists in the making of standards. *Kluwer Academic Publishers Group (Dordrecht; Norwell, MA)*
- Schmid, Calvin F. (1976). The role of standards in graphic presentation. *ASA Proceedings of the Social Statistics Section*, 74-81. American Statistical Association (Alexandria, VA)
- Seltzer, William (1996). Statistical standards and national, regional, and global requirements and capabilities. *ASA Proceedings of the Section on Government Statistics*, 12-21. American Statistical Association (Alexandria, VA)
- Shapiro, Gary, Pollock, Jesse, Harley, M. Diana and Beach, Mary Ellen (1985). Statistical standards at the Census Bureau: Past, present, and future. *Proceedings of the Bureau of the Census Annual Research Conference, Volume 1*, 481-491. U.S. Bureau of the Census (Suitland, MD)
- Singer, Eleanor, Dawn R. Von Thurn and Esther R. Miller (1995). Confidentiality assurances and response: A quantitative review of the experimental literature. *Public Opinion Quarterly* **59**, 66-77.
- Slater, Melaney and Christensen, Howard (2002). Applying AAPOR's final disposition codes and outcome rates. *ASA Proceedings of the Joint Statistical Meetings*, 540-545. American Statistical Association (Alexandria, VA)
- Smith, Tom W. (1999). Standards for final disposition codes and outcome rates for surveys. *Federal Committee on Statistical Methodology Research Conference, Vol. 5*, 49-58. U.S. Bureau of the Census (Suitland, MD)
- Spencer, Bruce (1986). Conceptual Issues in Measuring Improvement in Population Estimates. *Proceedings of the Bureau of the Census Annual Research Conference, Volume 2*, 393-407. U.S. Bureau of the Census (Suitland, MD)
- Statistics Canada (2003). *Statistics Canada Quality Guidelines, Fourth Edition*. Available through: <http://www.statcan.ca/english/freepub/12-539-XIE/12-539-XIE03001.pdf>
- Statistics Finland (2006). *Quality Guidelines for Official Statistics: Handbooks 43b*. Helsinki, Finland: Statistics Finland. Available through: http://tilastokeskus.fi/tk/ti/laatuutilastoissa/alku_en.html
- Statistics New Zealand (2007). *Quality Standards in Official Statistics*. Available through: <http://www.stats.govt.nz/about-us/corporate-reports/post-election-briefing-2002/quality-stds.htm>
- Stephens, Ken (1977). Standards: A new frontier for quality. *Journal of Quality Technology*, 9, 70-81
- Tukey, John (1986). Comments on "Deriving appropriate standards for Census undercount adjustment." *Proceedings of the Bureau of the Census Annual Research Conference, Volume 2*, 408-410. U.S. Bureau of the Census (Suitland, MD)
- Tupek, Alan R., Arends, William, Casselberry, Jay, Cecco, Kevin, Cohen, Stephen, Flanagan, Patrick, Harris-Kojetin, Brian A., Madans, Jennifer, O'Connor, Karen and Seastrom, Marilyn (2004). Updating federal standards for statistical surveys. *ASA Proceedings of the Joint Statistical Meetings*, 1795-1799. American Statistical Association (Alexandria, VA)
- United Kingdom Office of National Statistics (2002). National Statistics Code of Practice: Statement of Principles. Available through: http://www.statistics.gov.uk/about/national_statistics/cop/downloads/StatementRD.pdf
- United States Office of Management and Budget (1977a). *Statistical Policy Directive Number 1: Standards for Statistical Surveys*.
- United States Office of Management and Budget (1977b). *Statistical Policy Directive Number 2: Publication of Statistics*.
- United States Office of Management and Budget (2006). *Standards and Guidelines for Statistical Surveys*. Available at:

http://www.whitehouse.gov/omb/infoereg/statpolicy/standards_stat_surveys.pdf

Wainer, Howard (2000). Testing the disabled: Using statistics to navigate between the Scylla of standards and the Charybdis of court decisions. *Chance*, 13, 42-44

Webb, James E. (1965). Decision-making and statistical standards. *The American Statistician*, 19, 16-20

Yandle, B. (1999). Bootleggers and Baptists in retrospect. *Regulation* 22, 3.

Zemach, Rita (1977). Comments: "Professional standards of statisticians in state and local government." *ASA Proceedings of the Social Statistics Section*, 203-204. *American Statistical Association (Alexandria, VA)*

Zukin, Cliff (2006). The future is here! Where are we now? How do we get there? AAPOR Presidential Address, May 19, 2006. *Public Opinion Quarterly* 70, 426-442.