NISS

Survey Costs: Workshop Report and White Paper

Alan F. Karr and Michael Last

Technical Report Number 161
September 2006

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

Contents

1 Executive Summary

This document is both the report of a workshop on survey costs, held on April 18–19, 2006 in Washington, DC and a white paper articulating a research agenda derived from breakout and plenary discussions at the workshop.

The workshop was organized by the National Institute of Statistical Sciences (NISS), at the request of the Office of Research and Methodology at the National Center for Health Statistics (NCHS) and held in facilities provided by the National Center for Education Statistics (NCES). Alan Karr, Director of NISS, chaired the organizing committee, whose other members were Lawrence Cox (NCHS), John Eltinge (BLS), Graham Kalton (Westat), Daniel Kasprzyk (MPR), Myron Katzoff (NCHS), Partha Lahiri (University of Maryland), Judy Lessler (Chatham Research Consultancy), Marilyn Seastrom (NCES), Alan Tupek (Census) and Doug Williams (Williams Consulting).

1.1 Purpose

The purposes of the workshop were to:

- Articulate clearly the survey cost problems faced by the federal statistical agencies.
- Construct a research agenda for survey cost methodology and modeling that addresses these problems.
- Identify the key next steps in pursuing the research agenda.

1.2 Findings

The principal findings of the workshop were as follows:

1. Existing theory, methodology and software tools are not adequate to address current, let alone emerging, problems associated with survey costs.
2. The need for research to address the gaps is pressing and increasing.
3. The three most promising, and clearly inter-related, research thrusts, in terms of significant and immediate impact, are
    - Agent-based or other simulation models of surveys, for such purposes as prospective evaluation of policies and interventions and sensitivity analyses. Quantifiable indicators of cost and quality are an essential component of this thrust. See §7.1.1 for details.
    - Statistical modeling as a means of reducing costs. Examples include improved modeling of non-response bias in order to compensate for lower response rates resulting from cost considerations; using statistical modeling to produce valid statistical estimates at high geographical resolution without correspondingly large sample sizes; and employing statistical modeling to "borrow strength" across surveys in order to reduce costs. See §7.1.2 for details.
    - Survey integration—integrating data across surveys, which reduces costs by reducing duplication. Challenges include the need for powerful statistical modeling to characterize relationships not directly present in the data; practical impediments; temporal aspects; and data quality differences among surveys. See §7.1.3 for details.

Other research topics include paradata and dealing with multiple stakeholders. In particular, many survey costs are borne by agencies while quality principally benefits users of the data. See §7.1.4 for details.

4. There are substantive concerns to be borne in mind during the research. These include:

   - Access to sufficient and sufficiently high quality paradata on survey costs and effort.
   - Access to qualitative expertise about survey administration that resides in program managers, field managers and interviewers.
   - Capturing a full spectrum of fixed and variable costs. engagement by data collecting organizations.
   - High-level drivers of the problems, including federal statutes, Congressional mandates to collect data and data confidentiality; and scientific generalizability.

   See §7.2 for details.

5. Potential funders of the research include not only the federal statistical agencies but also the National Science Foundation and National Institutes of Health. See §7.3 for details.

6. Next steps include formation of working groups to undertake further discussion and exploratory research focused on particular needs. See §7.3 for details.

2 Workshop Goal

The principal goal of the workshop was to articulate a research agenda for survey cost methodology and modeling, including

In-process modeling, or responsive design. How should survey administrators respond during conduct of a survey to changes in budget or scope, or information about response rates?

Leveraging multiple data collections, in particular, by tradeoffs between cost and statistical modeling. Specific issues include designing targeted (small, in terms of sample size or data scale) data collections that leverage large-scale data collections and designing the large-scale data collections so that they can be leveraged.

Tools for principled tradeoffs between cost and quality. A central issue is that, whether made a priori or in-process, such tradeoffs must be made on the basis of predicted cost and predicted quality. Components of the problem include characterizing uncertainties and dependencies and use of decision theory and optimization.

To this end, the workshop focused specifically on *what is not possible currently*, that is, on research needs and issues, including:

- What are the fundamental problems?
- What are the high-leverage gaps, and where are the entry points to begin to address them?
- What techniques from statistics and other disciplines are most suitable for addressing the problems?
- What kinds of collaborations are needed and desirable?

3 Tutorials

The workshop began with and introduction by Alan Karr, followed by tutorials by John Eltinge (BLS) and David Banks (Duke University).[1]

3.1 John Eltinge: Survey Basics, Including Costs

The four principal parts of the tutorial focused on:

- Survey basics;
- Components of data quality and risk;
- Literature on survey costs;
- Methodological questions on survey cost structure and optimization thereof.

Important points raised in the tutorial and associated discussion:

- Legacy databases and production systems are a significant constraint.
- Principal sampling issues include incomplete candidate frames, nested population structures, rare subpopulations not reflected in frames (but sometimes with network structure), heterogeneity across units (leading to use of sampling weights), and heterogeneity across identifiable and subpopulations.

  As a result, exact models may not be known, raising issues of appropriate conditioning. Classical i.i.d. assumptions fail. Multiple stakeholders may have different inference needs. Consequently, classical criteria for performance of survey procedures are centered on relatively modest design-based properties.
- There are costs associated with modeling of survey data as well as data collection, which include costs of model fitting, use of auxiliary data, modification of production systems, dissemination of results and exposition of risks.
- Risks associated with surveys include model failure, misinterpretation of reported results, perceived reduction in value, damage to agency reputation and failure of data quality components such as accuracy, timeliness, accessibility, relevance, interpretability and coherence (Brackstone, 1999).
- Fixed costs of surveys are large and not well identified.
- Generalizability of cost information is a major issue.
- There are strong linkages among cost, information and data quality across multiple classes of stakeholders. For example, degradation in data quality (due, e.g., to model failure or delayed publication of results) may impose substantial costs on some data users.

[1] All tutorials and presentations are available on the NISS web site, at www.niss.org/affiliates/surveycost200604/surveycost-workshop200604.html. They can be downloaded individually or in a single PDF document.

3.2 David Banks: Simulation and Decision Theory

Among the provocative ideas raised in the tutorial and discussion were:

- Major research needs exist with respect to privacy protection, data quality requirements, small area estimation and use of administrative records. Each of these is driven to a significant extent by federal regulation.
- Challenges for survey methodology are falling response rates, total survey error (TSE), fast fielding, fast analysis and cost-effectiveness. New thinking by newly engaged people is needed.
- Decision theory and simulation are potential new tools for survey managers.
- Possible future solutions (some admittedly radical):
  - Elimination of federal surveys that are not needed.
  - Move from random samples to rotating panels of respondents.
  - Build a large simulation model that can be used for administration of surveys[2] or even as an alternative to surveys.
  - Build (smaller) simulation models of particular populations.
  - Incorporate (fast) statistical analyses into response designs of surveys.
  - Consider alternative forms of analyses, such as data mining, to reduce traditional emphasis on TSE.
  - Use dynamic programming to allocate interviewer effort.

  The cost implications of these kinds of strategies are not understood at this time.

4 Presentations

The three presentations were by Robert Groves (University of Michigan), Judith Lessler (Chatham Research Consultancy) and Alan Karr (NISS).

4.1 Robert Groves: Survey Budgets, Cost Models, and Responsive Survey Designs

The presentation had four principal parts:

- Why is there such a sparse literature on survey costs?
- Responsive survey designs.
- Observations and examples about survey costs.
- Towards a research agenda on survey costs.

The presentation introduced a term new to many—*paradata*, which are data about survey administration and operation, including costs. According to Groves, "Development and use of paradata is in its infancy."

Points raised in the presentation and discussion:

[2] See also §6.3.

- The extant survey cost literature (e.g., Groves (2004)[3] and Sudman (1967)) is largely confined to sample design optimization models, project management tools, costs of reducing non-response and two-phase samples for non-response reduction.
- Barriers to a richer literature include separation of sample design as a statistical problem from field design as a management problem and lack of paradata.
- Responsive survey designs (Groves and Heeringa, 2006):
  - Pre-identify a set of alternative features potentially affecting costs and quality.
  - Identify a set of indicators of the cost and quality properties.
  - Monitor indicators in initial stages of data collection.
  - Alter the survey based on cost/quality tradeoff decision rules.
  - Combine data from separate phases into a single estimator.

  Propensity-to-respond models are central to responsive designs. Such models have the potential to reduce costs by, for example, not wasting resources on respondents whose characteristics indicate that they are unlikely to respond even after intensive follow-up.
- How do sequential methods relate to responsive designs?
- Survey costs are rarely known accurately at design time.
- The impact of some "main effects" (e.g., survey mode) on costs is thought to be understood, at least by some people in some contexts.
- Much current attention to costs is on the interview process rather than on differences among sample strata.
- It is a challenge to get interventions such as responsive designs implemented by those in the field.[4]
- A continuous production model for surveys is needed, in which costs are associated with inputs that are linked by the model to outputs such quality. How detailed such models must be is not clear.
- Good paradata structures (that is, data structures for paradata) do not exist.
- Interviewer effort comprises 70% of variable costs. Therefore, real-time disaggregated effort allocation paradata are essential. Such data may become available for CAPI.
- Statistical models for forecasting respondent behavior resulting from various allocations of effort are a key need. Non-linearities abound.

[3] A reprint of Groves (1989).

[4] Whether there is a divide between survey managers and statisticians was discussed at length. One example such a divide of is statisticians not understanding the importance of interviewer effort and skill.

4.2 Judith Lessler: Leveraging Existing Data

The presentation focused on two central questions:

- Can the cost of surveys be reduced by making better use of existing administrative and survey data?
- Can we create an integrated methodology for use of primary and secondary data?

Lessler proposed the creation of an *integrated methodology for primary and secondary data* (IMPSD).

Key points in the presentation and discussion:

- Components of potential solutions—public use data files, administrative and geographical data, software tools, and ontologies and metadata standards—exist, but have not been assembled into solutions.
- Current deficiencies include surveys conducted without sufficient consideration of what is known already, lack of microdata and metadata from published surveys and studies and the growing tendency to investigate many research questions, leading to high respondent burden.
- An IMPSD entails:
  - Clear(er) specification of research questions.
  - Specification of what differences in survey results would yield different policy decisions.
  - Addressing needs of multiple stakeholders.
  - (Meta-)Analysis of secondary data.
  - Specification of remaining questions.
  - Design of primary data collection to "fill in the gaps."
  - Design of systems for data aggregation and integration.
- The NIH data sharing policy (National Institutes of Health, Office of Extramural Research, 2006) should be adopted for all government surveys.
- Quality of secondary data may be a major problem. There is also a question of reproducibility—should data that have not shown to be reproducible be used as secondary data? Should reproducibility be checked as part of the primary data collection?

4.3 Alan Karr: Principled Cost-Quality Trade-offs

The presentation outlined an abstract—until data and specifics are present—framework for comparing decisions at the design, intervention or analysis stage of a survey. Components of the framework:

- Definitions and quantified measures of cost and quality. For discussion of data quality metrics, see Karr et al. (2006b). Most of these, however, ignore *inferential* uses of the data.[5]
- Using concepts and techniques from optimization or microeconomics (cost-quality frontiers) to inform decision processes.

[5] Data utility measures based on inferential uses seem to have been proposed only in the context of data confidentiality (Gomatam et al., 2005; Karr et al., 2006a).

Key points from the presentation and discussion:

- The formulation can handle multi-dimensional indices of cost and quality.
- Computational difficulties may be severe. Efficient algorithms may not exist, although reduction of the decision space to the cost-quality frontier may help significantly. Techniques such as simulated annealing and genetic algorithms may depend on structures that are not present.
- In many situations, cost and quality are not deterministic functions of decisions, but must instead be treated stochastically. Dependences may be very hard to characterize. (Para)Data from which to estimate distributions of cost and quality may be lacking. Proper handling of uncertainties and dynamics requires Bayesian methods.

  Consequently, the power of analytical approaches may be limited: complications may be overwhelming, especially when dynamics are considered, sensitivity analyses may not be feasible, and the analytics may be disconnected from the domain science.
- Agent-based simulation models may be preferred to analytical approaches. The basics of the approach are:
  - Abstractions for agents (respondents, interviewers, ...), system responses (cost, quality) and interventions (e.g., increase incentives, increase intensity of non-respondent follow-up).
  - Simple, local models of agent interactions that produce the "correct" macroscopic behavior.
  - "Simple" models that relate agent behavior to interventions and "environmental" effects.

  The TRANSIMS metro-area traffic simulator (Los Alamos National Laboratory, 2006) was mentioned as one example.

  Validation of agent-based models would be a major challenge.
- The effects on data confidentiality of responsive design or *ad hoc* changes in surveys as a result of budget reductions or other considerations (for instance, major disruptions in data collection) are not understood.

5 Three-Minute Madness

NISS-sponsored workshops traditionally include an "*n*-minute madness" session at which every participant has the opportunity to articulate one issue that is particularly important to him or her. In this case, *n* was equal to three.

Among the issues, several of which amalgamate points raised by more than one participant:

- The need for tools to make principled, disciplined and potentially large reductions in sample sizes.
- The need to characterize the statistical consequences of reductions in sample size.
- Bi-directional interactions between survey costs and response rates, as well as the extent to which non-response bias analyses can compensate.
- Conducting the necessary research in production settings may be problematic.

- Is there a clear understanding of what strategies, for instance, to respond to budget reductions, have worked or not worked in the past?
- How much benefit is there from coordination—either intra- or inter-agency—across surveys?
- Are tools from computer science, in particular, ontology-based methods for data integration, relevant?
- How can cost-quality formulations be extended to include prioritized data collection needs? How can data collection needs be prioritized?
- Processing of data is a significant component of cost, especially when legacy software or hardware is involved.
- How can national and local data needs be balanced?
- Decentralized surveys such as the Behavioral Risk Factor Surveillance System (BRFSS) (Centers for Disease Control, 2006), which is run by states, present additional challenges. BRFSS also differs from most surveys in that it is used principally for surveillance rather than estimation.
- Should there be a distinct validation/verification for surveys, as an assessment of quality?
- To what extent must survey models (§7.1.1) incorporate agency and contractor effects?
- Technology may not be a reducer of cost so much as an improver of quality.

6 Breakout Discussions

Five breakout discussions were held, on paradata (led by Jonaki Bose, BTS), real-time monitoring of surveys (led by Marilyn Seastrom, NCES), agent-based modeling of surveys (led by Alan Tupek, Census), budget reductions (led by Doug Williams, Williams Consulting) and borrowing strength across surveys (led by Graham Kalton, Westat). A sixth discussion, on mapping data uses onto survey designs, was initiated, but it became clear that there was insufficient momentum, and the group re-distributed itself.

Summaries of the report from each group and associated discussion follow.

6.1 Paradata

This group focused on paradata as defined by Groves (§4.1)—data about the survey/data collection process itself. Their report emphasized:

- The need for more, and *uniform*, paradata.
- Usefulness of paradata regarding interviewer characteristics, although these data might be hard to obtain.
- Need for resources to analyze paradata.
- Need for examples of situations where good paradata have made a difference. NCHS' National Survey of Family Growth (NSFG) (National Center for Health Statistics, 2006b) was mentioned as one possible example.
- Desirability of developing strategies that rank survey units by response propensities (see §4.1).

The group raised engaging the Interagency Household Survey Nonresponse Group (IHSNG) as a next step.

6.2 Real-Time Monitoring

This group addressed the benefits and burdens of real-time monitoring of survey field operations. Points raised in their report:

- Rapidly increasing capability for real-time monitoring, using such technologies as CATI and GPS.
- Examples of what has been done, such as call logs that report the demeanor of respondents.
- Issues of how best to use real-time monitoring information. Examples mentioned included efforts at respondent conversion, variable incentives and mode switches (e.g., from telephone to face-to-face).
- The need for more cost detail.

There was considerable discussion of the realities and prospects for obtaining detailed cost information from—but for different reasons—either agencies or contractors. In the former case, details are less likely to be tracked, while in the latter, cost information would typically be deemed proprietary. Whether collection and provision of cost information to agencies could be incorporated into survey contracts was discussed at length, with no consensus.

6.3 Survey Models

This group focused on the desirability and feasibility of creating mathematical/statistical/simulation models of surveys, as recommended by Groves (§4.1), Lessler (§4.2) and Karr (§4.3). Key points in their report:

- Construction of an agent-based simulation model for surveys should proceed forthwith. Its main purpose would be to serve as a tool/environment to explore the effects of various decisions, interventions and exogenous events. Such experiments cannot be carried out in the field. Examples include changes in the number of interviewer supervisors, changes in interviewer training and sick interviewers.
- The simulation modeling effort should *start simple*. Complexity should be added sparingly and as necessary, in response to agency and research needs.
- Models must incorporate factors known by survey managers to be important, even if the path to doing so is unclear. The most frequently mentioned factor was interviewer skill/training/expertise.
- The National Health Interview Survey (NHIS) (National Center for Health Statistics, 2006a) is one possible initial context.

6.4 Budget Reductions

The impetus for this group is the increasing need for survey administrators to respond to mid-course budget reductions. Key points in their report:

- Historically, agencies have "lived for today," typically responding to budget cuts with reductions in sample size, even though these often do not reduce costs as much as expected.
- Alternative strategies, such as curtailing (detail of) surveys and survey integration (see also §4.2) need to be considered.
- A software tool to evaluate decisions (see also §6.3) is essential. Its development should include research on relationships among things that both can be changed and do affect cost and quality.

- It is essential to address both variable and fixed costs. Eltinge (§3.1) noted difficulties in identifying and, therefore quantifying, fixed costs.

6.5 Borrowing Strength

This group discussed issues raised in Lessler's presentation (§4.2), many of which have been noted elsewhere. Principal points in their report:

- Borrowing strength in ways that reduce respondent burden can both reduce cost and increase quality.
- Small area estimation should be explored as a means of borrowing strength across space, time and subsets of the population. Dependence on auxiliary data needs to be understood.
- Methods for creating useful synthetic data in other contexts such as confidentiality (Abowd and Lane, 2004; Raghunathan et al., 2003; Reiter, 2005; Rubin, 1993) should also be explored.
- Challenges to integrating surveys include accounting for differences in data quality and associations between surveys.

7 Research Agenda

At the end of the workshop there was no single, dominant priority for future research. However, there was consensus that:

- Existing theory, methodology and software tools are not adequate to address current, let alone emerging, problems.
- The need for research to address the gaps is pressing and increasing.
- The right research would have significant and immediate impact.
- The research challenges are substantial, and should be of interest not only to statistical scientists, but also to those interested in such fields as operations research, optimization and agent-based simulation.
- Without high-quality paradata and access to qualitative information about survey adminstration, the research will be speculative and its impact will be diminished.

A set of research issues on which there was broad interest expressed at the workshop is presented in §7.1. Some thoughts about next steps appear in §7.3.

7.1 Topics for Research

There were three threads of research about whose urgency and approachability there was strong consensus among workshop participants. These are discussed in §7.1.1–7.1.3.

7.1.1 Survey Models

The first of these is the need for *agent-based or other simulation models of surveys*. This theme, even though general, was clearly the most pervasive. Needs and reasons for developing such a model were articulated by Groves (§4.1), Lessler (§4.2), Karr (§4.3) and the Survey Models breakout discussion group (§6.3). The three most important of these are:

Prospective evaluation of policies and interventions, as an alternative to costly or, in most cases infeasible, "physical" experimentation.

Sensitivity analyses, to determine which controllable factors actually do affect survey costs and data quality, and to quantify the effects. The "production model" analogy articulated by Groves (§4.1 is compelling. It brings in quality/process control ideas of Deming (see, e.g., Deming (1986)): identify sources of variability, classify them as controllable or not, and manage those that are controllable.

Clarification of paradata needs, which is especially important in light of(perceived and real) difficulties in obtaining high quality paradata (§7.2).

Intractability of analytical approaches, as noted by Karr (§4.3), especially when stochastic elements must be taken into account.

There numerous challenges to development of such models, including:

- The need to construct quantifiable indicators of cost and quality. Without these, the approach outlined by Karr (§4.3), in particular, cannot be implemented. A complication is uncertainties associated with the indicators.
- Controlling the complexity of models, especially initial models. The TRANSIMS model used as an example by Karr (§4.3) cost more than $50 million to develop over nearly ten years, and was ultimately so machine- and data-intensive that it was largely unusable by the target users (metropolitan planning organizations—MPOs).
- Availability of high quality paradata, especially on cost and effort, at sufficient levels of detail.
- Ensuring a proper role for domain knowledge.
- Validation of models, given limited "real world" data.

7.1.2 Statistical Modeling as a Means of Reducing Costs

The second research issue for which there was consensus, and which was articulated in particular in the Borrowing Strength breakout discussion (§6.5), is to understand *how and to what extent statistical modeling can address cost-related issues*. The underlying rationale is that recently developed, and in some cases still emerging statistical technologies such as Bayesian models and computational methods, techniques for small area estimation and methodologies to generate high-utility synthetic data, can substitute for "raw" statistical power associated with sample sizes. This is only possible, of course, if correspondingly sophisticated models are also used to analyze the data.

Several examples of such possibilities were raised. For instance, can improved modeling of non-response bias compensate for lower response rates resulting from cost considerations? Here, the cost saving would result from pursuing non-respondents. A related issue is whether there are meaningful sample size—response rate tradeoffs.

A second example is whether statistical modeling—small area estimation, imputation, generation of synthetic data, . . .—can produce valid, useful statistical estimates at high geographical resolution without correspondingly high-resolution, and therefore large, sample sizes. It appears that the most significant "price" to be paid for employing such strategies is increased uncertainties, but this requires thorough investigation. The decision-theoretic framework described by Karr (§4.3) can, in principle, be used to assess tradeoffs between cost and uncertainty by viewing the latter as a measure of quality.

A third example is to apply design concepts such as matrix sampling in educational assessment (Childs and Jaciw, 2003) that would collect more information without increasing respondent burden.

A final example, discussed in more detail in §7.1.3, is how statistical modeling can allow borrowing strength across surveys in order to reduce costs.

7.1.3 Survey Integration

The third "consensus" need is for research aimed at reducing costs by what was variously termed "borrowing strength," "leveraging existing data" and "integrating data across surveys." Lessler's presentation (§4.2) and the Borrowing Strength breakout discussion (§6.5) addressed the issue directly, and it was present in several other contexts. The rationale is that current data collections are—to a significant extent, but one whose true extent and statistical implications remain to be elucidated in detail—duplicative.

The left-hand panel in Figure 1 shows two kinds of duplication between two surveys: duplicated respondents and duplicated attributes. The center panel shows the other extreme of surveys with no duplication. The right-hand panel shows two surveys in which the respondents are identical and there are no duplicated attributes. If Survey 1 has been conducted at some point in the past, and Survey 2, addressing Attribute Set 2, is to be conducted, then the right-hand panel in Figure 1 represents an ideal situation: no information already available is collected, and because the cases are the same, statistical modeling of relationships among attributes is not complicated by their having been collected in two different surveys.[6]

The research needed would address cases such as the left-hand and center panels in Figure 1. The cost savings are associated with Survey 2 not collecting information about the attributes in Set 1.

Among the challenges:

- The need for powerful statistical modeling, and potentially tools such as imputation for the data in the gray blocks, in order to characterize relationships between attributes in Set 1 and those in Set 2. This need is clearest in the center panel of Figure 1, where there are no cases in both surveys. There was reluctance among some participants to rely strategies that impute missing data values and the proceed to analyses.[7]
- Practical impediments.[8] For instance, the agency conducting Survey 2 may be unable to learn what respondents are present in Survey 1.
- Temporal aspects, as noted above. In right-hand panel of Figure 1, if Survey 1 occurs five years after Survey 1, then the attributes in Survey 1 may no longer be correct.[9]

[6] This is an oversimplication, since the values of Attributes in Set 1 may have changed since they are collected.

[7] The center panel in Figure 1 reveals a certain circularity to this process: the relationships needed to impute the missing values (gray blocks) are precisely those that are to be modeled.

[8] Some of which may be statutory.

[9] Indeed, not all respondents from Survey 1 may be available.

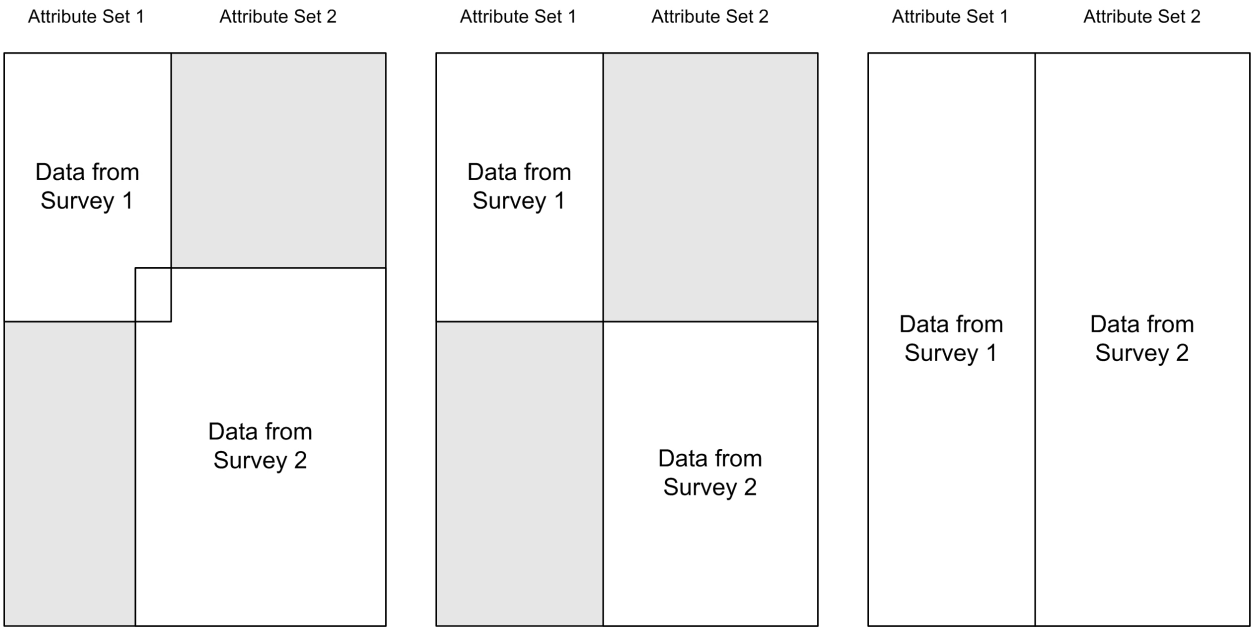

Figure 1: Graphical depiction of overlapping surveys, based on Lessler's presentation (§4.2). *Left:* surveys in which both records and attributes are duplicated. *Center:* surveys in which neither records nor attributes are duplicated. *Right:* surveys with identical cases but no duplicated attributes.

- Consequences of data quality differences between surveys. The "aging" just discussed is one aspect of this. At a later point in time, the (timeliness component of the) quality of the data from Survey 1 has diminished.

Models of the type described in §7.1.1 may be the only tools allowing full exploration of these questions. There is also evident overlap of this issue with the modeling issues in §7.1.2, since statistical modeling is the path to meeting these challenges.

An alternative approach to *ex post facto* integration is to append additional questions to existing surveys. While this approach does need to be explored, it entails problems of respondent burden and raises questions of who pays for what.[10]

7.1.4 Other Research Issues

Here we note briefly two other broad research topics that in general seemed either less pressing or more nebulous than those in §7.1.1–7.1.3.

The first is paradata (§4.1 and 6.1). What seems unclear at this point is whether there is an independent research agenda associated with paradata, or rather that the needs and challenges (e.g., paradata structures; see §4.1) will emerge from other research. Whether existing abstractions and tools for data quality[11] are

[10] Also, OMB was stated to be resistant to this approach.

[11] Which are argued in Karr et al. (2006b) to be inadequate for data of any sort.

adequate for paradata is one potential issue. Another is whether traditional concepts of uncertainty apply to paradata. And in any event, the need for the right paradata (§1.2 and 7.1.1) remains.

Dealing with multiple stakeholders is another major problem. In general, decision tools do not address this problem well. For survey costs, there is the additional complication that costs are borne by agencies while quality principally benefits users of the data.[12] How research such as that laid out in §7.1.1–7.1.3 can best accommodate multiple stakeholders is not clear.

7.2 Concerns

The research agenda presented in §7.1 was compelling and exciting to most workshop participants. At the same time, a number of concerns were articulated, some of which we list below. The word "concern" should be interpreted literally: these are not reasons not to undertake the research, but it is important to monitor and address them.

The main concerns:

- Availability of cost (para)data. This was, as implied in §6.1, the point raised at the workshop about with there was the greatest diversity of opinion. Some participants felt that paradata represent proprietary information that contractors will not release to agencies, let alone to researchers. Those at the other extreme felt that survey contracts could simply compel release of the information. Most felt that the truth lies somewhere between these extremes. Effort data may be an acceptable substitute, since associated costs could be treated parametrically, but many of the issues remain. Almost everyone agreed that paradata quality is an issue.
- Capturing a full spectrum of costs. Some parts of the workshop (e.g., §4.1) focused on costs of collection but Eltinge (§3.1) and others raised the need to consider not only other variable costs, such as processing, staff time to design and analyze surveys, . . ., but also fixed costs. There are also costs associated with making "incorrect" decisions, but these may be very hard to quantify. At a higher level, there are also opportunity costs: what is spent on one survey cannot be spent on other surveys, for other purposes entirely.
- Engagement by data collection organizations. There was a strong sense that immense knowledge about surveys, both quantified (see also the "availability of cost paradata item) and anecdotal, resides in the organizations that carry out survey data collections. How to access this knowledge by engaging such organizations in the research was much less clear.
- Whether data required by agent-based survey models (§6.3) exist, or can be expected to exist. There seemed to be agreement that the data do not currently exist in usable form. One important use of models, of course, is to clarify data needs.
- High-level drivers, including federal statues. Examples raised at the workshop include Congressional mandates to collect data at certain geographical resolutions, regulations that prohibit the kind of data sharing discussed in §7.1.3, and data confidentiality. Taken to an extreme, this concern implies that the research agenda in §7.1 should be undertaken in conjunction with high-level reviews asking such questions as What data, and with what level of uncertainty are really needed? By whom? For what purpose(s)? When?

[12] By contrast, a company that invests in improved data quality typically receives direct benefit (Redman, 2001). There is also indirect and more elusive benefit that a company benefits if its customers do.

- Understanding data uses. The short-lived "Mapping data uses onto survey designs" breakout discussion group was meant to address this and related issues. For example, understanding the "granularity" of policy decisions has direct implications for survey design, cost and analysis: low-granularity (e.g., yes/no) decisions may be able to tolerate relatively high levels of uncertainty in statistical estimates. Similarly, are precise, and expensive, county-level estimates required when decisions are made only at the state level?
- Generalizability. There is skepticism whether data quality is a science in the sense that results can be generalized across multiple contexts (Karr et al., 2001). There can only be more skepticism about some of the research issues in §7.1.

7.3 Next Steps

There was some, but limited, discussion at the workshop of desirable next steps. To some degree, the most desirable step is to "get on with the research" agenda in §7.1. There was also, however, the realization that doing the research in a high-impact way requires resources that may not be immediately available. Potential funders of the research include not only the federal statistical agencies[13] but also the National Science Foundation (NSF), for example, under the auspices of programs on Digital Government (National Science Foundation, 2006a) and Methodology, Measurement, and Statistics (National Science Foundation, 2006b).[14]

As a short-term measure and long-term complement to the research, there was considerable sentiment to form *Working Groups* that would continue to discuss and refine the research agenda. Ultimately, the groups might conduct or at least review the research. Working groups and volunteer leaders identified at the workshop were:

Paradata: No group leader identified. An interagency working group chaired by Nancy Bates is already focused on this area. Interested workshop participants are welcome to join.

Real-Time Monitoring: NCES personnel, led by Marilyn Seastrom

Survey Modeling: Alan Karr

Budget Reductions: Doug Williams

Borrowing Strength—Reducing Respondent Burden: John Eltinge

Borrowing Strength—Small Area Estimation: Partha Lahiri

Borrowing Strength—Integrating Surveys: Elizabeth Tighe

In addition, with support from the NCHS, NISS is currently undertaking exploratory research on survey models (§7.1.1).

Acknowledgements

Lawrence Cox (NCHS), John Eltinge (BLS) and Douglas Williams (then of NCHS) created the impetus for the workshop and this report.

[13] BLS, BTS, Census, EIA, NCES NCHS, data-collecting divisions of the NSF, and OMB were represented at the workshop.

[14] Specifically, under the sub-program for "Research on Survey and Statistical Methodology."

Financial support for the workshop and preparation of this report was provided by a contract awarded to NISS by the Centers for Disease Control and Prevention (CDC) on behalf of the Office of Research and Methodology at NCHS. The NCES generously provided meeting facilities for the workshop.

The authors thank Lawrence Cox, John Eltinge and Myron Katzoff (NCHS) for their very useful comments on drafts of this report.

References

Abowd, J. and Lane, J. (2004). New approaches to confidentiality protection: Synthetic data, remote access and research data centers. Available on-line at instruct1.cit.cornell.edu/~jma7/abowd-lane-barcelona-2004.pdf.

Brackstone, G. (1999). Managing data quality in a statistical agency. *Survey Methodology*, 25:139–149.

Centers for Disease Control (2006). Behavioral Risk Factor Surveillance System (BRFSS). Information available on-line at www.cdc.gov/brfss/.

Childs, R. A. and Jaciw, A. P. (2003). Matrix sampling of items in large-scale assessments. *Practical Assessment, Research and Evaluation*, 8(16).

Deming, W. E. (1986). *Out of the Crisis*. MIT Press, Cambridge, MA.

Gomatam, S., Karr, A. F., and Sanil, A. P. (2005). Data swapping as a decision problem. *J. Official Statist.*, 21(4):635–656.

Groves, R. and Heeringa, S. (2006). Responsive design for household surveys: Tools for actively controlling survey nonresponse and costs. *J. Royal Statist. Soc. A*. To appear.

Groves, R. M. (1989). *Survey Errors and Survey Costs*. Wiley, New York. Out of print. Currently available in 2004 Wiley-Interscience reprint.

Groves, R. M. (2004). *Survey Errors and Survey Costs*. Wiley-Interscience, New York. Paperback reprint of Groves (1989).

Karr, A. F., Kohnen, C. N., Oganian, A., Reiter, J. P., and Sanil, A. P. (2006a). A framework for evaluating the utility of data altered to protect confidentiality. *The American Statistician*, 60(3):224–232.

Karr, A. F., Sanil, A. P., and Banks, D. L. (2006b). Data quality: A statistical perspective. *Statistical Methodology*, 3(2):137–173.

Karr, A. F., Sanil, A. P., Sacks, J., and Elmagarmid, E. (2001). Workshop report: Affiliates workshop on data quality. Technical Report, National Institute of Statistical Sciences. Available on-line at www.niss.org/affiliates/dqworkshop/report/dq-report.pdf.

Los Alamos National Laboratory (2006). TRANSIMS: Transporation Analysis Simulation System. Information available on-line at transims.tsasa.lanl.gov/.

National Center for Health Statistics (2006a). National Health Interview Survey (NHIS). Information available on-line at www.cdc.gov/nchs/nhis.htm.

National Center for Health Statistics (2006b). National Survey of Family Growth (NSFG). Information available on-line at www.cdc.gov/nchs/nsfg.htm.

National Institutes of Health, Office of Extramural Research (2006). NIH Data Sharing Policy. Information available on-line at grants.nih.gov/grants/policy/data_sharing/.

National Science Foundation (2006a). Digital Government (DigitalGov). Information available on-line at www.nsf.gov/pubs/2004/nsf04521/nsf04521.htm.

National Science Foundation (2006b). Methodology, Measurement, and Statistics (MMS). Information available on-line at www.nsf.gov/pubs/2006/nsf06571/nsf06571.pdf.

Raghunathan, T. E., Reiter, J. P., , and Rubin, D. B. (2003). Multiple imputation for statistical disclosure limitation. *J. Official Statist.*, 19:1–16.

Redman, T. C. (2001). *Data Quality: The Field Guide*. Digital Press, Boston.

Reiter, J. P. (2005). Releasing multiply-imputed, synthetic public use microdata: An illustration and empirical study. *J. Royal Statist. Soc., Series A*, 168:185–205.

Rubin, D. B. (1993). Discussion: Statistical disclosure limitation. *J. Official Statist.*, 9:461–468.

Sudman, S. (1967). *Reducing the Cost of Surveys*. Transaction Publishers, Piscataway, NJ.

A Workshop Program

Survey Cost Workshop

National Center for Education Statistics, Washington, DC

Tuesday, April 18, 2006

9:00 AM	Tutorial I: John Eltinge, BLS: Survey Basics, including Costs
10:15	Break
10:45	Tutorial II: David Banks, Duke University: Decision Theory and Simulation
12:00 N	Lunch
1:00	Welcome and Introductions Alan Karr, NISS
1:15	Robert Groves, University of Michigan: In-Process Adaptation Discussion Leader: Alan Karr, NISS
3:15	Break
3:45	Judith Lessler, Chatham Research Consultancy: Leveraging Multiple Data Collections Discussion Leader: Partha Lahiri, University of Maryland
5:45	Adjourn for the Day

Wednesday, April 19, 2006

8:30 AM	Three-Minute Madness: All Attendees May Speak
9:30	Break
10:00	Alan Karr, NISS: Principled Trade-offs between Costs and Quality Discussion Leader: Myron Katzoff, NCHS
12:00 N	Breakout Discussions Over Lunch
2:00 PM	Final Panel Discussion Breakout Discussion Leaders Workshop Organizers
3:30	Workshop Adjourns

B Workshop Participant List

Name	Organization	E-Mail
Irma Arispe	National Center for Health Statistics	iaa9@cdc.gov
Jana Asher	AAA	jasher@aaas.org
David Banks	Duke University	banks@stat.duke.edu
Nancy Bates	Bureau of the Census	nancy.a.bates@census.gov
Helene Berard	Statistics Canada	helene.berard@statcan.ca
Darryl Bertolucci	Federal Aviation Administration	
Jonaki Bose	Bureau of Transportation Statistics	jonaki.bose@dot.gov
Ronda Britt	National Science Foundation	rbritt@nsf.gov
Lisa Broitman	National Center for Health Statistics	lbroitman@cdc.gov
Jack Buckley	National Center for Education Statistics	jack.buckley@ed.gov
Rolfe Carlson	University of Michigan Survey Research Center	rolfec@isr.umich.edu
Carolina Casas-Cordero	University of Maryland	ccasas@survey.umd.edu
Bor-Chung Chen	Bureau of the Census	bchen@info.cenusu.gov
Shijie Chen	RTI International	schen@rti.org
MoonJung Cho	Bureau of Labor Statistics	Cho.Moon@bls.gov
Sadeq Chowdhury	NORC at the Univ. of Chicago	chowdhury-sadeq@norc.org
Lawrence Cox	National Center for Health Statistics	lcox@cdc.gov
Kathy Creighton	Bureau of the Census	kathleen.p.creighton@census.gov
Byron Crenshaw	Bureau of the Census	byron.crenshaw@census.gov
Marcie Cynamon	National Center for Health Statistics	mlc6@cdc.gov
Ramon Dickerson	NuStats	rdickerson@nustats.com
John Eltinge	Bureau of Labor Statistics	eltinge.john@bls.gov
James L. Esposito	Bureau of Labor Statistics	esposito.jim@bls.gov
Julie Gershunskaya	Bureau of Labor Statistics	gershunskaya.julie@bls.gov
Lee Giesbrecht	Bureau of Transportation Statistics	lee.giesbrecht@dot.gov
Karen Goldenberg	Bureau of Labor Statistics	goldenberg.karen@bls.gov
Robert Groves	University of Michigan Survey Research Center	grovesb@umich.edu
Brian Harris-Kojetin	Office of Management and Budget	bharrisk@omb.eop.gov
John Jankowski	National Science Foundation Science	jjankows@nsf.gov
Graham Kalton	WESTAT	grahamkalton@westat.com
Alan F. Karr	NISS	karr@niss.org
Myron Katzoff	National Center for Health Statistics	mjk5@cdc.gov
Meena Khare	National Center for Health Statistics	mkhare@cdc.gov
Karol Krotki	RTI International	kkrotki@rti.org
Fran cois Laflamme	Statistics Canada	francois.laflamme@statcan.ca
Guy Laflamme	Statistics Canada	guy.laflamme@statcan.ca

Name	Organization	E-Mail
Partha Lahiri	JPSM, University of Maryland	plahiri@survey.umd.edu
Brenda LaRochelle	National Center for Health Statistics	blarochelle@cdc.gov
Michael Last	NISS	mlast@niss.org
Thomas Leckey	Department of Energy	thomas.leckey@eia.doe.gov
Judith Lessler	Chatham Research Consultancy	lesslerjud@msn.com
Yan Li	JPSM, University of Maryland	yanli2002@gmail.com
Benmei Liu	JPSM, University of Maryland	bliu@survey.umd.edu
Edith McArthur	National Center for Education Statistics	edith.mcarthur@ed.gov
Ronald Meeks	National Science Foundation	rmeeks@nsf.gov
Peter Meyer	National Center for Health Statistics	pmeyer1@cdc.gov
Susan Mitchell	RTI International	smitchell@rti.org
William Mosher	National Center for Health Statistics	wmosher@cdc.gov
Gail Mulligan	National Center for Education Statistics	gail.mulligan@ed.gov
Kathy O'Connor	National Center for Health Statistics	kdo7@cdc.gov
Brian O'Hara	Bureau of the Census	brian.j.ohara@census.gov
Eugene Owen	National Center for Education Statistics	eugene.owen@ed.gov
Jennifer Park	National Center for Education Statistics	jennifer.park@ed.gov
Van Parsons	National Center for Health Statistics	vparsons@cdc.gov
Santanu Pramanik	JPSM, University of Maryland	spramanik@survey.umd.edu
Louis Rizzo	Westat	lourizzo@westat.com
Jay Ryan	Bureau of Labor Statistics	ryan.jay@bls.gov
Adam Safir	RTI International	asafir@rti.org
Beth Schlein	Bureau of the Census	beth.s.schlein@census.gov
Marilyn Seastrom	National Center for Education Statistics	marilyn.seastrom@ed.gov
Iris Shimizu	National Center for Health Statistics	ims1@cdc.gov
Anne Stratton	National Center for Health Statistics	astratton@cdc.gov
Yves Thibaudeau	Bureau of the Census	yves.thibaudeau@census.gov
Elizabeth Tighe	Brandeis University	tighe@brandeis.edu
Bac Tran	Bureau of the Census	bac.tran@census.gov
John Tsapogas	National Science Foundation	jtsapoga@nsf.gov
Clyde Tucker	Bureau of Labor Statistics	tucker.clyde@bls.gov
Alan Tupek	Bureau of the Census	alan.r.tupek@census.gov
Akhil Vaish	RTI International	avaish@rti.org
Paula Weir	Energy Information Administration Dept.	paula.weir@eia.doe.gov
Andrew White	Institute of Education Sciences	andrew.white@ed.gov
Doug Williams	Williams Consulting	horseshoelane@comcast.net
Raymond Wolfe	National Science Foundation	rwolfe@nsf.gov
Abera Wouhib	National Center for Health Statistics	awouhib@cdc.gov
Tommy Wright	Bureau of the Census	twright@census.gov