NISS

National Institute of Statistical Sciences
PO Box 14006, Research Triangle Park, NC 27709–4006
Tel: 919.685.9300 FAX: 919.685.9310
www.niss.org

Survey Cost Workshop Presentations

September 21, 2006

Contents

1 Background

This document consists of presentations made at a workshop on survey costs, held on April 18–19, 2006 in Washington, DC. 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 Larry 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).

The companion document "Survey Costs: Workshop Report and White Paper," by Alan F. Karr and Michael Last, is available on the NISS web site:

http://www.niss.org/affiliates/surveycost200604/surveycost-workshop200604.html

2 Presentations

2.1 Alan Karr: Introduction

NISS

Survey Cost Workshop
Introduction and Goals

Alan F. Karr
National Institute of Statistical Sciences
karr@niss.org
April 18, 2006

Thanks To

- **NCHS:** funding
- **NCES:** facilities
- **Organizing Committee:** Larry Cox (NCHS), John Eltinge (BLS), Graham Kalton (Westat), Dan Kasprzyk (MPR), Myron Katzoff (NCHS), Partha Lahiri (University of Maryland), Judy Lessler (Chatham Research Consultancy), Marilyn Seastrom (NCES), Al Tupek (Census), Doug Williams (formerly NCHS)
- **Presenters and Discussion Leaders**
- **Attendees** (!!)

Purposes of the Workshop

- Articulate a research agenda for survey cost methodology and modeling, including
  - In-process modeling (responsive/adaptive design)
  - Leveraging multiple data collections
  - Tools for principled tradeoffs between cost and quality

Goals

- Identify
  - Fundamental problems
  - High-leverage gaps
    - Entry points begin to address them
  - Techniques from statistics and other disciplines for addressing the problems
  - Necessary and desirable collaborations

Format

- Two tutorials
  - Surveys and costs: Eltinge
  - Decision theory and simulation: Banks
- One session on each of three main topics
  - In-process modeling (responsive/adaptive design
  - Leveraging multiple data collections
  - Tools for principled tradeoffs between cost and quality
- Session consists of
  - Initial presentation
  - Discussion involving all attendees

Participation Opportunities

- Topic session discussions
- Two-Minute Madness (Wednesday, 8:30)
- Breakout discussions (Wednesday lunch)
- Final discussion (Wednesday PM)
- Breaks

Intended Final Product

- Workshop report that will stimulate
  - Additional attention and engagement, among
    - "Problem owners"
    - Academic (and other) research community
  - Research
  - Resources for research

2.2 John Eltinge: Tutorial 1: Survey Basics, Including Costs

Tutorial I:
Survey Basics, Including Costs

John L. Eltinge
Bureau of Labor Statistics
Eltinge.John@bls.gov
April 18, 2006

NISS Workshop on Survey Costs

Disclaimer:

The views expressed in this paper are those of the author and do not necessarily reflect the policies of the Bureau of Labor Statistics.

Acknowledgements: Many thanks to David Chapman, Larry Cox, Pat Getz, Bob Groves, Rachel Harter, Steve Haslett, Larry Huff, Sylvia Leaver, Judy Lessler, Bill Mockovak, Steve Pedlow, Joe Sedransk, Clyde Tucker, Rick Valliant, Doug Williams, Kirk Wolter and Ibrahim Yansaneh for helpful discussions

2

Overview:

I. Sample Surveys and Administrative Record Systems

II. Components of Data Quality and Risk

III. Literature on Survey Costs

IV. Two Classes of Methodological Questions

3

I. Sample Surveys and Administrative Record Systems

A. Goal of Government Statistical Agencies and Other Large Survey Organizations:

Provide the best available information on a given topic for the lowest reasonable cost

4

B. Information: Point ests, inference?
Cost: To agency? To data user?

C. Traditional View of Sample Surveys

Superpopulation model $\xi(\theta)$
generates a finite population U
of size N with characteristics
$(Y_i, X_i), i = 1, \ldots, N$

5

1. Goal: Estimation and inference for

$\gamma = g(\theta)$ superpopulation quantity

or the corresponding finite pop quantity defined through an estimating function

$$E_N(\theta_U) = \sum_{i \in U} f(Y_i, X_i; \theta_U) - v(\theta_U) = 0$$

e.g., Binder (1983, *Int. Stat. Rev.);* Scott and Wild (1986, Biometrics)

6

Examples: Finite population means, totals, quantiles, regression coefficients, parameters of a generalized linear model

Historical focus of most statistical agencies:
Simple descriptive quantities (means, totals, ratios) for large aggregates (full population or large subpopulations)

Ex: Current Employment Survey:

Estimated total employment and one-month change:

- Essentially all non-agricultural U.S. employers

- Eleven large industrial "supersectors"

7

2. Ideally, we would take a census (100% sample) of all units in U compute the desired quantities, and publish results.

3. Seven practical constraints that make (2) unfeasible:

a. Direct use of information from administrative record system not entirely feasible:

- Definitional or aggregation issues

- Diminishing returns (as measured by inferential quality) from very large sample sizes

- Constraints on processing systems

Solution: Base estimation and inference on a sample of units

8

b. Candidate frames (specification of prospective sample units): incomplete

Example: New construction

Example: Aggregation

Solution: Use multiple frames, some with nesting (area frames, list frames) and sample separately from each frame

9

c. Nested structure of population:

May not be able to identify units of interest directly from the available frames, or cost may be prohibitive

Solution: Use cluster sampling or other forms of multistage sampling

Ex: Sample counties, then neighborhoods, then houses

10

d. Subpopulation membership (possibly rare) not reflected in frame

Solution: Two-phase sampling

Large sample with cheap measures

Follow-up smaller sample of "interesting" units

Epidemiological variant: Case-control studies

11

e. Membership in rare subpopulation not reflected in frame

and

significant network structure in subpopulation membership

Example:
Wildlife sampling, some human social networks

Solution: Adaptive or network sampling

12

f. Heterogeneity across population units:

Example: Sizes of establishments

Solution: Sample units with unequal probabilities (e.g., probability proportional to size)

13

g. Heterogeneity across identifiable subpopulations:

Examples:
Industry, size class, occupation

Solution: Stratified sampling (partition into subpopulations and sample separately from each subpopulation)

14

4. Resulting complications:

a. Generally impossible or inefficient to draw a simple random sample from U

Alternative: Select a sample S of size n through a complex sample design that involves the use of one or more of:
- Stratification
- Unequal selection probabilities
- Clustering or other forms of dependent selection (two-phase, adaptive)

15

b. Consequently, observations are not iid

c. Multiple stakeholders: No uniform consensus on basis for estimation and inference

Model $\xi(\theta)$ generally not truly known and often the subject of controversy (esp. regarding appropriate conditioning)

16

3. Criteria for estimator performance:

a. At a minimum, we want good properties when performance is evaluated with respect to the sample design:

$$E_p(\hat{\theta}_S) \cong \theta_U$$

i.e., performance "in repeated sampling under the specified design"

17

b. Note minimalist approach:

i. Limited assumptions:
How we drew the sample
- Reduced (eliminated?) risk of model failure

ii. (Almost) no assumptions on population U

iii. Modest claim for performance:
wrt repeated sampling from *this* population
- Should be minimally acceptable to a wide range of stakeholders

18

c. In its most pure form, effectively ignores issues with:
- Nonresponse
- Measurement error
- Loss of efficiency (under specified model constraints)

Thus, need to introduce some amount of modeling into any serious discussion of performance, but this generally is done with considerable caution

19

d. Ideally, prefer good properties when performance is evaluated wrt either the sample design, or the underlying superpopulation model, or both

$$E_{p\xi}(\hat{\theta}_S) \cong \theta$$

as well as under moderate deviations (via sparse effect models?) from specified superpopulation model

Similarly for variance ests, inference methods
- Asymptotics usually through triangular-array type arguments: increasing N, n, conditions

20

4. **Primary approach for statistical agencies: Point estimation method through solution of weighted estimating equation:**

$$\hat{E}_n(\hat{\theta}_S) = \sum_{i \in S} w_i f(Y_i, X_i; \hat{\theta}_S) - v(\hat{\theta}_S) = 0$$

where weights w_i are proportional to the inverse of selection probabilities (with modifications for auxiliary information)

21

5. Examples:

Population total: $\hat{Y} = \sum_{i \in S} w_i Y_i$

Mean of subpopulation (domain) D:

$$\hat{\bar{Y}}_D = \left(\sum_{i \in S \cap D} w_i \right)^{-1} \sum_{i \in S \cap D} w_i Y_i$$

22

6. Justification of a given procedure (sample design, collection method and estimation method) generally involves a combination of:

a. Optimization of formal criterion (loss function, weighted likelihood function)

b. Performance evaluated with respect to:
- Sample design
- Specified model, and deviations therefrom

c. Compatibility with production systems

23

D. Related Comment on Costs and Risks Related to Modeling

1. Costs:

a. Labor for model fitting and monitoring

b. Access to, and use of, auxiliary data X (Ex: Multistate metropolitan areas)

c. Modification of production systems

d. Dissemination of results and exposition of risks for stakeholders

24

2. Risks (beyond standard measures of error)

a. Model failure: Greatest interest by stakeholders may coincide with conditions under which models may be most problematic
- Change-points in economic conditions
- Special subpopulations

b. Misinterpretation by stakeholders
- Highly exploratory data analysis, implicit multiple inference (FDR, other risk measures)

c. Reduction in perceived value for stakeholders

d. Resulting reputational risk for statistical agency

25

E. Parallel Developments on Costs and Data Quality Related to Design of:

1. Instruments

2. Fieldwork

3. Microdata review

4. Dissemination

Cf. Groves (this afternoon)

26

II. Components of Data Quality and Risk

A. Strong Links Between Perceptions of Quality and Utility

1. Varies widely across stakeholders

2. Suggestion: To structure discussion, borrow framework from literature on adoption and diffusion of technology

Ex: Rogers (1995) and references therein

27

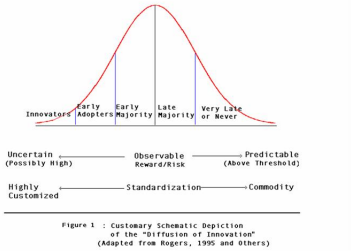


Figure 1 : Customary Schematic Depiction of the "Diffusion of Innovation" (Adapted from Rogers, 1995 and Others)

B. (Brackstone, 1999; many other variants)

Accuracy	Relevance
Timeliness	Interpretability
Accessibility	Coherence

C. Risk: Failure in one or more components of data quality

Implicitly reflect costs to some data users

29

III. Literature on Survey Costs

A. Broad Overviews

Pearson, R.W. and R.F. Boruch (1986). *Survey Research Designs: Towards a Better Understanding of Their Costs and Benefits.* New York: Springer.

Groves, R.M. (1986). *Survey Errors and Survey Costs.* New York: Wiley

United Nations Statistical Division (2005) http://unstats.un.org/unsd/hhsurveys/

30

B. Specific Case Studies:
Bibliography available

1. Tend to be very focused on one specific cost component

2. Consequently, any one study is of limited benefit for broad discussion of cost-benefit trade-offs

31

C. Important Limitations on Available Survey Cost Information

1. Large fixed costs, often not well-identified

   a. Human/intellectual capital investment
   cf. "capacity building" in UNSD (2005)

   b. Legacy systems (sample, instrument, field, production)

32

2. Aggregation effects

a. Operational constraints

b. Filters imposed by project management procedures, incentives

c. Reporting constraints

3. Side comment:
Incorporate more detailed variable cost accounting into OMB 83-I process?

33

IV. Two Classes of Methodological Questions on Survey Cost Structures and Optimization Thereof

A. Empirical Evidence on Survey Costs and Survey Efficiency

1. Gaps in current information

34

2. Extent of generalizability of available cost information

a. Global cost structures (simple dominant factors, consistent with underlying theory)
- Customary scientific ideal

b. Local cost structures (survey or module specific)

Cf. March, 2005 Workshop on Total Survey Error

35

B. Improved Methods to Optimize Survey Cost Effectiveness

1. Methods to collect and analyze cost information
("designing for cost reduction";
cf. "designing for nonresponse" or "designing for small domain estimation")

2. Characterize and quantify linkage among cost, information capacity, and data quality

36

3. Tools for cost optimization of survey procedures subject to complex and uncertain cost structures (cf. Karr – tomorrow)

Ex: Leaver (2005)
– Consumer Price Index

Ex: Adaptive sampling-based data review?

Ex: Drill-down data review

Ex: Genetic algorithms

37

4. Optimize overall procedure design, in light of:

a. Uncertain and spotty cost information (Critical question: extent to which we should condition on, or integrate over, components of uncertainty?)

b. Previously absorbed fixed costs
(cf. Lessler – this afternoon)

c. Constraints on data collection and processing that are often cost-driven (Constraints often also involve a substantial component of uncertainty.)

38

V. Summary

A. Classical sample design and randomization inference

B. Role of models

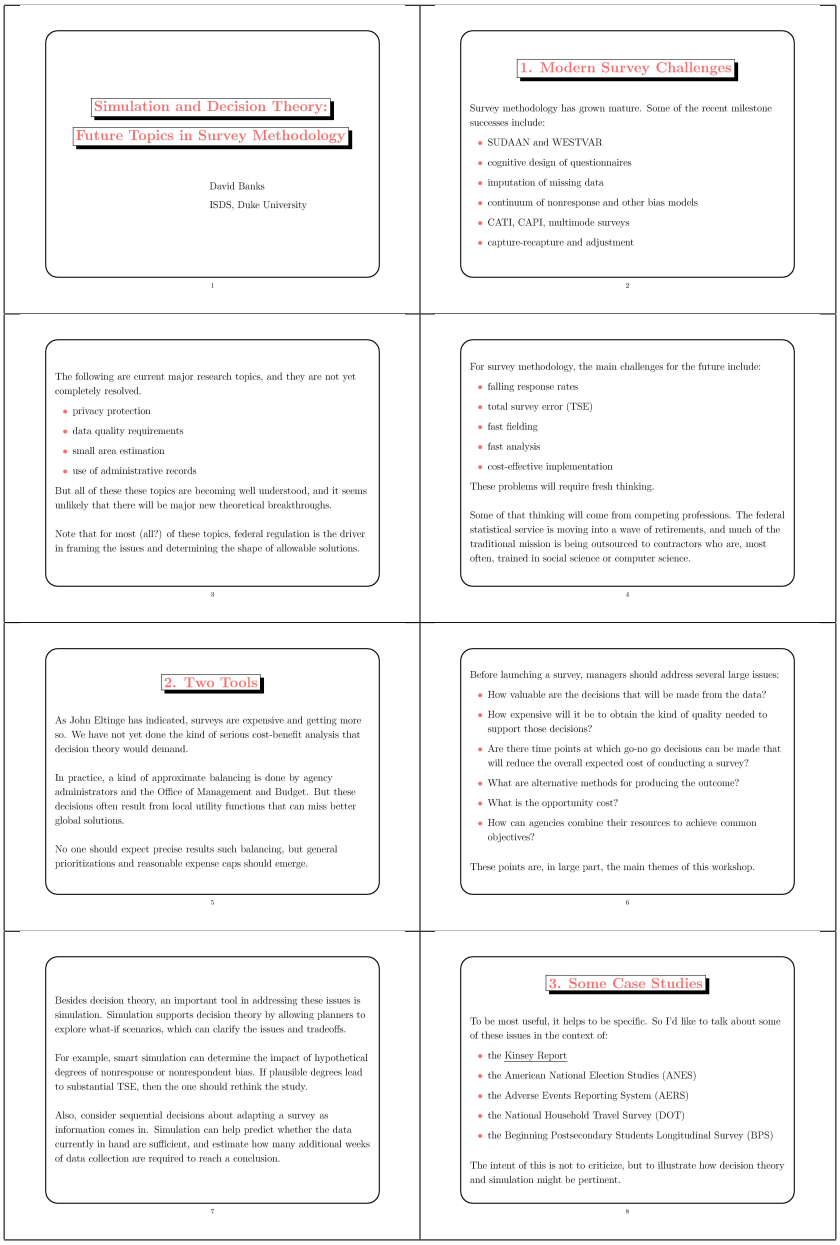
C. Components of data quality & risk

D. Previous literature on survey costs

E. Two classes of methodological questions

39

2.3 David Banks: Tutorial 2: Simulation and Decision Theory: Future Topics in Survey Methodology



Simulation and Decision Theory:
Future Topics in Survey Methodology

David Banks
ISDS, Duke University

1

1. Modern Survey Challenges

Survey methodology has grown mature. Some of the recent milestone successes include:

- SUDAAN and WESTVAR
- cognitive design of questionnaires
- imputation of missing data
- continuum of nonresponse and other bias models
- CATI, CAPI, multimode surveys
- capture-recapture and adjustment

2

The following are current major research topics, and they are not yet completely resolved.

- privacy protection
- data quality requirements
- small area estimation
- use of administrative records

But all of these these topics are becoming well understood, and it seems unlikely that there will be major new theoretical breakthroughs.

Note that for most (all?) of these topics, federal regulation is the driver in framing the issues and determining the shape of allowable solutions.

3

For survey methodology, the main challenges for the future include:

- falling response rates
- total survey error (TSE)
- fast fielding
- fast analysis
- cost-effective implementation

These problems will require fresh thinking.

Some of that thinking will come from competing professions. The federal statistical service is moving into a wave of retirements, and much of the traditional mission is being outsourced to contractors who are, most often, trained in social science or computer science.

4

2. Two Tools

As John Eltinge has indicated, surveys are expensive and getting more so. We have not yet done the kind of serious cost-benefit analysis that decision theory would demand.

In practice, a kind of approximate balancing is done by agency administrators and the Office of Management and Budget. But these decisions often result from local utility functions that can miss better global solutions.

No one should expect precise results such balancing, but general prioritizations and reasonable expense caps should emerge.

5

Before launching a survey, managers should address several large issues:

- How valuable are the decisions that will be made from the data?
- How expensive will it be to obtain the kind of quality needed to support those decisions?
- Are there time points at which go-no go decisions can be made that will reduce the overall expected cost of conducting a survey?
- What are alternative methods for producing the outcome?
- What is the opportunity cost?
- How can agencies combine their resources to achieve common objectives?

These points are, in large part, the main themes of this workshop.

6

Besides decision theory, an important tool in addressing these issues is simulation. Simulation supports decision theory by allowing planners to explore what-if scenarios, which can clarify the issues and tradeoffs.

For example, smart simulation can determine the impact of hypothetical degrees of nonresponse or nonrespondent bias. If plausible degrees lead to substantial TSE, then the one should rethink the study.

Also, consider sequential decisions about adapting a survey as information comes in. Simulation can help predict whether the data currently in hand are sufficient, and estimate how many additional weeks of data collection are required to reach a conclusion.

7

3. Some Case Studies

To be most useful, it helps to be specific. So I'd like to talk about some of these issues in the context of:

- the Kinsey Report
- the American National Election Studies (ANES)
- the Adverse Events Reporting System (AERS)
- the National Household Travel Survey (DOT)
- the Beginning Postsecondary Students Longitudinal Survey (BPS)

The intent of this is not to criticize, but to illustrate how decision theory and simulation might be pertinent.

8

3.1 Comments on the Kinsey Report

Regarding the Kinsey Report, John Tukey famously declared that "A random selection of three people would have been better than a group of 300 chosen by Mr. Kinsey."

But this needs to be judged in the context of the research. Tukey approached the problem from the perspective of wanting to estimate, say, the proportion of active homosexual males in the population. Whereas Kinsey, and to a large extent the national audience, were intrigued that a great many people were having much more interesting sex than they had ever suspected.

From the standpoint of Kinsey's research goals, his sample succeeded.

9

The Kinsey Report underscores many of the issues:

- How valuable are the decisions that will be made from the data? Very—it made us all more free.
- How expensive will it be to obtain the kind of quality needed to support those decisions? Very—we still do not have solid numbers on sex behavior. Biases are the main problem.
- Are there time points at which go-no go decisions can be made that will reduce the overall expected cost of conducting a survey? The biases would have invalidated almost any such conventional random sample in 1948.
- What are alternative methods for producing the outcome? Kinsey's self-reports, Humphrey's Tea Room Trade method, continuum of nonresponse models, eliciting expert opinion, measures of homophily, and so forth.

10

3.2 Comments on ANES

The American National Election Survey has been running since 1948. It is supported by NSF and headquartered at Michigan. The survey does face-to-face interviews with a "representative" sample, before and after the presidential election.

The questions are submitted by social scientists, and then selected for inclusion by a panel. Researchers can pay to have additional questions added.

ANES is teaming with the National Longitudinal Survey of Youth, which will begin including questions on political attitudes.

11

- How valuable are the decisions that will be made from the data? Moderate—the main direct users publish research papers. But the people to whom politicians listen read these.
- How expensive is it be to obtain the kind of quality needed to support those decisions? Moderate. Given the time, scope, and detail, and the fact that investigators can "buy" questions, this is a bargain. There is fairly high response rate, and the TSE is probably below average.
- Are there time points at which go-no go decisions can be made that will reduce the overall expected cost of conducting a survey? Not in a given year.
- What are the opportunity costs? The focus on elections is narrow, and the survey has a long four-year cycle.

12

3.3 Comments on AERS

The Adverse Events Reporting System is used to identify drug side-effects. It relies upon volunteer self-reports or physician reports. Thus the data often show the influence of high-profile media coverage (fen-phen and heart disease; ER's show on Prozac and rage).

But data mining techniques (DuMouchel, 1999) have been quite successful in finding problematic interactions. In part this is because even if a sample is only "representative," as opposed to random, one can learn a lot. In fact, for many problems non-random samples are better (e.g., the labeling problem in semisupervised classification).

Note that DOT has self-report data in DIMS, and the CPSC gets such data on fires.

13

- How valuable are the decisions that will be made from the data? Fairly important—the main problems emerge in the clinical trials, so AERS is most useful for rare problems.
- How expensive will it be to obtain the kind of quality needed to support those decisions? Very—but it would probably be worthwhile to figure out what a computerized system that draws on administrative records would cost, and account for multiple uses.
- Are there time points at which go-no go decisions can be made that will reduce the overall expected cost of conducting a survey? One might consider phasing in drugs, so that they receive less scrutiny as time passes.
- What are alternative methods for producing the outcome? Clever statistical modeling of self-report data, as done by DuMouchel, seems to extract a lot of usable signal.

14

3.4 Comments on NHTS

The National Household Travel Survey is done to study changes in travel practices. It looks at how much people travel, how they travel, and so forth. It is a mail/phone survey of about 66,000 households run by Westat and Morpace.

Some communities pay for "add-ons" which are higher-resolution samples that can be used for local road planning and other purposes.

The response rates were fairly low, the survey was burdensome, OMB worried that it was not generalizable, and the NHTS was expensive.

15

- How valuable are the decisions that will be made from the data? Probably not so much. Social scientists are interested in how our travel affects our work and personal time, but it is hard to see that there is major benefit.
- How expensive will it be to obtain the kind of quality needed to support those decisions? Very. Data quality is a real issue, since the respondents keep diaries and report over the phone. People willing to do this are not typical, so there are biases.
- Are there time points at which go-no go decisions can be made that will reduce the overall expected cost of conducting a survey? Since the survey does not feed a specific decision, this is hard to do.
- What are alternative methods for producing the outcome? Maintain a stable of respondents, let people volunteer information.

16

3.5 Comments on BPS

The NCES Beginning Post-Secondary Students Longitudinal Survey looks at the kinds of school experiences that people have after high school. It offers about $30 as an incentive; the questionnaire is a bit burdensome but not especially sensitive.

As a longitudinal survey, it poses special challenges. The main users seem to be sociologists. The results may also feed forward into policy decisions about student loan programs.

17

- How valuable are the decisions that will be made from the data? There is moderate value from a better understanding of the barriers to education and educational paths.
- How expensive will it be to obtain the kind of quality needed to support those decisions? This is an expensive survey. I suspect the biases are not large, since the population is generally amenable. But some subgroups (e.g., working mothers going back to school) have a different participation rate.
- Are there time points at which go-no go decisions can be made? Not many. As a longitudinal study, it is hard to curtail the first portion. For the second, it is possible to reallocate sample to focus on questions that have become hot.
- What are alternative methods for producing the outcome? Administrative records are the main alternative. But they do not probe the reasons why some students drop out.

18

4. Possible Solutions

As mentioned, the main challenges for the future include:

- falling response rates
- total survey error
- fast fielding
- fast analysis
- cost-effective implementation

19

A good solution will be a cocktail with many components.

First, it may be that not all of the federal surveys are truly needed. Some have a legislative requirement, some are used to determine the allocation of funds, some are done because key stakeholders think they are interesting, and some because of agency turf defense of historical roles.

This will entail macro-level decision theory, and a global grasp of how federal surveys support essential policy. We all recognize how hard this is to achieve, but somebody in federal government ought to know this.

20

Second, it may be that we should move away from random surveys and use rotating panels of respondents, chosen by quota sampling to be representative. One model for this is the Nielsen families.

The advantages include:

- One can build trust, which enables effective use of random response and tolerance of sensitive questions.
- As a longitudinal study, one reduces variance in trend estimation.
- One can change the questions in response to current events.
- One can weed out respondents who slack or lie.
- The analysis builds on previous responses (e.g., one doesn't have to endlessly capture demographic variables).
- One can ask follow-up questions.

21

The disadvantages include:

- The Magic Town effect.
- The difficulty of estimating uncertainties.

The magnitude of the first is unclear, and partially addressed by the rotation of the panel. The second can be partially addressed through half-sampling, if the panel is large enough.

I suspect that the TSE for a seasoned stable is less than is found in all but the Census.

22

A third strategy is build a large simulation model. There are 290 million people, and administrative records provide a solid anchor for much of that.

This would be insanely ambitious project, but a virtual population would allow thoughtful development of models for changes and choices, and identify survey questions that would be most illuminating (cf. the label buying problem in semisupervised learning). It would lead to a complete reinvention of agent-based modeling technology.

Of course, LexisNexis-Seisint and ChoicePoint may already have done this...

23

Fourth, and more realistically, one can build simpler simulation models of populations. The actors would have response probabilities that depend upon covariates, and survey planners could see how different assumptions about household bias, nonresponse bias, and so forth affect the inference.

Fifth, one can do fast analysis. This helps because one track response rates and trends by subgroups. If one subgroup has a low response rate but homogeneous results, then it may not affect the conclusion. But if another group has a low response rate and is sufficiently disparate that more resolution is needed, then one can adaptively allocate more recruitment effort.

Both of the analyses above can be performed at the question level, with results fed back to survey administrators.

24

Sixth, one can use dynamic programming to allot interview effort. As the survey proceeds, one can decide, at regular time points, which kinds of respondents are most needed in order to reduce the overall uncertainty.

Dynamic programming is not easy, and it would be wonderful to do even just one-step myopic optimization. Simulation would play a key role in doing that.

In short, although the problems are hard, there are solution strategies. The strategies may appear insanely ambitious, but realistic compromises will still advance the practice.

25

5. Conclusions

Classical survey methods, in government and industry, are faltering. Surveys grow more expensive, but biases are reducing our certitude for all but a few flagship efforts.

The traditional emphasis on sampling errors is a straitjacket. Data mining often does quite nicely without random samples.

Modern decision-makers need rapid information. This implies flexible real-time adaptation in surveys, for both questions and procedures.

Decision theory should determine how we allocate our time and design our instruments. Simulation is a necessary tool for applying complex decision theory.

26