STATISTICAL POLICY WORKING PAPER 35

FCSM Statistical Policy Seminar

Challenges to the Federal Statistical System in Fostering Access to Statistics

Federal Committee on Statistical Methodology

Statistical Policy Office
Office of Information and Regulatory Affairs
Office of Management and Budget

October 2004

The Federal Committee on Statistical Methodology
(October 2004)

Members

Brian A. Harris-Kojetin, Chair, Office of Management and Budget

Wendy L. Alvey, Secretary, U.S. Census Bureau

Lynda Carlson, National Science Foundation

Steven B. Cohen, Agency for Healthcare Research and Quality

Steve H. Cohen, Bureau of Labor Statistics

Lawrence H. Cox, National Center for Health Statistics

Zahava D. Doering, Smithsonian Institution

Robert E. Fay, U.S. Census Bureau

Ronald Fecso, National Science Foundation

Dennis Fixler, Bureau of Economic Analysis

Gerald Gates, U.S. Census Bureau

Barry Graubard, National Cancer Institute

Brian Greenberg, Social Security Administration

William Iwig, National Agricultural Statistics Service

Arthur Kennickell, Federal Reserve Board

Nancy J. Kirkendall, Energy Information Administration

Susan Schechter, Office of Management and Budget

Rolf R. Schmitt, Federal Highway Administration

Marilyn Seastrom, National Center for Education Statistics

Monroe G. Sirken, National Center for Health Statistics

Nancy L. Spruill, Department of Defense

Clyde Tucker, Bureau of Labor Statistics

Alan R. Tupek, U.S. Census Bureau

G. David Williamson, Centers for Disease Control and Prevention

Expert Consultant

Robert Groves, Joint Program in Survey Methodology

Table of Contents

PROGRAM

FCSM Statistical Policy Seminar
Challenges to the Federal Statistical System in Fostering Access to Statistics
November 6-7, 2002

Wednesday, November 6

Welcome and Introduction: **Katherine Wallman, OMB**

Keynote Address: Jay Hakes, Jimmy Carter Library

1. Ensuring Data Confidentiality
Organizer: **Lawrence Cox,** NCHS ; Chair: **Edward Sondik,** NCHS
Speakers:
Lawrence Cox, NCHS; **Ramesh Dandekar**, EIA
A Disclosure Limitation Method for Tabular Data That Preserves Accuracy and Ease-of-Use
Stephen H. Cohen, BLS; **Wilbur Hadden**, NCHS
Data Access: Issues with Public Use Data Files and Data Centers
Discussant:
Fritz Scheuren, NORC

2. Achieving Timeliness in a "Real Time" World
Organizer**: Edward Spar,** COPAFS; Chair**: Louis Kincannon, USCB**
Two panels of producers, suppliers and users of economic and demographic data
Panel 1 on Economic Data:
Steven Landefeld, BEA;
Forrest Williams, DOC;
Maurine Haver, Haver Analytics
Panel 2 on Demographic Data:
John Kavaliunas, USCB;
Linda Jacobsen, Claritas;
Kimball Brace, Election Data Services

3. Enhancing the Design, Access and Analytical Utility of Federal Surveys Through Coordinated Efforts Between Sponsors, Stakeholders and Data Users
Organizer: **Steven B. Cohen,** AHRQ; Chair: **Philip Fulton,** ERS
Panel Discussion:
Joel Cantor, Rutgers University;
Doris Lefkowitz, AHRQ;
Pat Doyle, USCB;
Steven B. Cohen, AHRQ;
Rick Brown, UCLA;
Jane Gentleman, NCHS

4. E-Government and New Dissemination Paradigms

Organizer: **Cathryn Dippo, BLS; Chair: Lawrence Greenfeld, BJS**

Speakers:

David Roy, Statistics Canada

How the Internet is Transforming Client and Respondent Relationships at Statistics Canada

Marshall DeBerry, BJS; Valerie Gregg, USCB; Rachael Taylor; USCB

Fed Stats: Statistical Information Dissemination in the 21st Century

Discussant:

Marjorie Blumenthal, NAS

5. Improving Data Quality

Organizers: **Nancy Kirkendall,** EIA and **David Williamson,** ATSDR; Chair: **Mary Hutzler,** EIA

Panel Discussion:

Katherine Wallman, OMB;
Nancy Kirkendall, EIA;
Cynthia Clark, USCB;
Marilyn McMillen, NCES;
Brian Greenberg, SSA

6. Preserving the Past, Linking to the Future

Organizer: **Margaret Adams,** National Archives; Chair: **Steven Landefeld,** BEA

Speakers:

Thomas Brown; NARA

The History of the Custodial Program for Electronic Records at the National Archives and Records Administration

Margaret Adams; NARA

Evolution in Access Services for Electronic Records at the National Archives and Records Administration

Kenneth Thibodeau; NARA

Building for the Future: the Electronic Records Archives Program

Discussant:

Constance Citro; CNSTAT

Thursday, November 7

7. Benefits and Stewardship of Linked Survey and Administrative Data
Organizer: **Cynthia Clark,** USCB; Chair: **Susan Grad,** SSA
Speakers:
Nancy Potok, USCB
Data Stewardship and Accountability at the Census Bureau
Pamela White, Statistics Canada
Statistics Canada Policy on Record Linkage
Howard Iams, SSA
SSA Policy Applications of Administrative Data Linked to SIPP
Discussants:
Jim Spletzer, BLS;
Olivia Blum, Israel Central Bureau of Statistics

8. Capitalizing on Technology to Enhance Survey Reporting
Organizer: **William Nicholls**, Consultant; Chair: **Lynda Carlson,** NSF
Speakers:
Jon Krosnick,, The Ohio State University
Comparing the Quality of Data Obtained from Telephoned Internet Surveys: Field and Laboratory Results
Michael Link, RTI International
The Use of Responsive Virtual Human Technology to Enhance Interview Skill Training
Discussants:
Carol House, NASS;
Al Tupek, USCB

9. Providing Adequate Technical Support
Organizer: **Clyde Tucker,** BLS; Chair: **Gary Phillips,** NCES
Panel Discussion: Providing Training and Staff at Statistical Agencies to Create Metadata, to Consult with Users About Metadata and to Consult with Users About Available Analytic Tools
Stephen Dienstfrey, Schulman, Ronca and Bucuvalas;
D.E.B. Potter, AHRQ;
Carol Hert, Syracuse University;
Patricia Doyle, USCB;
Thomas Nardone, BLS;
Samuel Highsmith, USCB

10. Providing Small Area Estimates
Organizer: **Robert Fay,** USCB; Chair: **Lois Orr,** BLS
Speakers:
John Eltinge, USCB
Small Domain Estimation for the U.S. Current Employment Statistics Program: Management Implications of Multiple Stakeholders and Multiple Constraints
Douglas Wright, SAMHSA
State Estimates of Substance Use Rates from the National Household Survey on Drug Abuse (NHSDA)
Discussant:
Graham Kalton, Westat

1:00-2:15 p.m **Luncheon Hosted by COPAFS**
Speaker: **Katherine Wallman,** OMB
What's Hot, What's Not

11. Ensuring Citizen Privacy
Organizers: **Gerald Gates,** USCB; **David Williamson,** ATSDR Chair: **Thomas Petska**, SOI
Speakers:
Lawrence Greenfeld, BJS
Data Privacy and Confidentiality Issues and the Role of the IRB
Jennifer Madans, NCHS
Oval Pegs in Round Holes: Health Survey and the Common Rule
Discussants:
John McArdle; University of Virginia ;
Marjorie Speers; AAHRPP

12. Obtaining Respondent Cooperation
Organizer: **Brian Harris-Kojetin,** OMB; Chair: **Rich Allen,** NASS
Speakers: **Ruey-Ping Lu**, EIA
Response Rates Achieved in Government Surveys: Results from an OMB Study
Nancy Bates, USCB
The 2002 Response Rate Summit: Recommendations from an Expert Panel
Adrienne Oneto, USCB
National Health Interview Survey Response Rates: Influences and Interviews
Discussants:
Robert Groves, University of Michigan;
Richard Bitzer, USCB

Keynote Address

Welcome and Introduction of Keynote Speaker Jay Hakes

Katherine K. Wallman
Office of Management and Budget

It is a special pleasure for me to welcome today's keynote speaker, Jay Hakes, who currently serves as the Director of the Jimmy Carter Presidential Library in Atlanta, Georgia. When Jay first told me that he was taking this position, I was somewhat surprised – envisioning a rather dry building filled with the records of the Carter presidency. But Jay advised me that much more is involved – and indeed, a few highlights he recently provided to me bear that out. For example, the archival materials at the library provide the foundation for an upcoming "American Experience" on PBS – a biography of Jimmy Carter that will run On November 11 and 12 – which I now plan to watch. The Museum associated with the library has just finished hosting the American Independence Road Trip with Norman Lear's copy of the Declaration of Independence. along with other great original documents from the Revolutionary War period. From September 27 to January 5, 2003, the Museum is hosting "American Originals," a collection of major original documents including the Louisiana Purchase, Edison's patent on the light bulb, the surrender documents from World War II, and the arrest warrant for Susan B. Anthony illegally voting. The exhibit also includes the Emancipation Proclamation, which has not come to the Southeast since 1949. I am confident that Jay could entertain us for the next hour ... and far more ... with vignettes from his current endeavors.

But why, you may be asking, did I suggest that Jay Hakes serve as the keynote speaker for our biennial Federal Committee on Statistical Methodology Seminar. Let me explain. As many of you know, Jay served as the presidentially appointed, Senate confirmed Administrator of the Energy Information Administration from 1993 to 2000. During that period, he was a principal spokesman on energy issues, briefing policy officials throughout the Federal government (and around the world), testifying frequently before congressional committees, and interacting regularly with news organizations. At the heart of Jay's efforts were a strong and steady commitment to making the products of EIA, and indeed the statistical system, more readily understandable by and accessible to the many policy makers and publics we serve. Thus, while Jay oversaw the development of EIA's award-winning web site, he also laid the foundation for further efforts. For example, that site has just been deemed "best site for tracking economic trends" by Time magazine. And, as a member of the Interagency Council on Statistical Policy, Jay strongly encouraged and supported the birth and maturation of FEDSTATS. Always, it seemed to me, Jay Hakes challenged his own agency, and his sister agencies, to be a bit more creative, a bit more assertive, and a bit more responsive to those who could benefit from the information we statisticians provide. His insights and his proposals always were respected – and acted favorably upon – by his colleagues around the agency heads table. We learned a great deal from Jay Hakes; we were fortunate that he was keen to serve as the head of EIA.

And so, it is with great personal and professional pleasure that I introduce Jay Hakes to challenge us as we strive to foster access to Federal statistics.

A Gift to the American People:
Victories and Challenges in Providing Web Access to Federal Statistics

Jay E. Hakes
Jimmy Carter Presidential Library

It's great to be with you today. I'd like to thank Kathy and Ed for inviting me. I'm delighted to be back with many friends and former colleagues. I am here for a reason. It's because I've always done what Kathy Wallman told me to do.

From the somewhat distant perspective of a presidential library, I'd like to repeat what I said before I left Washington. The technical competence and independent integrity of the statistical agencies contribute to the foundations of our democratic system. Whatever the future holds for our country, we need to not only retain these values, but encourage their continued development.

For those of you who are interested in what I do now, I suggested you watch the "American Experience" on PBS next Monday and Tuesday nights. They have produced a major new biography on President and Mrs. Carter. Most of the material came either directly or indirectly from the archives at the Carter Presidential Library.

I continue to be fascinated by our various national energy policies and the attempts of some to suggest their policy is the first of its kind. So I'm doing historical research on this issue in my spare time. Right now at the library we have a letter and sword sent by the King of Siam to the President of the United States. It part of a collection called "American Originals" that includes the Louisiana Purchase, the Emancipation Proclamation, and the arrest warrant for Susan B. Anthony illegally voting – all on loan from the National Archives here in Washington. When the sword was mailed from what is now modern Thailand, James Buchanan was president. By the time it arrived, Abraham Lincoln had taken office. In the letter, the King offered the President elephants to breed for national transportation needs. Lincoln responded that he wasn't sure that elephants would breed in our climate. Furthermore, he said we had committed to steam power on our rivers and rails. I think it's fair to say our national energy policies go at least as far back as Lincoln.

Well, today I've been asked to speak in a general way about access to federal statistics to kick off this conference. I can do so as a former producer of federal information at the Energy Information Administration and the Council on Statistical Policy, a current collector and sharer of presidential archives, and a frequent consumer of federal information of many kinds. Though in Atlanta, I'm only a click away from what you produce. I assure you I use it frequently.

The key word for today is "access." This is a word that's achieved great cache in today's cyber world and in government circles. Maybe even too much cache:

- Access is the name, of course, of a popular Microsoft database.

- Adobe also has software named Access, which helps the blind and visually impaired read web documents.

- The State of Indiana calls its web site "AccessIndiana." In Arizona, it's "AccessArizona." In Idaho, it's "AccessIdaho." (I think you get the picture.)

- The web offers us access to wine, access to art, as well as, first and foremost, access to information.

- We can even find web sites that help us restrict access to unwanted information. A site called "NetNanny" can help if you have this problem. (I'm not making this up.)

- Access has been perhaps the most important word in the strategic plan of several federal agencies, including my former agency the Energy Information Administration and my current agency the National Archives and Records Administration, as they attempt to utilize electronic tools to accomplish their missions. If you look at the introduction to EIA's strategic written in 1994, it was all about access.

- More recently, I should also note that in September President Bush ordered the development of an interagency disability web site. The announcement promised the site would provide people with disabilities "access to a single point to go online for Government information and resources related to disabilities." Incidentally, the word "access" is used a couple of additional times in the announcement.

The federal statistical agencies have, of course, established strong web sites to encourage use of official data some time back now. The general site, FedStats, has always promised, "direct access to statistical data on topics of your choice."

Access to federal data involves more than just maintaining good web sites. But the change brought about by the web has been revolutionary. In fact, I find myself looking at the release of the Netscape browser as a fundamental turning point in the kinds of access we can and do provide today.

In my remarks, I'd like to talk about

- What access means for federal agencies,
- Some of the obstacles we've had to overcome to provide the access we have today,
- Some of the benefits we've gotten from our efforts, and
- Where we might best devote our future efforts.

It should come as no surprise that "access" can mean different things to different people.

I would make an important distinction between access that is grudging and passive and access that is expansive and active.

Grudging access can be associated with words like "bureaucratic" and "legalistic." At its worst, it's reflected in the attitude: "If this person has actually found out we have this stuff, I guess we might have to give it to them." Unfortunately, this kind of access is still the norm in a few places. (I won't name them, but I could.)

Expansive access, on the other hand, is associated with words like customer service, finding potential customers, and public education. Customers of government services are increasingly expecting this kind of access and increasingly they're getting it.

Easy access to government information is a hallmark of a democratic society. James Madison is often quoted for his comment: "popular Government, without popular information, or the means of acquiring it, is but a Prologue to a Farce or a Tragedy; or perhaps both." In today's complex world the range of issues and choices seems to have no end. Madison's sentiments point to the modern value of easy access to information, in a manner than goes well beyond the minimal requirements of the law.

The development of modern web sites began very recently, basically in the mid-1990's. So how did we get good statistical agency web sites so quickly?

Many in this room were involved in the early efforts. But some of you may have forgotten the obstacles we faced at the time:

- First, we didn't have a lot of young employees. So, if the stereotypes were correct, we shouldn't have been very web savvy.

- Second, there wasn't much, if any money appropriated for the specific purpose of developing web sites, so we could have easily justified inaction by a lack of resources.

- Third, our regular customers weren't demanding web-access in the mid-1990's, because they didn't have modems yet.

- Fourth, there were undoubtedly a few people in government who would have been very nervous about all this information going out if they had been alert enough to figure out what was going on.

- In addition, some employees were hesitant to move quickly. Some saw a focus on the web as a distraction from their "regular work." Others were wary of making information available to the masses in a way that experts wouldn't be there to "explain it." I even remember a discussion or two about the fairness of putting information up on the web, since it would be primarily the rich who would have the equipment to use it and would, as a result, gain superior access.

Another problem, at least at EIA, was a movement in congress in the mid-90's to have statistical agencies offset the cost of data collection by selling it. The House Budget Committee for two

consecutive years provided that EIA's appropriation would be cut in half on the assumption it could earn an equivalent amount from selling its products. (I think might have been meant as a compliment, albeit an unwelcome one.) The difficulty of private web sites earning money on the web from information they can copyright suggests the futility of trying to earn substantial revenues from electronic information that can't be copyrighted. One strategy would have been to restrict electronic access to increase the revenue potential of hard copies. This is an important point, because our brethren at the OECD and in some other countries went the direction of emphasizing sales of hard copies over electronic availability. Because they were forced to make revenue off their highly priced print publications, they couldn't give much information away free on the web. This policy, in effect, choked attempts to provide excellent customer service on the web.

Fortunately, the federal statistical agencies were able to overcome the obstacles I've described. Many of our employees were or became web savvy and found the challenge fun and exciting. They foresaw the need to provide good electronic access before customers were asking for it. I remember visiting the offices of sometimes-puzzled congressional staffers around 1995 demonstrating a data-rich CD-ROM at a time they had neither CD drives nor modems.

The web, of course, leapfrogged everything, and CDs never became the major player we thought they would. But when customers were ready to use web products, we were already providing them. If agencies had waited until they were asked to provide web sites, the necessary lead times for development would have put them well behind the curve.

Fortunately, it didn't cost a lot of money to develop products for the web. As a result, the financial constraints normal for government projects were minimized. The House Appropriations Committee ignored its direction from the Budget Committee to replace appropriations with sales revenues. The most obvious result of all these developments was very good web sites – content rich, relatively integrated, well tagged for search engines, with good navigability. I might also add that these government sites are much better than most business sites, despite some myths to the contrary.

What are the Benefits?

I've done a quick count of some benefits of developing good web sites. I'm sure there are many more, but I selected ten worth mentioning:

- Most obviously, people all over the country were able to access official data in a timely manner. Since the costs of to the taxpayers for this service was low, I call this a great gift to the American people. Previously, obtaining hard copies was at best slow and at worst virtually impossible. With our new web sites, we said: "A high school student in Altoona, PA, has much data available as a cabinet secretary had five years previously." But I also think of the reporter in California who working on a story at 5 Pacific Time, after offices in Washington are closed. Now the data are still available to them because the web sites are always open. I also remember staff from the National Economic Council at a conference in Buenos Aires tapping regularly into statistical agency web sites. I assure that in the days of

hard copies in the suitcase, these data would simply have been too bulky to travel and thus ignored.

- Second, web sites facilitated communication with those close by. A recent television ad portrays two business two construct a major business deal over the web, only to find they're located across the street. Have you seen it? A lot of times we don't recognize that people just down the hall at the Labor Department, at the Justice Department and elsewhere are using our information much more than when they had to use hard copies.

- Third, web access has helped statistical agencies get credit for the work they do. In the "old days," clever repackagers would sell federal data to clients for big price tags and often neglect to mention the source of the data. This gave the impression that the data would still be produced, even if the statistical agencies went away. Now the easiest way to get federal data is directly from federal web sites. You know what? I don't feel sorry for the repackagers, because the good ones will always be able to find good ways to add value and give credit where credit is due.

- Fourth, web sites give us important feedback from customers. Software is readily available track what parts of your site people are using and what parts get less usage. For example, EIA found that people liked summaries, which encouraged it to do more of them.

- Fifth, the Fedstats gateway to federal statistical sites made it easier for the layman to find federal statistical. To order a Ford automobile, you don't have to know the plant where it was manufactured. You should be able to find federal data without knowing which agency conducted the survey.

- Sixth, electronic dissemination helped agencies meet the requirements of the Government Performance and Results Act. With GPRA, we were expected to account for the outcomes of our activities. With web sites, we could produce good evidence of high and rising usage. We could also do sample surveys more easily on the value of the data to the users.

- Seventh, government experts were able to devote their attention to higher level matters. Before web sites, highly skilled analysts were spending time faxing data tables. It's hard to imagine now why that might be necessary. Experts can spend their time answering the hard questions that better utilize their special talents.

- Eighth, Fedstats helped fend off what I regarded as an unnecessary and probably counterproductive reorganization of federal statistical agencies. Periodically, people in the congress look at bringing at least some of the statistical agencies into one super statistical agency, a sort of Department of Homeland Statistics, so the right hand will know what the left hand is doing. When reorg gained some momentum in 1997, the work on Fedstats helped demonstrate to the Congress that reorganization was largely unnecessary. The statistical agencies were already talking to each other and had solved at least some of the problems that reorg was supposed to resolve.

- Ninth, it's fun to win awards. How many times does a government agency win awards? Yet many federal sites have won awards for there web sites. The most recent was in this week's Time magazine. EIA was picked as one of the 46 best web sites for business. In fact, it was of ten sites to receive a star. The description said: "For free research on a crucial industry, try this site from the Department of Energy, which forecasts future prices and trends for oil, gas and other petroleum products. In addition to statistical tables, the EIA produces clearly written reports that spell out in plain English what the numbers mean. It also features profiles of the energy sector in various countries and regions."

- Tenth and finally, the electronic world got us used to color graphics. Color became a standard feature because it was cheap and it made our data easier to understand. Color is a habit that's hard to break. I, for one, never what to go back to the days of one-color line graphs.

What about the future?

There are still many access issues for the future:

- We have so much information on our sites that *navigation* remains a major challenge. It is still possible for a fairly savvy user to not find something that's actually there. The battle to fully integrate sites hasn't been fully won. You might say people could call a help desk. But if they don't see something, it's hard to ask for it. You have to remember another thing. Men never ask for directions.

- There is also a strong argument for going back and putting up some of "the old stuff" on the web. This project would take some resources. However, now that we've shown what we can do with the "new stuff," we might be in a better position to argue for putting up available data and reports produced "pre-Netscape."

- It is in the public interest and our institutional interest that we be as visible and accessible as possible. We're doing very well with this, if Google searches are a good indicator. If you type in the subject matter covered by the federal statistical agencies with the words "statistics" or "data," in most cases, the relevant federal agency will show up first or second in a Google search. If you're not, that's a serious matter that needs to be addressed. Are your files well tagged? Are you meeting customer needs?

In my view, access was the leading goal of the 1990's. Electronic access helped us perform our missions better and helped build an expanded customer base for our efforts. We still need some of "our best and our brightest" working to make our information even more accessible.

In my opinion, however, the statistical agencies have won enough of battle to provide access to adopt a new dominant goal for the current decade. In my view, our greatest focus now should be (plastics? No, that was 1967) -- *credibility.*

Let me read from an August op ed piece in USA Today. It's written by a journalism professor at Duke about the news profession, but I think it applies to us as well:

> At the start of each college semester, I ask my students: "What is it that a news organization has to sell?" After all the predictable answers – news, facts, information – we arrive at the only one that matters: credibility. Unless news consumers are getting the news they need, presented with fairness and balance, they will find other ways to keep abreast of current events.

I think this is even more the case for statistical agencies.

Credibility is an area where statistical agencies have always done well and have a competitive edge over other providers of information. For instance, admitting when you make a mistake is one of the most important aspects of credibility. We've already done that and need to continue this practice.

There is a lot of inaccurate information being distributed over the web, even on sites that look respectable. For instance, I wanted to do some web research on the gift of a sword by the King of Siam to the United States – the story to which I referred earlier. Several sites say the King gave the sword in recognition of Lincoln's Emancipation Proclamation. The only problem with this assertion is that the King, as I mention earlier, mailed the sword while Buchanan. It was only because of the slowness of transportation in the 1860's that Lincoln was the recipient. There's a lot of information on the web that sounds good, but is in error.

Federal statistical agencies need to differentiate themselves from other sites. They should not mislead users who can now use the data for everything from policy speeches to investment decisions. This is an issue of both substance and image. With more people using our products, we are more vulnerable if any weaknesses in our systems are seen to be the cause of public misinformation.

Since the issue of credibility isn't the major thrust of my speech, I will only list a couple of the issues involved, and you'll have to invite me back if you want to discuss them in more detail:

- Response rates and quality. We all know that it's been getting harder to get the public to respond to requests for information. We have also seen reports that energy trading companies intentionally misreported data to private, but respected firms who compile and report data. The purpose of inaccurate reporting was to manipulate energy markets. We need to attack the problems of under and misreporting very aggressively. We cannot become resigned to these problems and begin treating them as necessary evils. We have to find solutions. I will be interested to see what suggestions come out of the conference in this regard.

- Timeliness. When major decisions hinge on official data, it is unfortunate when those data don't reflect current reality. Timeliness can be a threat to quality of data, if we're rushing out shoddy information. However, timeliness is an essential part of quality. As the computer has

been used to provide access, it can be used to reduce the time it takes to process data. If our processing times are not dramatically different than they were ten years ago, they are certainly taking much too long.

- I would add another issue that might be a bit sensitive. Cabinet-level departments sometimes see data as weapons to be used in behalf of advocacy of policy positions. In general, of course, this is fine. However, as the manufacturer of the bullets, the statistical agencies might be confused with the shooter of the bullets. As a result, statistical agencies need at times to retain a respectful distance from agencies in which they are housed. The credibility of the data is too valuable to risk.

If these comments look a lot like the agenda for this conference, I would point out that I completed my outline before I saw the titles for the panels. It sounds like there may be some consensus about the strategic issues we need to address. In the early 1990's, there was a lot of high-quality federal information for which the audience was too small. Now the audience has been greatly enlarged through web access. To keep and further enlarge that audience, we need to protect the quality standards we have and raise the bar for what quality means.

Session 1.

Ensuring Data Confidentiality

A DISCLOSURE LIMITATION METHOD FOR TABULAR DATA THAT PRESERVES DATA ACCURACY AND EASE-OF-USE

Lawrence H. Cox
National Center for Health Statistics
Centers for Disease Control and Prevention
LCOX@CDC.GOV

Ramesh A. Dandekar
Energy Information Administration
Department of Energy
Ramesh.Dandekar@EIA.DOE.GOV

Abstract

Disclosure limitation in tabular data traditionally has been accomplished by subjecting cell values to any of three methods: rounding, perturbation, or complementary cell suppression. If outputs are two-dimensional tables arranged independently or hierarchically, all three methods rest on sound theory and efficient computational algorithms that can be implemented up to the level of a census or large survey. Beyond two-dimensions, the reverse is true: the close connection between mathematical theory and efficient computational algorithms breaks down and computational requirements escalate. Each method is effective for disclosure limitation in contingency (count) data wherein disclosure is associated with small cell values. For magnitude data such as sales or expenditures data, disclosure can be associated with cell values of any size, rendering rounding and perturbation ineffective or inferior to cell suppression in most situations. Unfortunately, cell suppression can create patterns of missing cell data that may destroy information important to certain users and are difficult to analyze properly by all but sophisticated users. These factors create a complicated and undesirable situation from both a statistical and policy perspective: none of the current methods assures the creation of a complete, accurate, disclosure-limited data product that is as easy to use as the original (pre-disclosure limited) data, created in a flexible manner at reasonable computational effort. We present an alternative method designed to preserve these properties. We refer to this method *controlled adjustment of tabular data*, or controlled tabular adjustment. It is a method for large-scale controlled data perturbation based on linear programming. We discuss issues of expected importance to data producers and data users and illustrate how these can be accommodated flexibly within the controlled tabular adjustment framework.

1. Introduction

In this paper, we outline a new methodology for disclosure limitation in statistical data presented in tabular form. We focus on properties and applicability of the method and omit technical details, available in Dandekar and Cox (2002). Similarly, we do not repeat definitions or review the extensive literature on statistical disclosure limitation, also available in Dandekar and Cox (2002) and elsewhere, except as needed to provide relevant context in which to examine the new method. Precise development of terminology and concepts is eschewed to favor a descriptive presentation.

A Typical Situation

A National Statistical Office (*NSO*) collects data on individual entities (persons, businesses, farms, hospitals, ...), processes the data, and releases information in the form of *statistical data products* to the public and decision makers. Traditional data products are large systems of predetermined tabulations (as from the Economic Censuses), public use or specialized microdata files (as from demographic surveys), and special tabulations. Emerging new forms of data release include tabular or analytical (e.g., regression) output from statistical data base query systems.

Statistical disclosure occurs if a third party (the *intruder*) can use released data products to associate an individual entity with either:

- a tabulation cell (in tabular data from a census or survey)
- an individual record (on a microdata file)
- a response to a query (to a statistical data base query system)

and, - can deduce or infer one or more of the entity's confidential attributes.

This has been called *attribute disclosure*. In certain situations or programs (e.g., Statistics of Income), association alone may constitute disclosure (*identity disclosure*). The NSO usually takes into account the possibility that the intruder will use *auxiliary information* (public knowledge, matching file, ...) to achieve disclosure, but often must do so without complete knowledge of sources or specifics pertaining to this (potential) information. An exception is tabular economic statistics wherein the best informed intruder is often a competitor contributing to one or more tabulations involving the *target* of the disclosure.

Achieving Disclosure

Confidential attributes are often deduced via mathematical manipulation of released data. Tabular data are organized by categorizing respondent data within *elementary tabulation cells* defined by one or more variables (e.g., Age by Race by Sex in the Current Population Survey, North American Industry Classification System (NAISC) by Metropolitan Statistical Area (MSA) in the Economic Censuses, Age by Sex by International Classification of Disease (ICD) code in national health surveys). Each elementary tabulation cell is assigned a *cell value* corresponding to a statistic of interest. For *categorical data*, the cell value equals the number of respondents in the cell. For *magnitude data*, the cell value equals the sum over all respondents in the cell of a quantity of interest (e.g., income, number of doctor visits, total quantity of a commodity shipped by a manufacturer). Cell values of elementary tabulation cells are then aggregated to produce values for tabulation cells at successively less refined levels of detail (e.g., for States, the entire United States, larger industry groupings, broader Age categories). For survey (as opposed to census) data, there may be an intermediate step at which the individual data are weighted. Because this organization–from individual data to elementary tabulation cells to more general tabulation cells–is based on addition, it can be realized mathematically as a system of linear equations.

Disclosure occurs if the intruder can work backwards from aggregated data to deduce individual respondent data. In certain cases, this can be accomplished by linear algebra. By the same token, disclosure occurs if the intruder can estimate individual respondent data to within an unacceptable narrow (*prohibited*) range (what is meant by "narrow" is determined by the statistical agency and often varies from agency to agency and sometimes from survey to survey). Narrow estimation, whenever possible, can be accomplished by linear programming.

Disclosures as above are achieved by deterministic means, so that respondent data are identified within a range. *Probabilistic disclosure* determines if, within an acceptable range, there is high probability that respondent data lie within a smaller prohibited range. Probabilistic disclosure is only beginning to be addressed in the literature and is beyond the scope of this paper.

The paper is organized as follows. Section 2 describes typical mechanisms for quantifying statistical disclosure in tabular data products. The new method is introduced in Section 3. Two questions are central in the evaluation of a disclosure limitation method. Does the method provide the required degree of disclosure limitation, that is, has it reduced the risk of disclosure to a sufficient extent? This question must first be answered in the affirmative. The second question is then: Has the method preserved important analytical properties of the data? The first question is addressed in Section 3, the second in Section 4. Section 5 provides concluding comments.

2. Quantifying and Limiting Statistical Disclosure in Tabular Data

Quantifying Disclosure

For categorical (count) data, statistical disclosure occurs when an individual can be correctly associated with a specific set of characteristics or *attributes*. The concern is that known or publicly available attributes of the respondent (e.g., sex, age category, profession, industrial classification, geographic area where a person lives or business, medical or insurance services are offered) can be used to identify the respondent in the data and from there link the subject to its confidential attributes (e.g., illegal drug use, income category, disease incidence, corporate cost, sales or employment practices information, medical insurance costing or reimbursement policies). A clear problem exists if the respondent is categorized in a tabulation cell containing only a small number of respondents, viz., the cell value is *small*. Or, further, if a small cell or cell complement can be so-identified. What is meant by "small" is determined by the policies and practices of the NSO and/or survey. For example, the U.S. Census Bureau has in the past used values such as "5" for census data and "15" for survey data. Statistics New Zealand and the Statistics of Income Program use "3".

Disclosure in categorical data is thus defined by a *threshold rule*: a cell or cell combination (or complement) is a disclosure if its value is less than a predetermined threshold value n (e.g., n = 3, 5, 15). Consequently, narrow estimation is defined to be an estimate of a cell value, computed by deterministic means such as linear programming, that is less than n. Because cells for which there are no respondents or data, and consequently have cell value equal to zero, are often well-known, *zero cells* are typically exempted from the notion of "small". Typically, the NSO makes the numeric value of n publicly available.

This rule can be expressed quantitatively in the following manner. A nonzero cell or cell combination **X** is a *primary disclosure cell* under the *n-threshold rule* if:

$$S(X) = n - m > 0$$

where m denotes the number of respondents in the tabulation cell or cell combination. The prohibited range under the n-threshold rule is thus the interval (0, n). Because inferences equal to zero or n are permitted, this is an *open interval*, viz., the endpoints are excluded. Conversely, a range estimate for a cell that strictly contains the prohibited interval must be acceptable.

For magnitude data, disclosure amounts to narrow estimation of a quantitative attribute corresponding to the respondent. For, e.g., manufacturing or business data, it is often easy to associate individual respondents to particular tabulation cells (e.g., type of good manufactured or goods sold and location of factories or retail outlets are well known). The NSO may consider this information to be publicly available. What the NSO must protect from disclosure are the quantitative attributes of the respondent (e.g., sales, cost or pricing data). Here disclosure is a bit more complex because the most likely intruder may be a competitor whose data are also contained in the cell total. It is instructive to proceed from an example from manufacturing statistics.

Assume that four companies contribute their individual Total Value of Shipments (TVS) to the Manufacturing Census, and that the respective contributions, measured in some appropriate units, are 55, 40, 3 and 2 units. The *true* cell value is therefore 55+40+3+2 = 100 units. If the cell value is published, Contributor #2 can subtract its contribution (40) from the published total (100) to infer that its largest competitor had TVS at most 60 units. This estimate is therefore accurate to within 9% of the actual contribution. If the NSO regards 9% as "too close" (and, typically, an NSO would do so), then releasing this cell value would result in disclosure (to Contributor #1 by Contributor #2).

A typical disclosure rule for magnitude data is the *p-percent rule*, illustrated above: no estimate of any respondent by another respondent can come within p-percent of the first respondent's contribution to the cell. In contrast to categorical data where the threshold *n* is made publicly known, the NSO typically keeps the value *p* confidential as an additional safeguard to confidentiality.

It results that the greatest threat to a respondent by another respondent or third party is that illustrated above: where Contributor #1 is the target and Contributor #2 is the intruder. The p-percent rule can be represented quantitatively in the following manner. A cell **X** is a disclosure under the p-percent rule if:

$$S(X) = \sum_{i=3}^{m} x_i - (p/100)x_1 > 0$$

where x_i denotes the contribution of the i-th largest respondent (ordered from largest to smallest) to cell **X**. For simplicity, we assume all respondent contributions are nonnegative. Clearly, all cells with only one or two respondents satisfy the rule.

The prohibited range for primary disclosure cell **X** under the p-percent rule follows directly from the quantitative disclosure rule, as follows. The upper endpoint of the prohibited range should be the smallest value of a (hypothetical) cell containing **X** for which the quantitative rule fails to hold. This value is precisely the cell value of **X** plus S(X). Computation of the lower endpoint of the prohibited range is more complicated, and NSOs often replace it by the cell value of **X** minus S(X).

Limiting Disclosure

There are several disclosure limitation methods available for tabular data. For convenience, we characterize these either as *perturbative* methods or *suppression* methods.

Perturbative methods modify some or all of the true cell values to make it impossible or unlikely that the intruder can narrowly estimate the original primary disclosures. *Random perturbation*, which has been practiced by NSOs in the United Kingdom, amounts to adding or subtracting a small randomly determined integer value (possibly zero) to original cell values. In this way, the intruder cannot with certainty conclude that a published small value corresponds to a true small value. The NSO may or may not make the perturbation values and/or the perturbation probabilities publicly known.

Rounding is a form of perturbation for which all cell values are rounded either down or up to an adjacent multiple of some rounding base B (under the n-threshold rule, B = n). In this way, the intruder cannot with certainty conclude that a published cell value corresponds to a small original value. As B = n and as it is obvious when data have been rounded, no attempt is made to conceal the rounding base B. Random rounding is performed using a randomization method that ensures that expected values of rounded entries equal original entries. The rounding probabilities are uniquely determined, so no attempt is made to conceal them. A variant is *minimum distance rounding*, e.g., with respect to minimum sum of squared differences between rounded and original entries.

Simple conventional rounding (e.g., base B=5, round 0, 1, 2 down to 0 and round 3, 4, 5 up to 5) does not preserve additivity (e.g., $3 + 4 = 7$ but $5 + 5 \neq 5$). For one- and two-dimensional tables, random perturbation and random and minimum distance rounding can be performed in a manner that preserves additivity. This is *controlled rounding* (Cox 1987). Unfortunately, controlled rounding is not always possible in three- or higher-dimensions or for linked tables.

Complementary cell suppression is a third disclosure limitation method for tabular data. Under complementary suppression, primary disclosure cells are suppressed from publication, viz., the corresponding values are replaced by a suppression symbol, denoted **D**. Because (narrow) estimates of suppressed cell values can be obtained by manipulating aggregation equations between cell values, it is often the case that additional, nondisclosure cells, called *complementary suppressions*, must also be suppressed to prevent narrow estimation of primary disclosures. Combining two or more data categories (known as *collapsing*) can be viewed as (wholesale) complementary suppression. Complementary suppression is a complex theoretical, computational and operational undertaking.

Perturbation, rounding and suppression all are suitable disclosure limitation methods for categorical data. Because perturbation and rounding produce more useable results, these methods generally are preferred to suppression for for disclosure limitation in contingency tables. As an illustration, Figure 1 presents an original contingency table under the 5-threshold rule, alongside the table after controlled rounding and complementary cell suppression.

Perturbation and rounding in general are ineffective for disclosure limitation in magnitude data, for two reasons. First, magnitude data typically are skewed, necessitating changes of different magnitudes to individual cells. Second, perturbation and rounding are designed to introduce small changes into cell values, whereas rules like the p-percent rule often dictate larger changes (e.g., 5%-20% of cell value). Consequently, complementary cell suppression has become a defacto standard for

disclosure limitation in tabular magnitude data, despite it being difficult to perform and control, its computational demands, and its removal of useful data and thwarting statistical analysis. It is not that data producers or users like complementary suppression--there simply has been no realistic alternative.

20	11	28	2	19	80
12	12	21	3	12	60
39	11	3	20	17	90
4	1	13	20	2	40
75	35	65	45	50	270

20	10	30	0	20	80
15	10	20	5	10	60
40	10	5	20	15	90
0	5	10	20	5	40
75	35	65	45	50	270

20	***D***	28	***D***	19	80
D	12	21	***D***	12	60
39	11	***D***	20	***D***	90
D	***D***	***D***	20	***D***	40
75	35	65	45	50	270

Figure 1: *Original, Rounded, and Suppressed Two-Dimensional Contingency Table*

Figure 2 illustrates complementary cell suppression. Assume that the six cells in bold are primary disclosures. To simplify understanding, assume each primary disclosure requires protection equal to 10% of its value, viz., the prohibited range for a cell of value 200 is the open interval (180, 220). Alongside the original table is one possible suppression pattern to protect this table. In lieu of suppression symbols **D**, we provide best-possible (*exact*) interval estimates for suppressed cells. Note that, for the six primary disclosure cells, each exact interval contains the prohibited range, as required.

200	40	50	**200**	120	610
20	70	60	**100**	120	370
40	90	**250**	100	30	510
100	**150**	30	80	**150**	510
360	350	390	480	420	2000

200	***[0,60]***	50	**[180,240]**	120	610
[0,60]	70	60	**[60,120]**	120	370
40	90	**[100,280]**	100	***[0,180]***	510
60,120]	**[130,190]**	***[0,180]***	80	**[0,180**	510
360	350	390	480	420	2000

Figure 2: *Table of Magnitude Data Before and After Complementary Cell Suppression*

Complementary cell suppression leaves some data fixed but removes other data. For the naive user, the missing data appear to be removed entirely. The more sophisticated user could compute exact interval estimates for the missing data (see Figure 2) and impute the missing values based on these intervals. Indeed, some practitioners, e.g., Gordon Sande, have suggested that NSOs release the exact intervals as in Example 2 to assist all users. Sophisticated users might employ missing data strategies, e.g., the E-M algorithm, to impute the missing data. Indeed, a largely unexplored problem with cell suppression is the ability of such strategies to narrowly estimate original (confidential) values.

3. The New Method–Controlled Tabular Adjustment

Our objective is to develop a method for statistical disclosure limitation in magnitude data that preserves analytical properties of original data and offers acceptable theoretical and computational properties and performance in multi-dimensional settings. It should be an improved alternative to complementary cell suppression. An useful analogy is between controlled rounding and cell suppression in two-dimensional contingency tables. Controlled rounding can be performed optimally and efficiently in two-dimensions and produces a table "nearby" the original table devoid of missing entries. Suppression is more difficult to perform optimally and, while keeping some values fixed, removes other values. Most would agree that, for two-dimensional contingency tables, controlled rounding is an improved alterative to complementary suppression.

Our objective is to provide analogous improvement for magnitude data in two and higher dimensions. Applications can be as large as a national census or survey such as Censuses of Manufacturing or Retail Trade that contain many thousands of tabulation cells, at many levels of aggregation (viz., totals/subtotals/sub-subtotals/...../detail), and span several to many logical dimensions (viz., classification variables such as geography, NAICS, size categories, ...). Relying on heuristic methods, complementary cell suppression has been made to work in such applications since the 1970s at the U.S. Bureau of the Census and Statistics Canada but at the cost of *oversuppression* of data and patterns of missing data that can be difficult to analyze.

From the outset, it should be clear that our proposed method is NOT complementary cell suppression (**CCS**). Both methods are designed to provide disclosure limitation in tabular data. As we present our method as an improved alternative to complementary cell suppression, it is worthwhile to summarize the principal features of CCS. We focus primarily on magnitude data, that being the area most in need of an alternative to suppression.

Disclosure in tabular data is based on the risk of identifying confidential information pertaining to an individual respondent. Disclosure rules characterize this risk by labeling each tabulation cell either as a *primary disclosure cell* or not. Using the disclosure rule, each tabulation cell **X** considered for release is examined for disclosure. For categorical data, the disclosure rule might be the n-threshold rule, e.g., n = 5. For magnitude data, the disclosure rule might be the *p-percent rule*.

To characterize the disclosure risk associated with publishing primary disclosure cells, a *protection interval* $[L_X, U_X]$ is assigned to each primary cell **X**. The protection interval is computed directly from the disclosure rule and the contributor data corresponding to cell **X**. Estimates of the value of **X** breaching this interval are *disclosures*; interval estimates of the value of **X** that contain (are at least as broad as) the protection interval are *acceptable*. This characterization is important–it provides both quantification of risk and a decision rule for determining when *sufficient disclosure limitation* has been achieved. Complementary cell suppression then can be performed to achieve sufficient disclosure limitation. A simplified synopsis of complementary suppression is as follows.

Under complementary suppression, each primary disclosure cell is suppressed from publication (and replaced by a symbol **D**). The system of tabulation equations naturally defines a system of linear equations **S** among the cell values, in which the value of a cell **X** corresponds to a variable **x**. Initially, variables corresponding to the non-primary disclosure cells are replaced by their true values, so that only the primary disclosure cell values are represented by variables. Linear programming analysis can be applied to the system **S** to obtain exact interval estimates $[\min_S \{x\}, \max_S \{x\}]$ of the value of each suppressed primary disclosure cell **X**. If any of these intervals fails to contain the corresponding protection interval, then disclosure occurs. It is then necessary to suppress additional, *nondisclosure cells* until all protection intervals are contained in the corresponding exact intervals. This amounts to replacing selected true values of non-primary disclosure cells with variables until the exact interval test is met for each primary cell. We do not describe this process further, except to emphasize that it is equivalent to solving a typically large integer linear program and that the computational effort and time required to do so can be prohibitive. From the standpoint of analysis, once complementary suppression is complete, most users can only guess values of primary disclosure cells at best to within the protection limits, and, for nondisclosure cells, to within arbitrary limits.

Returning to Figure 1, after attempting complementary cell suppression in the rightmost table, exact interval estimates are given by Figure 3. Note that two of these estimates (both equal to [0, 4]) actually fail the exact interval test (because their right-hand endpoints lie in the protection interval), necessitating further disclosure analysis and complementary suppression (not shown here).

20	[8,12]	28	[1,5]	19	80
[11,15]	12	21	**[0,4]**	12	60
39	11	[1,16]	20	[4,19]	90
[1,5]	**[0,4]**	[0,15]	20	[0,15]	40
75	35	65	45	50	270

Figure 3: *Exact Interval Estimates After Complementary Cell Suppression*

We next describe the new disclosure limitation method for magnitude data, using the example provided in Figure 4. Assume that the cells in Figure 4 represented in **boldface** are the primary disclosure cells and, for ease of understanding, that the protection interval corresponding to each primary disclosure cell is the interval corresponding to +/- 10% of the true cell value **x**, viz., the interval (0.9x, 1.1x). The endpoints of a protection interval are called the lower/upper *protection limits*.

200	40	50	**200**	120	610
20	70	60	**100**	120	370
40	90	**250**	100	30	510
100	**150**	30	80	**150**	510
360	350	390	480	420	2000

Figure 4: *Table of Magnitude Data with Six Primary Disclosures (Protection Required for Each Primary Disclosure = +/- 10%* of Cell Value)

The new method is based on adjusting many and potentially all cell values in a manner that: 1) provides sufficient disclosure protection for the primary disclosure cells, 2) preserves the additive structure of the tabulations, and 3) minimizes individual adjustments and any of several sensible measures of overall adjustment towards preserving analytical properties of the data. This can be accomplished in many ways that are explored in the next section. As a starting point for introducing the new method, here we offer the following adjustment schema:

- replace the value of each primary disclosure cell with a *safe value*, viz., a value that does not represent disclosure (this is the *instantiation* step); an obvious choice is
  * a value at or beyond either of the primary cell's lower or upper protection limit
- assign nonnegative variables y^-_i, y^+_i to each non-primary cell value or total **i**
  * these variables represents potential downward/upward adjustment to the cell value
- represent the additive tabulation relationships (viz., from detail to sub-totals, sub-totals to higher-level sub-totals, ..., and ultimately to grand total) as a *system of linear equations*, denoted **S**
- augment **S** with *capacity constraints* on the cell adjustment variables **y** to ensure that values of nondisclosure cells do not change too much; sensible capacity constraints
  * constrain each **y** to be within a small percentage of the true cell value
  * constrain each **y** to be within estimated measurement error of the true cell value
- impose a linear *cost function* **c** on **S** that represents a sensible measure of overall change to the data; standard possibilities include
  * sum of absolute deviations from original values
  * average percent deviation from original values
  * sum of logarithms of 1 + deviations
- use linear programming on **S**, **c** to instantiate remaining values in a manner that
  * assures all additive tabulation relationships are preserved
  * minimizes the measure of overall change **c**

The linear program performs these tasks automatically. Linear programs are computationally efficient even for large problems. Massively large problems require specialized techniques.

The schema outlines a method for *controlled tabular adjustment* (**CTA**). CTA transforms a tabular system *with* disclosures to one *without* disclosures. The schema describes the method sufficiently for understanding the remainder of this paper. A formal mathematical statement of the CTA schema follows. Understanding this model is not required to follow the remainder of the paper.

Mathematical Model for Optimal Controlled Tabular Adjustment

Notation

i = 1, …, p: denote the p primary disclosure cells
i = p+1,…, n: denote the n-p nondisclosure cells
M = coefficient matrix of the tabular system **S**

I_i = binary (zero/one) variable denotes selection of lower/upper protection limit at which to instantiate primary disclosure cell i = 1,…, p
y_i^- = potential downward adjustment to cell value i
y_i^+ = potential upward adjustment to cell value i

$LPROTECT_i$, $UPROTECT_i$ = lower/upper deviation required to protect primary disclosure cell i = 1,…,p
* these values are derived directly from the disclosure rule and the cell contributions
LB_i, UB_i = lower/upper bound (capacity) on downward/upward change to cell i = 1,.., n
* these values are determined by analytical or data quality requirements
c_i = cost per unit change in cell i
* these values are determined by NSO policy/practice

Mixed Integer Linear Program (MILP) for CTA (simplified)

Minimize: $\sum_i c_i(y_i^- + y_i^+)$
Subject to:

For i = 1,…, n:

$\mathbf{M}(\mathbf{y}^+ - \mathbf{y}^-) = \mathbf{0}$
$0 \le y_i^- \le LB_i$
$0 \le y_i^+ \le UB_i$

For i = 1,…, p:

$y_i^- = LPROTECT_i * (1 - I_i)$
$y_i^+ = UPROTECT_i * I_i$
$y_i^-, y_i^+ \ge 0,\ I_i \in \{0, 1\}$

Figure 5 illustrates a possible controlled tabular adjustment of the table with disclosure presented in Figure 4. This solution was obtained "by-hand" and therefore is not optimal. Using the cost function equal to absolute-sum-of-deviations, viz., $c(y) = \sum_i (y^-_i + y^+_i)$, an optimal CTA is given in Figure 6.

200	40	50	**200**	120	610
20	70	60	**100**	120	370
40	90	**250**	100	30	510
100	**150**	30	80	**150**	510
360	350	390	480	420	2000

195	35	55	**220**	115	620
30	65	65	**90**	125	375
45	95	**225**	105	35	505
90	**165**	35	75	**135**	500
360	360	380	490	410	2000

Figure 5: *Table of Magnitude Data with Six Primary Disclosures, Before and After CTA*

189	36	45	**220**	120	610
22	70	56	**90**	132	370
37	81	**275**	90	27	510
110	**165**	27	73	**135**	510
358	352	403	473	414	2000

Figure 6: *Optimal Controlled Tabular Adjustment of Figure 4 With Respect to Minimum Sum-of-Absolute-Deviations*

The sum-of-absolute deviations in Figure 5 equals 240; the optimal value, from Figure 6, equals 198. For simplicity, no capacity constraints were imposed. There are many adjustments with this optimal cost. A different cost function can produce a different optimal solution. In the next section we argue that, for practical purposes, there is little discernible difference between two adjustments like those in Figures 5 and 6.

The mathematical model describes a *mixed integer linear program* (**MILP**) because the variables I_i are binary integers. Integer programs are very hard to solve efficiently, except for small problems. In general, we do not recommend the pursuit of an optimal MILP solution. Instead, the use of *heuristic* methods

to instantiate the primary disclosure cell values is recommended. This reduces the problem to linear programming. Heuristics are discussed in the next section and in detail in Dandekar and Cox (2002). Comparisons with optimal solutions are made in Cox and Kelly (2003).

In summary, controlled tabular adjustment, produces a system of tabular cell values that

- is additive to all sub-totals and totals
- for nondisclosure cells, the instantiated values
  * are close to original values individually
  * minimize an overall measure of deviation from true values
- for primary disclosure cells, the instantiated values
  * do not represent disclosure
  * are better than what the user gets under **CCS**
- is as easy to analyze as original data

This new disclosure limitation methodology

- is computational efficient
- can be repeated many times using different constraints and costs to simulate/examine a range of releasable data tabular products
- consequently, can be run, examined, and fine-tuned to specific survey conditions by NSO subject-matter analysts
- obviates the need for complementary cell suppression

Whereas complementary cell suppression is a *turn-key system* in that it allows little interaction by subject analyst, controlled tabular adjustment is more of an *expert system* or *expert assistant* (such as in medical diagnosis or architectural design) to augment the capabilities of the subject analyst. In the next section we examine some of the potential pros and cons of this new method and its potential for preserving analytical properties of the original data.

4. Properties of the CTA Method and Data Analysis Issues

This discussion is organized around questions that naturally arise.

Each disclosure primary cell is instantiated with a value at or near its lower or upper protection limit. Is this easy to do? Does how this is done make any difference?

As discussed in Section 3, instantiating the primary cells optimally requires solving a mixed integer linear program. This is computationally demanding for small problems and impossible for large problems. The use of heuristics for the instantiation is indicated. Random instantiation of the primary cells can be done quite easily. Unfortunately, experience (Cox and Kelly 2003) demonstrates that random solutions tend not to be close to optimal. However, computing, say 100 randomly instantiated solutions and choosing the best one often works well.

Other heuristic approaches include ordering the primary cells from largest to smallest value and assigning the lower/upper deviation in alternating fashion. More are emerging.

It is important to note that the meaning of optimality in this context is less clear than for example for mathematical optimization problems based on an actual dollar cost. Consider Figures 5 and 6. Is there really a meaningful difference between the two adjustments? In the literature and among practitioners, there is no consensus on the form of "best" cost function would take (e.g., minimize total absolute deviations, or minimize total percent deviation). Whereas an optimal solution establishes a gold standard mathematically, it cannot incorporate all the subjective information an analyst might incorporate. We expect that the ability to produce a variety of near-optimal solutions for analyst review and refinement will be seen as more valuable than exhibiting a mathematical optimum.

Primary disclosure cells may be changed quite a bit. Won't this bias data analysis?

Certainly changes other than small changes to a cell value biases that value and enough changes of this magnitude can bias analysis of a subdomain or the entire data set. Changes to primary disclosure cell values are determined by the disclosure rule and the cell data, and percent deviation will vary from cell to cell and survey to survey. Under typical NSO scenarios, the percent deviations are likely to be in 0% to 15-20% range. Changes at the upper end of this range certainly are liable to create bias. Empirical studies have shown that, without further attention to this issue, a small bias is introduced in the regression of instantiated values on original values. A worst case is would be if every primary disclosure value were adjusted upwards by a fixed percentage p, for then the regression coefficient would equal $(1 + p/100)$. But, under this scenario, correlation would equal one. Empirical studies demonstrate small change in correlation among instantiated and original primary disclosure values.

As the only alternative to CTA for disclosure limitation is complementary cell suppression, it is appropriate to assess the effects on analysis of CTA in comparison to those of complementary suppression. Complementary cell suppression forces the user to estimate the true value only within an interval at least as broad as the protection interval. If the user could estimate any closer value with confidence, then confidentiality would be breached. Therefore, instantiation of either the lower or the upper protection limit for each primary cell leaves the user with no more bias than suppression. Indeed, CTA provides the user with a unique value, enabling analysis by even the most unsophisticated user.

Still, this could result in bias. Closer examination reveals that the NSO can in fact release a closer value that still is safe. The user (and the intruder) have no way of knowing whether the original value was instantiated down or up from the true value. Thus, releasing a value in the protection interval provides the intruder no reliable means to obtain a narrow interval estimate the contribution of a target respondent. (An exception is single-respondent cells that must be treated separately.) The NSO could instantiate values for primary disclosure cells by random selection from values in the protection interval with respect to an appropriate distribution. This can be done with little or no bias. Because this results in smaller adjustments to primary disclosure cells, it requires smaller changes to individual cells and overall, thereby better preserving analytical properties of the data set. This approach does raise a policy issue as the perception that the NSO is releasing nearby values may be problematic.

Can CTA assure only small changes to nondisclosure cells?

The NSO can constrain changes to nondisclosure cells to be as small as desired. If solutions satisfying these requirements exist, the linear program will find them. If solutions do not exist, because this method is computationally efficient, it is then possible to either reinstantiate the primary cells and run the linear program again, relax some or all of the variable constraints and run again, or both.

It is important to note that constraints can be variable-specific, meaning that a variable for which no change is appropriate can be fixed at its original value and/or looser constraints can be assigned to unimportant/unreliable cell values.

In the typical case where the disclosure cells do not dominate the system, tightly constrained solutions should be available. A strong advantage of our approach is that all of these considerations can be expressed formally within a single linear program that in many situations can be run multiple times to represent different scenarios or desiderata.

Consider also the obverse: if it is inordinately difficult to balance protection with efficient selection of local changes in CTA, then it must be at least this difficult to obtain a pattern of complementary suppressions that is useful/tractable for analysis.

What are the likely effects of CTA on data analysis?

It is important here to acknowledge that first one must specify "which analysis": analytical scenarios and issues tend to be data-dependent. A census or survey offers myriad possibilities for analysis. Census and survey data are also subject to various sources and levels of error, whose effects on analysis are largely unknown. An approach as we have offered that minimizes or controls change at both the individual cell and overall is an important feature.

Change must be examined at three levels: for the primary disclosure cells, for the nondisclosure cells, and overall. Effects on the primary disclosure cells were discussed in an earlier subsection. These effects are no worse than for complementary suppression and, if our suggestions are followed, can be improved considerably by judicious choice of instantiations.

Nondisclosure cells are changed by only a small percentage. Empirical studies show that regressions and correlations are good. Arguably, if changes to nondisclosure cells are confined to within measurement error, then original and adjusted data are for all intents and purposes indistinguishable statistically.

In most settings, the primary disclosures are only a small part of the overall tabulations, and do not tend to dominate the larger values. This results in very small change to regressions and correlations among all cells, borne out by empirical studies.

CTA provides complete data, so analysis is as easy and simple as for original data. The ability to control change to individual cells allows analytically unique or important cells to be treated favorably. Conversely, less important cells can be allows to vary to a greater extent. If our suggestions are followed, changes to primary disclosure cells are no worse than complementary suppression, easier to deal with analytically, and may be expressed as random draws from known distributions.

The release of model-generated microdata in lieu of original data for disclosure limitation purposes has been suggested. How does this methodology relate to that?

The difficult thing to control in tabular data is the tabulation structure among the cells and cell values. Models for *synthetic microdata* based on microdata, as suggested by Rubin, do not have to contend with these issues at their typical levels of complexity. It has been suggested that synthetic microdata could be released under the *multiple imputation* paradigm by releasing multiple versions of the tabular system. For tabular data, this is likely to reverse disclosure limitation as the tightly defined cell and tabulation structure would force averages across multiple files of "synthetic tabulations" to be very near original cell values, thereby increasing risk of disclosure.

Are there other potential approaches for controlled tabular adjustment?

Doubtless other approaches will emerge. One statistical approach would be to develop algorithms for *iterative proportional fitting* in complex tabular settings. A potential drawback is that limited empirical experience indicates that predicted values tend to be closer to true values. Also, development of such algorithms for complex, multi-dimensional tabular systems may be tricky.

Linear programming, as used here, finds *extremal solutions* among all possible (*feasible*) solutions. Except for purposes of optimizing the linear cost function, there is no particular reason to favor extremal solutions. Indeed, although very efficient, there are limitations to linear programming vis a vis problems size. An approach that sought or exploited feasible solutions in general would be advantageous in these situations. Kelly et al.(2003) are exploring search algorithms for moving from feasible to better or near-optimal solutions using *Tabu search*. Direct search procedure have the additional advantage of lifting the requirement that cost functions be linear. This enables comparison of original and adjusted data based, e.g., on correlation, chi-square, etc.

5. Concluding Comments

Controlled tabular adjustment potentially offers an improved alternative to complementary cell suppression in terms of data analysis, simplicity of the theoretical model, interaction of the methodology with subject matter analysts, flexibility in use and operational/computational performance. Instantiation of values for multiple-respondent primary disclosure cells from known distributions and of nondisclosure values within measurement error would assure both confidentiality protection and consistency of analytical results.

We have offered a new approach to disclosure limitation in tabular data that enables variations and refinements to meet a wide range of survey, analytical and computational settings. It is the first step in replacing data suppressed to preserve confidentiality in tabular magnitude data released by NSOs with nonconfidential data suitable for analysis. Future research and evaluation areas for this new methodology include acceptance of synthetic data products by producers and users, good heuristics to obtain near-optimal solutions, integerization of continuous outputs for contingency tables, examination of effects on data analysis, limitations/opportunities for interaction with subject analysts, identification/development of supplementary information to improve analytical outcomes and account for bias, exploring limitations of/alternatives to linear programming solutions (e.g., nonextremal feasible solutions), and incorporation of nonlinear cost functions related to statistical analysis.

Our approach utilizes linear programming as a means to preserve additive tabular structure. Analytical properties are preserved and biased controlled to the extent possible by imposing appropriate constraints on individual cell adjustments. Optimality of the final solution in many cases is only an artifact in the sense that no meaningful difference can be discerned between optimal and near-optimal solutions, including nonextremal feasible solutions. This flexibility enables the development of other methodologies, including branch-and-bound and direct search, to perform controlled tabular adjustment. We look forward to further developments and refinements for controlled tabular adjustment.

Disclaimer

The opinions expressed herein are solely those of the authors and should not be interpreted as representing the policies or practices of the Centers for Disease Control and Prevention, the Energy Information Administration, or any other organization.

References

Cox, LH (1987). A constructive procedure for unbiased controlled rounding. *Journal of the American Statistical Association* **82**, 520-524.

Cox, LH and J Kelly (2003). An empirical study of heuristic and optimal methods for controlled tabular adjustment. Manuscript.

Dandekar, RA and LH Cox (2002). Controlled tabular adjustment: an alternative to complementary cell suppression for disclosure limitation of tabular data. Manuscript.

Kelly, J et al. (2003). Controlled tabular adjustment using tabu search. Manuscript.

Issues and Impediments to Expanding Access to Confidential Statistical Agency Data: Restricted Data and Restricted Access

Stephen H. Cohen
Bureau of Labor Statistics

Wilbur Hadden
National Center for Health Statistics

Abstract

The Federal Statistical Agencies collect a wealth of confidential economic, demographic and social data. These data are collected to meet requirements in legislation or the code of federal regulations. The agencies publish estimates from that data in various tabular forms on paper or on the Internet. However, analysts still are interested in the wealth of potential additional tabulations that are not published by agencies and in developing statistical models of the data.

Historically, responding to these interests the agencies publish micro data sets for demographic surveys, but the agencies are limited in what data can be released by requirements to protect the confidentiality of data providers and survey respondents.

More recently, agencies have created data centers. Data centers are secure sites where analysts can access confidential data in a setting that ensures the integrity of confidential micro data. Some agencies have developed routines that allow analysts to submit computer programs remotely across agency firewalls to access confidential microdata. This paper will explore the advantages and issues associated with each type of data access.

Introduction: Access to Statistical Agency Restricted Data

The Federal Statistical Agencies collect a wealth of data on America's society, economy, institutions, and environment. These data are collected to meet specific or general requirements in legislation or the code of federal regulations. The agencies publish estimates from that data in a wide variety of media and formats from specific tabular forms on paper to interactive query systems on the Internet. There is routine reporting of statistics which accumulate over time into time series monitoring trends. There are special studies of topical interest. And there are detailed analyses published in scientific journals. Initial publication could be a press release followed by bulletins that present much statistical data and analysis. Usually there are still many additional possible tabulations and analyses not published due to the lack of resources within the agency.

Outside the federal statistical agencies there are many institutions with interests in science and public policy that have resources to support tabulation and analysis of data produced by these agencies. And in a free society there is compelling interest in making data available to the public for analysis and publication. Indeed, most of the federal statistical agencies devote considerable resources to the preparation and publication of public use microdata data files (PUMS). At this point, however, the agencies encounter conflicting requirements. Data rich PUMS files contain records representing individuals or establishments. The detailed attributes on these records include some of the complex characteristics of individuals or establishments that make them unique, and thus create the possibility that someone might recognize an individual or organization in the data file. But, many of these data are collected under pledges of confidentiality. In some cases, agency employees releasing identifiable information are subject to severe legal penalties.

The agencies use statistical disclosure control techniques to protect individual identification. These techniques involve data modification or partial suppression to avoid the release of data so

detailed that individual respondents can be identified. Agencies have policies and rules governing the publication of statistical tables and analyses. For instance, in publishing total counts or amounts, agencies inspect tables to be sure that at least 3 organizations contribute to the total and that no one organization contributes more than one-half. This restriction makes it impossible for one organization to deduce a competitor's response from a published table. In publishing PUMS agencies remove obvious identifiers and use statistical disclosure control techniques to protect the identity of individual respondents. These techniques involve data modification or partial suppression, such as coding continuous amounts into categories and grouping all extreme cases into cells less than or greater than cut-off amounts.

Threats to Data Confidentiality

Modern computing power plus the information explosion has increased the vulnerability of federal statistical agency data to re-identification. Let's examine this issue in more detail.

There is an unprecedented growth in the size, detail and variety of data collections as computer technology and disk storage space become increasingly affordable. Latanya Sweeney has summarized this as a tendency to collect more, collect specifically, and collect it if you can. Although federal statistical agencies are probably less likely to respond opportunistically in the current environment, they are certainly not immune, and some of our greatest achievements of recent decades are part of this trend. For instance, in 1960 our system of economic statistics was mainly producing national estimates; now we are getting estimates for some statistics down to the county level. This is the result of increases in the size of data collections like the Current Population Survey.

An example of an increase in detail is the birth certificate. The fields on birth certificates in the mid 1900's included little beyond fields identifying the child, the child's parents, and the place of birth. There were a few demographic and medical fields for birth order, weight, and health status. Today, in addition to the basic information collected in the mid 1900's, birth records include additional information on parents such as their education and place of residence at the birth date, on the mother's health, risk factors and health care, and on the infant's health and delivery. Eight States have open vital records files and twnty-five have restricted access procedures.

An example in the private sector is storage of customer transactions in supermarkets utilizing loyalty cards. Food Marketing Institute reported in 1998 that 6 out of 10 supermarket companies collect or plan to collect detailed information on consumer purchases compared to 3 out of 10 in 1993. In terms of supermarket collection a consumer can opt out by not participating in a loyalty card program but can not opt out of mandatory government programs such as birth certificate records.

On the collection side there is no doubt that we are moving toward an environment where society could collect and store data on all persons; one of the fields added to the birth certificate is a check-off box requesting a social security number for the infant. On the access side technology is making the transfer of data very easy. In the past to view a paper record you had to travel to the record repository or have someone copy and mail the information. Computers and public use

files made data available to select individuals with programming skills and access to computer systems. Today the power of the personal computer, software and the internet permits personal data to be transmitted across the street or around the world. CD-rom and DVD technology make inexpensive storage and distribution widely available and reduce access time. Distance has been replaced by the speed of one's connection to the internet; and there is no reason to believe that this will long remain a limiting factor.

Today on the internet it is easy to identify data bases that have detailed personal information about people, companies, etc. The ability of a user to take public statistical agency datasets and link them with other easily available data limits the amount of detail that can be included on PUMS files.

Making Micro Data Available: Restricted Data, Tradition Methods, PUMS

Agencies also release public use datasets for researchers to further analyze on their own. When agencies release public use datasets for researchers to further analyze the amount of detail that can be released must be limited. Obvious identifiers such as name, address, and social security number are not released. Sensitive data elements such as annual salary are typically top coded or only reported in fairly wide bands. Geographic detail is often restricted at areas that have population totals over 100,000, 250,000, or even greater.

Economic establishment data are never released as public use datasets. Geographic identifiers on demographic and social statistics must be suppressed or aggregated to levels that limit the analysis possible. No identifiers are included on PUMS files that would enable a researcher to link agency data with other data.

Bureau of Labor Statistics (BLS) releases special CPI data sets for researchers when requested. These sets often include longitudinal prices within establishments. We have found that the interaction between variables on these datasets must also be evaluated to ensure that a knowledgeable person can not isolate out individually identifiable respondent information.

For example, the National Center for Health Statistics (NCHS) is pleased to offer downloadable public-use data files through the Centers for Disease Control and Prevention's (CDC)/FTP file server. The web site offers the following documentation of downloading PUMS files:

> Users of this service have access to **data sets**, **documentation**, and **questionnaires** from NCHS surveys and data collection systems. Downloading instructions are available in "readme" files.
> Public-use data files are prepared and disseminated to provide access to the full scope of the data. This allows researchers to manipulate the data in a format appropriate for their analyses. NCHS makes every effort to **release data** collected through its surveys and data systems in a timely manner. Descriptions of NCHS data systems and activities are found in the section **Surveys and Data Collection Systems**. Public-use data files that are not listed below can be obtained through other sources. Ordering instructions and the various formats available (e.g., CD-ROM and data tape) are provided in the **Electronic Products** web pages. Users of NCHS public-use data files must comply with **data use restrictions**

to ensure that the information will be used solely for statistical analysis or reporting purposes.

Since the statistical agencies can only produce a limited amount of potential outputs, the full potential of these data are not realized. One way of satisfying both concerns, the desire of researchers to have access to such files and the desire to prevent disclosures, is for the agency or research organization to release files under highly controlled conditions. This article will explore four methods of restricted access procedures that are used to allow researchers to access confidential data:

- Licensing Agreements
- Research Fellowships and Post-Doctoral Programs
- Research Data Centers
- Remote Access.

The later two methods will be explored in detail in this paper.

Licensing Agreements

A licensing agreement is a formal agreement that permits confidential microdata to reside on a researcher's personal computer in their home institution. These agreements are formal legal documents between the agency and the host organization that specify the conditions under which the specific data set licensed may be used and the penalties for violation.

There are several common themes that run through the licensing agreements.

The principal investigator (PI) must demonstrate that the data are required for research; i.e., public use data, if it exists, are not adequate. The goals of the research that require non-public data must be stated in the application. The licensor must approve the goals of the research before the application process can proceed.

License agreements specify which people in the licensee's institution must sign the form. For an academic department it is typically a Dean and not the department chairman.

The agreement also includes a statement concerning which law(s) protects the data (e.g., Privacy Act of 1974). The PI must supply a list of names of people who will be authorized to use the data. Those people must be informed of their responsibility not to share the data with people outside the group. The PI must indicate the group's experience, if any, with handling other licensed datasets.

A data security program must be developed and implemented. The licensee's institution must allow inspections of the area where the data are used and stored.

Inspections of licensee's institution are used to enforce the data security program and access restrictions. The inspections can be unannounced. Penalties for violations of aspects of the agreement are listed on the form (e.g., denial of use of other data from the licensor, fines, prison

terms, etc.). There is a requirement that no attempt will be made to determine the identity of respondents. In general, the licensee is not allowed to link the licensed data to other microdata files.

Articles, reports, and statistical summaries generated from the data must be reviewed by the agency before they are published or otherwise communicated. The results must adhere to the agency's disclosure limitation practices (e.g., all non-zero cells in a publicly released table must represent some minimum number of respondents).

Some examples of datasets released under licensing agreements include: National Center for Educational Statistics (NCES)'s Schools and Staffing Survey and The Early Childhood Longitudinal Study; BLS's Census of Fatal Occupational Injuries and The National Longitudinal Survey of Youth; and National Science Foundation (NSF)'s Survey of Doctorate Recipients and Survey of Earned Doctorates.

To date, statistical agencies have found no flagrant violations of the licensing agreements that would warrant requesting the U. S. Department of Justice (DOJ) to prosecute an individual. The question to ask is: Would DOJ consider a confidentiality breach a serious enough offense to prosecute? If not, what message would we be sending to our respondents about the seriousness of the stewardess of the data entrusted to us?

Fellowships and Post Doctoral Programs in Principal Statistical Agencies

Research Fellowships and post-doctoral programs provide unique opportunities for researchers to address some of the complex methodological problems and analytic issues relevant to agency's programs. Fellows and Post-doctoral candidates conduct research in residence at an agency, use agency data and facilities, and interact with agency staff. They adhere to the same confidentiality agreements as regular employees.

Research fellows have to have a recognized research record and considerable expertise in their area of proposed research. The American Statistical Association (ASA) administers the ASA/NSF Research Fellowship Programs, with some support from the NSF for three Federal statistical agencies: the Bureau of the Census (BOC), the BLS, and the NCES. The ASA also administers a Research Fellowship Program for the NCHS and the Bureau of Economic Analysis (BEA).

Restricted Data Access: Research Data Centers

Research Data Centers (RDCs) are secure facilities designed to provide outside researchers access to confidential microdata files. Initially these facilities have been located only at an agency's headquarters. After gaining sufficient experience with these centers agencies may expand them to additional locations. The BOC, for instance, has expanded its RDC program to various sites around the country. RDCs are both physically and electronically separated from agency's central data stores and routine operations.

After an agency has decided to create a center by gaining agreement from within and outside, decisions have to be made about which data will be made available for access. These decisions include the survey files that will be available for analysis and the data elements collected that will be made available. Some files, such as Internal Revenue Service tax files, may be considered too sensitive to allow non-agency personnel access. Permissions may need to be need to be obtained from survey sponsors (some of which may be in other government departments), providers of administrative data underlying the agency's programs, and possibly higher levels within the agency's Department (such as departmental legal offices). Files should have adequate documentation on definitions, data fields, etc.

The specific details that make RDCs possible varies from agency to agency subject to the legal protections of data. Access to certain sensitive identifiers such as name, address, social security number may not be allowed. Outside researchers might have conditions placed on use that are more restrictive than internal staff. The BOC has authority to make researchers special sworn employees, which subjects them to the same penalties as agency employees for confidentiality breaches. Other agencies do not have this authority and must, as a result, be more restrictive in making data available. Agencies might restrict access for the sake of research only or to projects that generate specific benefits to the agency's programs; this is one of the requirements at the BOC, but not at NCHS.

In choosing site locations care must be exercised to ensure that the selection process is fair. Solicitation announcements should be made in the federal register in addition to distribution to likely candidate organizations. It might be advisable choose the sites with a partner such as the NSF as the US BOC did. The evaluation process should be fair and objective. As RDCs impose considerable costs on the agency, and the agency must decide which options to use to recover the costs associated with RDCs. Costs can be recovered by charging researchers directly or charging the host organizations which can recover their costs by charging laboratory fees. The BOC and the NCHS charge researchers directly at headquarters. BOC charges hosts for remote sites.

The RDCs must be secure facilities not only physically but also procedurally. All materials researchers remove from the facilities must be reviewed for confidentiality. The computer facilities must have no network or internet links to or from the outside and the "A" drive and/or other write media disabled. The site must have an on-site employee or contractor who is trained in security and the datasets.

The NCHS has as RDC only at its headquarters while BOC has remote locations in addition to its Washington, D.C. headquarters. The NCHS RDC is a secure monitored facility where

external researchers may be allowed access to internal restricted data files for approved projects. Restricted data files are those that contain information, such as lower levels of geography (e.g., state, county, or Census tract), but do not contain direct identifiers (e.g., name or social security number). Restricted data files may be used in the RDC by researchers wishing to control for geographic area in their models or they may be used to merge additional data onto the NCHS collected data files for enhanced analyses (e.g. The NCHS contextual data file.) To gain access to the NCHS RDC researchers must follow the strict procedures that govern the use of the RDC:

- researchers must submit a research proposal
- no materials may be brought into the RDC
- no materials, printed or electronic may leave the RDC without a disclosure review
- researchers must sign a Researcher Affidavit of Confidentiality
- the RDC is open only when staff are available for supervision
- use of the RDC is subject to space availability, consistency with the NCHS mission and
- the feasibility of the proposed project.

Except for very unusual circumstances, researchers are not allowed access to files with direct geographic identifiers. Should a researcher request an NCHS data file merged with external data, RDC staff will merge the files then remove the geographic identifiers leaving the researcher access to a files that consists of the NCHS data merged with the additional data. Should the researcher need clustering variables to stratify on geography, RDC staff will construct a set of dummy geographic indicators.

Expanding the number of research data centers beyond agency headquarters has been limited by the expense of developing and maintaining a center and by the difficulty of meeting confidentiality restrictions. Even recognizing that user fees might recover certain costs, everything isn't recouped. There are non-center costs of developing survey documentation, creating center files, training staff on file structure and data limitations, replacing on-site staff, maintaining equipment, etc. And there are issues in management and organization. For instance, NCHS' confidentiality law forbids the public release of confidential data and thus requires that an RDC be staffed by Center employees. Regardless of the staffing, an authority structure has to be created that maintains and enforces agencies' culture of confidentiality.

Restricted Data Access: Remote Access

For many researchers, working at an RDC is a burden because of travel away from his/her host institution. Remote access overcomes, almost, the expense and inconvenience of distance. With remote access researchers outside the statistical agency submit analytical programs through e-mail or the internet to an RDC to run on RDC computers storing confidential microdata files. Here, too, many decisions need to be made. Decisions need to made on the languages that will be supported, medium to be used to submit the programs and review procedures for the output generated. Usually, remote access in not a method that can produce tabulations not previously released.

At NCHS SAS was chosen as the analytic language because it is in wide use and is sufficiently well structured that an automated scanning system could be used. A number of functions

available in SAS have been disabled because they are capable of producing output that present an unreasonable risk of disclosure. These commands might result in a case listing or produce unstructured output that cannot be inspected by the system. The current NCHS remote access system operates by e-mail but an internet-based system is under development and testing. The internet-based system offers a user-friendlier interface and is capable of improved turn-around time.

The RDC staff will construct a dummy data file configured exactly like the real data (univariate distributions are the same, variable locations and lengths are the same, and paths are the same) that the researcher can use for developing and debugging programs prior to sending them to the remote access system. The use of the dummy data file results in fewer iterations on the remote access system thus increasing overall efficiency. The remote access system operates entirely automatically: the system scans the e-mail for arriving computer programs, validates the user, scans programs for forbidden commands, verifies that programs are not trying to access unauthorized data files and, if no problems are found, executes the program against the real data. After execution, the system scans the analytic output generated by users' program for disclosure problems. Questionable output is routed to an RDC staff person for manual resolution. Users can submit requests to the remote access system 24 hours a day although output is only returned during normal working hours because staff randomly spot check the system to ensure that the system is working properly in all respects. Generally users receive their output within a few hours after submitting their e-mail.

Issues in Making Data Available

There are various laws governing confidentiality of data in the federal statistical system. BOC, NCHS, NCES, and Bureau of Transportation Statistics (BTS) each have agency-specific laws specifying the protection of their data. These laws, as illustrated above, are not consistent with each other. Other statistical agencies are covered by more general provisions in exemption B4 of the Freedom of Information Act (FOIA), and the Trade Secrets and Privacy Acts.

Following the various laws, the various agencies have various policies. There is a lack of uniformity in policy across the agencies. Instruments such as licensing which are available to one agency are not available to another. Each agency has to develop procedures customized to their own data and their own legal environment to protect their data and to respond to requests for access. This inhibits the development of protection policies by making it more difficult for agency officials to find common ground either for discussion and policy development or for actual cooperation in the creation of institutions like RDCs. The differences in the legal context of institutions is one reason why it is that the BOC and NCHS have developed RDCs while NCES has developed licensing agreements and the BLS has limited its access program to IPA and fellowship awardees. These differences also mean that the administrative and legal means of enforcement differ across agencies. Because of this variety any one agency has less relevant legal experience and the general legal environment for protecting statistical data is more uncertain than it might be.

The variety of laws governing various statistical agencies also inhibits cooperation among statistical agencies at levels other than policy making. Some examples of this are well-known.

Agencies are prevented from sharing some data on sampling frames, for instance, with the result that one agency is unable to take advantage of advances within another agency, inconsistent data sets are created, and survey costs are increased. Agencies are also limited in their capacity to share data for research purposes. In this case the scientific community and the public are denied the benefit that might flow from linking data across agencies.

The legal restrictions on sharing data also limit the ability of the statistical agencies to share RDC resources. BOC employees or special sworn BOC agents can only view Census Bureau data. Thus if BOC data were located in another agency each RDC staff member would have to be a sworn Census agent and ensure all researchers met the BOC restrictions before gaining access to the data. With each agency having its own legal requirements, an RDC that has to maintain different procedures for different agencies becomes unwieldy.

The public is rightly concerned about the capacity to link data, but the complex legal situation does not facilitate the statistical agencies efforts to explain the risks and protections to the public. Public opinion research shows that the public is skeptical of the government's promises to protect privacy and cynically believes that there is wide-spread data sharing among agencies. The statistical system, which institutionally is committed to protecting respondents, is not getting credit for its position while the public is not getting the benefits of data sharing, of which it thinks it is bearing the costs. This is a lose-lose situation.

The public is not alone in its concern that the confidentiality of statistical data is increasingly threatened. This conference is evidence of concern within the statistical community. As mentioned above the threats to confidentiality are increasing. These threats, however, are but dimly perceived. There has not been very much research focused on the resources available to someone attempting to reidentify entities on PUMS. Even the elementary strategies a data intruder might employ have only been superficially explored. These studies have shown that certain data sets do have limited vulnerability, and that there are data resources that an intruder might use. That is, demonstration projects have shown that in certain files persons targeted because they had met rare criteria might be identified through matching these rare criteria in other publicly available data sets. These studies suggest that there is a need to review and catalog the growing accumulations of data and evaluate them from the perspective of their potential value to a data intruder.

Efforts of the Federal statistical system on detecting a fixed disclosure risk are ad hoc. Problems are fixed as issues are raised. However, there are not many efforts by the agencies in the statistical system to systematically test their PUMS against as many data sets as publicly available as research for identification risk.

For example, recently the BLS was concerned that the National Longitudinal Public Use Datasets were vulnerable for reindentification using birth records. BLS contracted with a researcher to see if he could identify individual respondent data from birth records. The researcher used Massachusetts records along with birth information on the file to verify that with considerable expense it is possible to reidentify some records. BLS decided to suppress detailed birthdate information to ensure adequate protection to the data. However, we need a program to

study all the variables with all publically available datasets to ensure no undetected problems exist.

Research into the vulnerability of published data to reidentification will also support a growing stream of research into techniques of disclosure limitation. The purpose of this stream is to produce techniques that statistical agencies can use to raise barriers to reidentification. This research is important because lacking proven disclosure limitiation techniques statistical agencies will be placed in the unhappy situation of having to withhold data sets from surveys that once were published. That is, rather than continue to expand the public availability of data, agencies will have to retrench and put more of their data under access restrictions such as RDCs or remote access.

The most commonly applied techniques of disclosure limitation in microdata files, recoding schemes and data swapping, are applications of pragmatic, ad-hoc methods. Statistical research has, at this point, largely described the statistical properties of these methods. This research has also defined the problem in statistical terms and established methods for evaluating disclosure limitation techniques. With this as a foundation there is a new stream of research emerging into new methods based on statistical theory. A great deal of work needs to be done in this area, however, before this research produces results with practical application.

There is a continuous demand for more information and more detailed information. In responding to this demand agencies are exploring new ways of producing tables and publishing data using CD-ROM and internet technologies. They are also discovering some of the limitations of existing methods of disclosure limitation in published tables. This is another area in which pragmatic and ad-hoc methods have been analyzed with statistical theory and theoretically motivated methods are beginning to emerge. There is slim hope that these methods can satisfy users demands for information, but there is the greater possibility that these methods can be applied in automated systems such as remote access to restricted data and internet query systems like the Bureau of the Census American Factfinder system.

One last area where research is needed is statistical disclosure in models. Although statistical models generally are not sufficiently precise to lead to the statistical disclosure of confidential information, tables can, in fact, sometimes be expressed in statistical models that then inherit the same problems of potential disclosure inherent in the tabular form. Little research has been done on the vulnerability faced by statistical agencies on allowing researchers to publish intricate models. However, most research on restricted data involves publishing models.

For example, suppose one fits a simple regression model of a dependent variable against three independent variables where the model fit of the independent variables with a dependent variable is exceedingly high. Suppose in a population there exist only one entity with a specific set of values on those independent variables. Then it is possible via the model to determine the exact value of that dependent variable fairly closely.

Another issue with models that needs exploration is the risk to disclosure of sensitive dependent variables using readily available micro data that can be applied to the model's independent variables.

Conclusion

We have explored in this paper four methods Federal statistical agencies use to allow researchers access to confidential micro data: PUMS and restricted access methods. These methods have been devised to allow researchers access to the richness of statistical agency data for further analysis than the agencies can do themselves. It also opens up possibilities for re-analysis using a different approach. That builds up credibility for analysis performed by the statistical agencies.

PUMS have been produced by the agencies for demographic statistics for years. However, the richness of data found on the Internet has shown us the vulnerability of re-identification is a real threat. Ad hoc adjustments have been made. However, we need to consider a systematic review of all PUMS by all agencies producing them for disclosure risk. PUMS for economic data is not a viable due to our inability to minimize disclosure risk while providing a useful file for analysis.

The power of the PC and Internet has allowed statistical agencies the ability to set up restricted access procedures: either remote data centers or remote access. However, these efforts are done by each statistical agency independently. We need to consider setting up a one-stop shopping RDC for access to sensitive research files like FEDSTATS for published series. This will require confidentiality legislation that will give the statistical agencies uniform laws to grant special sworn status to their data. Here too much work is needed. Models proposed by researchers to be published are usually assumed safe and not given a lot of disclosure review. Are they really safe?

References

Doyle, Lane, Theeuwes, Zayatz (2001), Confidentiality, Disclosure and Data Access—Theory and Practical Applications for Statistical Agencies, North-Holland

Massell, Paul, Overview of Data Licensing Agreements at U.S. Government Agencies and Research Organizations, CDAC Paper

Jabine, Thomas B.(1993), "Procedures for Restricted Data Access," *J. Official Statistics*, vol. 9, no. 2, pp. 537-589.

Massell, Paul B.(1999), "Review of Data Licensing Agreements at U.S. Government Agencies and Research Organizations," paper presented at the Workshop on Confidentiality of and Access to Research Data Files, sponsored by the Committee on National Statistics (CNSTAT), Washington, D.C.

Massell, Paul B., Laura Zayatz, (2000), "Data Licensing Agreements at U.S. Government Agencies and Research Organizations," *Proceedings of ICES-II (International Conference on Establishment Surveys).*

George.T. Duncan, Thomas B. Jabine, Virginia A. de Wolf (eds.), *Private Lives and Public Policies,* National Academy Press (1993), in "Chapter 6 : Technical and Administrative Procedures," pp. 141-179.

Jabine, Thomas B., "Procedures for Restricted Use Access." *Journal of Official Statistics*, 9:2, 1993, pp. 537-589.

National Center for Education Statistics, "Restricted Use Data Procedures Manual."

Reznek, Arnold., Joyce. Cooper, and J. Bradford Jensen. "Increasing Access to Longitudinal Survey Microdata: the Census Bureau's Research Data Center Program." *American Statistical Association 1997 Proceedings of the Section on Government Statistics and Section on Social Statistics.* Alexandria, VA, 1997, pp. 243-248.

Sweeney, Latanya, Information Explosion in Doyle, Lane, Theeuwes, Zayatz (2001), Confidentiality, Disclosure and Data Access—Theory and Practical Applications for Statistical Agencies, North-Holland. Pp 43-74

Alonso, William and Paul Starr (eds). 1987. "The Politics of Numbers." New York: Russell Sage Foundation.

Singer, Eleanor, Public perceptions of confidentiality and attitudes toward data sharing by federal agencies in Doyle, Lane, Theeuwes, Zayatz (2001), Confidentiality, Disclosure and Data Access—Theory and Practical Applications for Statistical Agencies, North-Holland. Pp 341-70

Greenia, Nick, J Bradform Jensen and Julia Lane, Business perceptions of confidentiality, in Doyle, Lane, Theeuwes, Zayatz (2001), Confidentiality, Disclosure and Data Access—Theory and Practical Applications for Statistical Agencies, North-Holland. Pp 395-429

Session 2

Achieving Timeliness in a “Real Time” World

Achieving Timeliness in a "Real Time" World

Charles Louis Kincannon, Director,
U.S. Census Bureau

Introductory Remarks

Data collectors, data disseminators, and data users contributed unique observations regarding the timeliness of data from the Federal statistical agencies. In particular, representatives from the data user community provided useful perspective of stakeholders' needs and uses for economic and demographic data. The obvious tension or inconsistency between the expectations of timeliness in a "real time" world and the requirements for accuracy and relevancy was immediately apparent. The "cost" to achieve the government standards of accuracy it appears to many, is often, timeliness.

Whether this is a tension or an inconsistency should be explored further. In many cases, it is an inconsistency, this means that Federal statistical agencies must do a better job with stakeholders to illuminate the entire data delivery process—from data collection to data use. However, there is often a tension between the requirements of data quality, or accuracy, and timeliness. If this tension can be resolved or overcome, data collectors, data disseminators, and data users may have to engage in a discussion of priorities. There is the possibility that the most important uses for some data may rank timeliness above restrictive measures of accuracy. Or, perhaps more likely, priorities established for processing certain data products with regard to others may need to be reassessed.

Achieving Timeliness in Real Time

John Kavaliunas, Chief, Marketing Services Office
U.S. Census Bureau

Abstract

The Census Bureau has made great strides in speeding up the process of releasing information. Not only have there been improvements in release dates, but new technology has enabled us to get the data into the hands of end users much more readily than ever before. What is on the horizon that will enable us to continue to meet rising user expectations?

Early Censuses

The results of the first census in 1790 were released as soon as the enumeration was completed, posted by the U.S. Marshals on tavern walls and in other public places. The results were sent to the President who sent a tabular statement of the results to Congress in late October, 1791. Of course there were only 3.9 million persons in the country at that time and the information collected was quite limited.

Throughout the 19th Century, additional questions were added to the census questionnaire and the size and the number of printed reports increased.

It took time to collect, tabulate, analyze and publish the increasing number of census reports. Timeliness of the data became a concern. General Francis A. Walker, Superintendent of the 1870 Census, wrote:

> So rapid are the internal changes of the country, oftentimes setting calculations at naught, so fierce and vast the growth of the Nation as a whole, that the hiatus in the statistical information at the command of the legislator, the pamphleteer, the journalist, and the social and political philosopher, becomes positively painful five or six years after the day of the census. (**Report of the Superintendent of the Ninth Census**, November 1, 1872.)

Things would get worse before they got better. By 1880, the number of census reports had risen to 22. This is a noteworthy number because it was more than all the reports printed from previous censuses combined. The first reports from the 1880 Census didn't appear till three years after the census was taken and the last reports weren't printed until 1888, prompting the Assistant Director of the Census Bureau, Frederick H. Wines, to write in 1899:

> Speed is to be greatly desired. Former censuses have required as much as nine years to complete the publication of their work, and their statistics have been to a certain extent out of date when they appeared. ("The Census of 1900," **Munsey's Magazine**, 1899).

The 20th Century

At the close of the 19th century, the introduction of new technology in the form of punch cards and electrical tabulating machines speeded up considerably the processing of decennial census results. Printed reports continued to be the mainstay of the dissemination process. However, the number of questions on the questionnaire continued to grow, as did the levels of geography for which the data were tabulated and, subsequently, the number of printed reports.

Census tracts, metropolitan districts, and the terms urban and rural first appeared in 1910.

Blocks were added in 1940; minor civil divisions and census county divisions became part of the standard geography in 1950, and data were first tabulated for block groups in 1970.

There were 1,003 individual reports in 10 different series published from the 1960 census, including the 421 reports in the report series, HC3 (*City Blocks*), which included data by block for all cities above 50,000 inhabitants, and for some 200 smaller places that had contracted for block statistics.

In the late 1960s, the Census Bureau experimented with releasing information on computer tapes and, by 1970, computer tape was a standard dissemination medium. While computer tapes speeded up the release process and much more data could be provided on tape than in a printed report, only large organizations like university research units, government agencies, and private companies could read the tapes and process the data. The public had to wait to hear about the data release and then find an organization that had the information.

The Democratization of Data

The introduction in the mid 1980s of personal computers and the adoption in 1985 of a new technology known as CD-ROM, made Census data even more accessible to the public. The widespread use of CD-ROMs to deliver 1990 Census data brought about, what then-director, Barbara Bryant, called the democratization of data.

But the revolution in information dissemination was just beginning. The real democratization of data didn't really occur until the introduction of the Internet, just a few years later. The Census Bureau launched its Internet site in May 1994. A little more than a year later, the Census Bureau announced that the Internet would be its primary means of data dissemination:

> The new dissemination plan will allow for quicker release of detailed data many people want. In the past, issuing tables and analyses in printed reports could add months to the process. And since we could only print a selection, users still might not get the data they wanted. A major advantage of this initiative is that it will allow users to receive data files on demand and to create their own reports rapidly... (CENSUS BUREAU EXPANDS ELECTRONIC DATA DISSEMINATION, Press Release dated August 9, 1995)

During the 1990s, the Census Bureau had already begun cutting back on the number of printed reports as well as the number of pages in the reports. In lieu of the traditional 200-300 page

reports, the Census Bureau began publishing short ***Briefs***, which summarized findings and included analysis, graphics and maps, but with a limited number of statistical tables. These tables were put on the Internet instead of in a report appendix. Another development was the use of Adobe Acrobat to convert reports into portable document format or *pdf*, enabling us to create web-based documents without having to go through the often lengthy printing process, which could add months to the release of the information. The planned number of printed pages from Census 2000 is about one-tenth of the 1990 census output, down from 400,000 pages to about 40,000.

Census 2000

On the one hand, the Census Bureau's release of Census 2000 information was somewhat comparable in timing to 1990. However, if we look at when the information was actually in the hands of the public, then the release of Census 2000 data is far and away the fastest ever. Technological advances such as the internet, the American FactFinder, File Transfer Protocol, and our ability to produce custom CD-ROMs enabled us to get the information to many more end users much faster than in 1990.

The number of end users of Census 2000 information is something we could not have imagined in 1992. Almost five million users visited the Census Bureau's Internet site during the month of October 2002 which, in terms of data releases, was an uneventful month. Compare this to the quarter of a million users who called or wrote the various Census Bureau call centers and regional offices in all of 1992.

Let's compare the timing of several key data releases.

Release of the Public Law Redistricting Data is mandated to be completed by April 1 of the year following the Census. While we met that deadline during both decades, it should be remembered that the Census 2000 version of the file, due to the multiple race tabulations and additional geography, is about 10 times greater than its 1990 counterpart.

Summary Tape File 1A was released on computer tape between April and August 1991, and on CD-ROM several months later (October- November 1991). Veterans of the 1990 Census data user community will remember that there were suffixes appended to the file names to indicate the geographic summary levels at which the data were provided. Summary Tape File 1A contained data down to the block-group level. Summary File 1B contained data for all 7 million blocks in the U.S. at that time and was released on computer tape in the fall of 1991. An extract version of this file on CD-ROM was finally made available in 1992, although due to recalls and other factors, data for some regions were not officially released until November of 1993 . The Census 2000 SF 1, containing data for blocks and block groups (i.e., no suffixed files) was released between June and August of 2001 with all states available on a single DVD in September 2001, some two months to 2 years earlier than its 1990 census counterparts.

Summary Tape File 3A, the first release of 1990 sample information, came out on computer tape in April and May of 1992; but the comparable CD-ROMs, all 61 of them, were not produced until that winter, with some not released until February 1993. The long-awaited 3B or ZIP Code

file was made available on tape or CD-ROM between April and June of 1993. By comparison, we released the entire SF 3 File in September of this year. The DVD is expected in late November, a half a year earlier than the comparable 1990 product. It should also be remembered that the Census 2000 version of this file is about 5 times larger than its 1990 counterpart (16,530 cells vs. 3,300).

The large 1990 Summary Tape Files 2 and 4 with detailed race, ethnic, and ancestry data, because of their complexity and size, were not produced on CD-ROM and were therefore only available to only a small number of state data centers and other groups that had the capacity to process these multi-reel files. In 2000, both files are much more accessible to end users, available to the public on the Internet through the American Factfinder, the file transfer protocol, and on CD-ROM and DVD.

Release Dates for Key Decennial Products		
Product	**1990**	**2000**
Redistricting Data	2/91-3/91	3/01
S(T)F 1	4/91-11/93	6/01-9/01
S(T)F 2	10/91-11/91	12/01-4/02
S(T)F 3	3/92-6/93	8/02-11/02
S(T)F 4	3/93-12/93	4/03-9/03(Planned)

Rising User Expectations

In a 1994 survey of users of Summary Tape Files 1 and 3 about three-fourths of respondents agreed with the statement that STF 1 on computer tape was available in a timely manner. However only 50 percent thought the STF 3 tape and CD-ROM products were available in a timely manner and more than one-third (36 percent) of CD-ROM users disagreed with that statement.

"Need to get the data out sooner. Business does not like to work with 3-year old data," wrote one survey respondent. "The quality of the product is excellent, but please try to work on release dates and delays," commented another.

So, how have users reacted to the timeliness of Census 2000 data? In a series of 12 focus groups with key customer segments in the winter of 2001-02 (that is, prior to the release of sample data), most participants said that the timing of 2000 release actually exceeded their expectations.

Nevertheless, they also said that while the release of data was considerably faster than in 1990, they wished the products could be released even sooner!

In Summary

Technological advances have shortened the time necessary for processing, but governments and society have demanded more data, more complex tabulations, and additional levels of geography. In 2000 the Census Bureau improved upon 1990 census release dates from several weeks to as much as a year or more. But perhaps more importantly, technological advances have put this information into the hands of more people than ever before.

With the timeliness of Census 2000 data releases exceeding public expectations, how do we meet the challenge of rising user expectations?

What will be the next technology to appear on the horizon? For the past two decades we've been able to quickly adapt new emerging technologies (CD-ROM and the Internet) to data dissemination. Will the Internet be even more pervasive in 2010 or should we look for still another technological advance?

Will the collection of data via the Internet improve the timely processing of information? Will real-time data collection result in real-time tabulation and dissemination?

And what about the re-engineered 2010 Census? Will annual data from the American Community Survey, available six to seven months after the end of the collection year and the collection and processing of responses to only a few basic short-form questions, make timeliness somewhat of a non-issue?

But these are topics for another presentation.

Session 3

Enhancing the Design, Access and Analytical Utility of Federal Surveys through Coordinated Efforts Between Sponsors, Stakeholders and Data Users

Influence of Sponsors, Stakeholders, and Data Users on Design, Access, and Analytical Utility of Census Bureau Demographic Surveys

Pat Doyle
U.S. Census Bureau[1]

The U.S. Census Bureau is unique among federal statistical agencies because it is simultaneously a sponsor of federal surveys it collects and a collector of survey data for other sponsors. The Census Bureau receives authorization and funding directly to carry out some programs but serves as a contractor to other federal statistical agencies in carrying out other programs. We at the Census Bureau refer to surveys sponsored by other federal agencies as reimbursable surveys, because we are reimbursed for our collection efforts much in the same way as a contractor would be.

The variations in authority and funding sources across surveys have a big influence on how the Census Bureau interacts with other federal agencies, stakeholders and data users. The nature of the interaction and the influence of these groups also varies according to the phase of the survey (design, development, administration, and dissemination), and there are different types of coordination efforts based on the relationship between the parties.

Aside from explicit coordination with agencies, stakeholders and users, there is implicit coordination that occurs as part of the budget process—either during the federal budget cycle or as part of the negotiation of the agreement governing the collection. Reimbursable projects are largely driven by sponsors' desires and budgets. Stakeholder and user input is filtered through the sponsor; and requests are honored if the sponsor agrees, funding exists, and it fits within Census policies and standards and within the goals of the survey. User/stakeholder inputs on Census-sponsored surveys are solicited in variety of forums—including conferences, user mailing lists (electronic or otherwise), and websites. Their implementation is conditioned on funding, as well as on how it fits within Census Bureau polices and standards and the survey goals.

Below, I describe the partnerships formed by the Census Bureau with a variety of government and non-government entities; the constraints faced in the development and refinement of demographic surveys; and the process through which sponsors, stakeholders, and users influence the design, access, and analytic utility of the data.

The Census Bureau and Its Partners

The Census Bureau has four different types of partners in the development and administration of surveys and censuses and in the development and dissemination of data: survey sponsors, Office of Management and Budget (OMB), Congress, and data users.

---

[1]This paper reports the results of research undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope to that given to official Census Bureau publications, and is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

Survey Sponsors: Many surveys conducted by the Census Bureau are authorized and funded through other government agencies, and the Census Bureau acts as the data collection agent (not unlike other non-government survey institutions). The funding agencies are the sponsors of the surveys and other data collection projects, and the Census Bureau works hand in hand with them to develop the survey and sample design. The Census Bureau oversees and implements the data collection and in some cases handles postcollection data processing and dissemination.

The National Crime Victimization Survey (NCVS) represents an example where the Census Bureau provides design, development, administration, and postcollection processing for the survey, based on funding provided by the Bureau of Justice Statistics. The National Health Interview Survey (NHIS) represents an example of where the Census Bureau is the data collection agent only, with the sponsor (National Center for Health Statistics) providing the sample design and selection, as well as postcollection processing and data dissemination.

These two surveys illustrate two different legal authorities under which the data are collected, which has an influence on how responsive the data sponsor can be to the influence of the sponsors, stakeholders, and users. NCVS data are collected under Title 13 and NHIS data are collected under Title 15.[2] While there are a lot of differences between the two titles, the important ones for this paper are the rules governing disclosure protection and release of data to users. All surveys conducted under Title 13 are subjected to the Census Bureau rules governing disclosure avoidance and surveys conducted under Title 15 are subject to the sponsoring agency legislation. In some cases the rules differ significantly between the Census Bureau and the sponsoring agency and in other cases they do not.

Office of Management and Budget (OMB): All surveys conducted by the Census Bureau are subject to OMB clearance and we work with OMB on a continuing basis to ensure the instruments we field do not unduly burden respondents while meeting the statistical information needs of the federal government. For general-purpose surveys sponsored directly by the Census Bureau (such as the Survey of Income and Program Participation [SIPP]), the relationship between the Census Bureau and OMB in content determination is very direct. Requests for clearance are prepared in full by the Census Bureau and submitted directly to OMB. Changes required by OMB for clearance are negotiated between Census and OMB. In addition, OMB may convene interagency working groups to debate the scope of the instrument for a particular survey and how that instrument meets (or does not meet) the agencies' needs.

Congress: In some instances, the survey or other data collection instrument is either mandated directly by Congress, or some aspects of its content are required by law (the prime example being, of course, the decennial census). In some of these cases (like the decennial census), we work directly with Congress to develop the instrument and determine the data collection project design. It is not unusual for this process to occur as part of the budget cycle (as is currently the case with the development of the American Community Survey), and the simultaneity of budget setting and survey design is often not conducive to careful, iterative instrument development.

---

[2] 13 USC Sec. 101; 15 USC Sec. 1517.

Users: User needs are preeminent in guiding content and data product development. However, the role of users as Census Bureau partners varies depending on the data collection effort and its sponsorship. Users have more direct interaction with the Census Bureau on Census-sponsored surveys; but their needs have to be weighed against budget constraints, federal government priorities for statistical information, and disclosure limitation requirements of public information. The Census Bureau solicits input from users of Census-sponsored surveys in a variety of formats—such as advisory committees, user groups, and conferences. We also welcome unsolicited comments from users and encourage them to contact us whenever they experience anomalies in Census-supplied information. All of our data products and announcements are accompanied by contact information to facilitate these unsolicited comments.

A user's role in reimbursable surveys is typically as a partner or constituent of the sponsor, although there are exceptions—particularly with hybrid surveys like the Current Population Survey, which has multiple components with different sponsors (including the Census Bureau) under an umbrella reimbursable survey. In the hybrid case, users' needs and comments often do come directly to the Census Bureau, but they also come indirectly through the funding source (or sources). As is true in other instances, the user's needs on reimbursable surveys have to be weighed against the sponsor's needs for the overall program and against budget and disclosure constraints.

Constraints

All efforts are made to comply with reasonable requests for changes or enhancements that conform to a data collection project's purpose and goals but, as noted, there are constraints. Regardless of concerted efforts to coordinate with our partners, there are circumstances when needs cannot be met due to insufficient funds. Given the nature of the budget cycle, these constraints are often unpredictable and are often significant (and sometimes both). This situation often allows the players in the budget process in some instances to be *the* most influential Census Bureau partners.

For example, in response to clear and large demand from agencies and users, the Census Bureau put forth a budget initiative on several occasions to reinstate an overlapping panel design for SIPP. The need for an overlapping panel design was identified as part of a larger recommendation to provide data to support a modernization of the official poverty measure—a recommendation from the National Academy of Sciences, reinforced by prominent researchers, user groups, and federal agencies. The budget initiative was rejected each time it was offered, so the highly sought after design change to SIPP has not been implemented. The Census Bureau remains committed to responding to user and stakeholder needs to provide data to improve the measurement of poverty but cannot comply without significant funds, the absence of which creates a solid barrier to cooperation.

The budget constraint can be minimized, of course, when the stakeholder can fund the enhancement. We recently initiated a project to extend the SIPP sample to target a larger segment of individuals receiving Supplemental Security Income or Social Security benefits, because the requestor (the Social Security Administration) was able to provide financial support for the data collection and was able to select the sample from their administrative records.

A second type of constraint is that sponsor, stakeholder, and user requests cannot be fulfilled if they are not in line with the Census Bureau's mission; if they have a negative impact on the Census Bureau's reputation; if they are not consistent with the production of high-quality data; if they do not address sensitive populations and topics thoughtfully; if they do not comply with Census Bureau policies governing content, development, administration, and testing; and if they do not work well within the larger purpose, scope, and design of the survey or data collection effort. Whenever a request comes in, we work with the requestor to adjust the specifics of the request in an attempt to conform to these constraints, if they do not at the outset.

A third constraint is respondent burden. Overall, of course, there are limits on how much time respondents can be asked to spend responding to federal surveys—which, in turn, places limits on the ability to respond to the needs of sponsors, users, and stakeholders. Often, that means it is difficult to expand lines of questioning that are not directly related to the specific purpose of the survey or to improve the precision of a particular estimate through increased probing of respondents. There are trade-offs in the burden metric, so that one can ask more questions—if the size of the universe for each question is restricted to the point where there is no increase in the time respondents take to respond to the survey, on average.

Once a change is agreed to in principle, it must be "proven in," which is a fourth constraint. We believe that pretesting is critical to the successful collection of the information needed, as it helps to ensure the instruments used to collect the data do accurately measure the intended concepts. Hence, the Census Bureau has a pretesting policy for data collection instruments that requires all questions to be field and/or cognitive tested before they are fielded in a production survey. The pretesting policy accepts, as a substitute for pretesting, proven success of a particular item in the field in a different context. However, since many requests are for data items that are substantively different from items on other surveys, this policy places limits on the introduction of new questions to meet user/sponsor/stakeholder needs when the cost and time requirements for pretesting exceed allowable limits or available resources.

The other major constraints to responding to user and stakeholders' needs are the protections the Census Bureau imposes on data collected on households. For any data collected under Title 13, we cannot and do not publicly disseminate any information that can be use to identify a respondent. This task of disclosure proofing these data is becoming increasingly difficult, with the recent explosion of publicly available information on individuals and of tools to easily locate and access that information. To protect against disclosure of respondent identities, we cannot issue public microdata files with low levels of geography or that identify unusual demographic or

economic events. If users need these data to carry out the analysis, we cannot respond by enhancing the public data products. We have other options, however, to provide sponsors, users, and stakeholders with what they need, when their requests cannot be fulfilled with public data.

- As noted, sponsors can elect to have data collected under Title 15, so that they can access the full array of information collected. This option is used when the sponsor—rather than the Census Bureau—selects the sample (as is true for the SSA project noted above). Since they already have the identities of the individuals selected into the sample, the identity of respondents selected by the sponsor cannot effectively be protected from the sponsor.

- Another option is to offer data users the choice of submitting a proposal to carry out their work at one of the Census Bureau-run research data centers spread across the U.S. If the proposal is accepted, the researcher can become a special sworn-status "employee" of the Census Bureau and thus be subject to all the laws and penalties for misuse of data. In that case, they are approved to work at a Census Bureau site using more detail than is publicly available. They work under the supervision of Census Bureau staff and can only remove results from the center that meet the Title 13 constraints.

- Finally, users requiring more detail than can be disseminated on a public-use microdata file can request a special tabulation of the nonpublic files and can receive the results in aggregate form (if they meet the Title 13 restrictions).

Process

In spite of the major influence of the budget and the presence of other constraints, the Census Bureau does adjust survey or sample design or postcollection processing systems to meet the needs of sponsors, stakeholder, and users. Sometimes, there is a lot of room for compliance with the request—particularly at the beginning of a long term program or at the point of a major redesign. On most occasions, however, only marginal adjustments are feasible.

To implement requested changes, the Census Bureau must coordinate with all of its partners in the survey, and the process of doing that varies by the type of mandate under which the data are collected. As noted, we have certain standards and policies that impact the relationship with sponsors. There are certain types of information the Census Bureau will not collect. For example, we will not collect information if it requires biological samples from respondents. In those cases, the sponsor will have to decide either to not include those data items or to seek another data collection agent.

In the case of reimbursable surveys, the sponsor's needs dominate. When a request to change a reimbursable survey comes to the Census Bureau, we work with the sponsor to see if we can refine the collection strategy in response to the request, and to see if we can do so within the budget constraints the sponsor faces. If not, the request is modified or rejected. If so, the final

decision to accept or reject resides with the sponsor, since it is the sponsor who largely determines the scope and major design features of the survey. The sponsor may not agree to the requested change, even if the request falls within the constraints noted above (in which case, the request is rejected). Of course, requests from stakeholders and users may go directly to the sponsors, who have various ways in which they interact with users and stakeholders (interagency or advisory committees, user groups, federal register notices) to gauge the appropriate direction to take for their surveys. In those cases, the changes are requested by the sponsor to the Census Bureau and negotiated as part of the ongoing working relationship between the Census Bureau and the sponsor.

A different process governs any request to enhance or otherwise change legally-mandated items on surveys or censuses. To accomplish change, we make recommendations to Congress based on our understanding of the legal requirements and based on guidance we receive from established advisory committees (whose purview includes that content). Congress will approve (or not) the recommendations and, when they do not, we revise and present new ones until the content becomes agreeable. This effort is largely carried out working with Congressional staff of the various committees that oversee the Census Bureau or have data needs. Stakeholders and users are represented in the process either through the advisory committees or their congressional representatives, and make requests for changes through these groups.

The Census Bureau is always open to (and frequently solicits) suggestions for enhancements to the surveys and projects we sponsor. Census-sponsored surveys have various mechanisms for soliciting input on content and design and for implementing requested changes. The staff maintain a presence at professional meetings and conferences on topics related to our data collection efforts. At these conferences, we often present updates on the status of Census-sponsored surveys and discuss research and other efforts that influence the survey design, execution, or dissemination. The Census Bureau maintains websites for the Bureau as a whole and for individual projects, and those websites provide contact information for individuals who can accept and process requests for changes. The Census Bureau has a marketing services office to encourage familiarity with and use of our products, and to support display and information booths at conferences and other meetings. This office also provides conferences to help users of data, particularly tabular or aggregate data, complete their analysis. The Census Bureau also includes formal notification of pending data collection efforts in the Federal Register for public comment.

Occasionally, a specific Census-sponsored project or survey will initiate a survey of users to determine the most desired content and design features. Some projects have committees of stakeholders (formal advisory committees, technical working groups, and OMB-sponsored interagency groups), through which comments and technical review and evaluations are sought. Interactions with these groups frequently lead to changes in some aspect of the survey or its processes. An example is the American Community Survey, which has formal advisory committees to which it needs to respond, as well as congressional committees and user groups.

At the project level, we often maintain open list serves and/or working groups that tend to be populated by heavy users or those with a strong interest in a particular data collection effort.

These provide forums for discussion of issues, sharing of techniques for understanding or analyzing data, and suggestions for changes. For example, SIPP has a user list serve, an interagency group to review topical modules, and a local users group that meets once or twice a year. The project managers at the Census Bureau will also arrange for periodic meetings with users to solicit input or to announce new products or services. These are convenient forums to gain information about a survey and to provide comments.

In addition to survey-specific groups, the Census Bureau often sponsors, cosponsors, or participates in meetings and seminars focused on specific themes, and these meetings frequently generate suggestions for improvements to both Census-sponsored and reimbursable surveys. The meetings tend to be topic- or function-based but can take on many forms and be sponsored by a variety of different organizations. For example, there are the two interagency committees on nonresponse sponsored by Federal Committee on Statistical Methodology; over the years, there have been Census-sponsored working groups formed by the Association of Public Data Users to discuss Census Bureau data products.

At the moment, the Census Bureau is considering the possibility of establishing one or a series of user conferences for users of microdata from our surveys. This would not overlap with the existing seminars on tabular data, because this series would be restricted to issues unique to using the microdata directly. We expect this series of conferences would yield good suggestions for survey enhancements.

Conclusion

Sponsors, stakeholders, and users have significant influence over the design, access, and analytical utility of Census Bureau demographic surveys. The reimbursable surveys are governed for the most part by the sponsors goals and budgets and the scope of the project is negotiated formally through a contractual arrangement that governs the transfer of funds to the Census Bureau. The U.S. Congress has a great deal of influence over census-sponsored activities, largely through the budget-setting process and through legal mandates for collection of information. The Office of Management and Budget has influence through the clearance process and assessment of the burden of collection on the general population. Users provide both direct and indirect feedback on the analytic utility of the information provided by the Census Bureau, which is then used to guide decisions on data file and survey design and content.

Of course, all requests for enhancements have to be screened to ensure they are consistent with the budget and scope of the survey, as well as Census Bureau and federal guidelines for collection and dissemination of data. These constraints limit the amount of change that can be included, but they do not prevent change altogether.

Enhancing the Design, Access and Analytical Utility of Federal Surveys Through Coordinated Efforts Between Sponsors, Stakeholders and Data Users

Steven B. Cohen, Center for Cost and Financing Studies
Agency for Healthcare Research and Quality

Introduction

Co-ordinated efforts between survey sponsors, stakeholders and data users have been demonstrated to yield synergies that have been quite successful in facilitating enhancements to the design, access and analytical utility of federal surveys. This paper provides several examples of effective co-ordinated efforts in achieving notable survey design and analytic enhancements to a national information resource to inform health policy, the Medical Expenditure Panel Survey (MEPS). Attention is given to the analytical enhancements and design efficiencies introduced to the MEPS as a consequence of the Department of Health and Human Services Survey Integration Plan. Examples are provided of additional content enhancements to the MEPS to support health care quality measurement that were achieved through coordinated efforts. Furthermore, the collaborative efforts between the Agency for Healthcare Research and Quality (AHRQ), the Bureau of the Census, the Bureau of Labor Statistics, the National Center for Health Statistics, CDC, the Centers for Medicare and Medicaid Services and OMB are discussed, with attention given to the design improvements realized and the enhanced state level estimation capacity achieved for the MEPS Insurance Component.

Background

The Medical Expenditure Panel Survey was designed to produce national and regional annual estimates of the health care utilization, expenditures, sources of payment and insurance coverage of the U.S. civilian non-institutionalized population. The MEPS includes a survey of medical providers, to supplement the data provided by household respondents. The design of the MEPS permits both person based and family level estimates. The scope and depth of this data collection effort reflects the data needs of government agencies, legislative bodies, and health professionals for the comprehensive national estimates needed in the formulation and analysis of national health policies. The survey is sponsored by the Agency for Healthcare Research and Quality (AHRQ).

The MEPS collects data on the specific health services that Americans use, how frequently they use them, the cost of these services and how they are paid, as well as data on the cost, scope, and breadth of private health insurance held by and available to the U.S. population. MEPS is unparalleled for the degree of detail in its data, and its ability to link health service medical expenditures and health insurance data to the demographic, employment, economic, health status, utilization of health services, and other characteristics of survey respondents. Moreover, the MEPS provides a foundation for estimating the impact of changes in sources of payment and insurance coverage among various economic groups or special populations of interest, such as the poor, the elderly, veterans, the uninsured, and racial and ethnic minorities (J. Cohen, 1997).

DHHS Survey Integration Plan and MEPS Enhancements and Efficiencies

As part of the Reinventing Government Part II (REGO II) activities, DHHS targeted improvement of the analytic capacity of its programs, filling of major data gaps, and establishment of a survey consolidation framework in which DHHS data activities are streamlined and rationalized. A Survey Consolidation Working Group was charged with developing a consensus plan for meeting these objectives (Hunter, Arnett, Cohen, et al., 1995; Arnett, Hunter, Cohen, et al., 1996).

A major concentration of the Survey Integration Plan was the redesign of the health care expenditure and insurance studies conducted by DHHS, which include the National Medical Expenditure Survey (NMES, the precursor of the MEPS), the Medicare Current Beneficiary Survey (MCBS), and National Health Interview Survey (NHIS). The proposed survey integration plan was designed to achieve significant cost efficiencies by eliminating duplicative efforts and reducing overall respondent burden. Furthermore, the analytic capacities of the component surveys were enhanced because their design features were integrated. To improve survey design capabilities, enhancements such as an ongoing longitudinal survey effort and the capacity to derive State-specific health care estimates were considered. Consideration was also given to including a periodic institutional component in the survey to provide national use and expenditure estimates for the population residing in nursing homes (Hunter, Arnett, Cohen, et al., 1995).

Enhancements and Efficiencies Through Survey Integration:
One attraction of the DHHS Survey Integration Plan was the enhanced analytic capacity to be achieved by linking the distinct surveys through design integration. Use of NHIS as a sample frame for MEPS increased the analytic content of the resultant linked surveys. Through design integration of DHHS surveys, inefficiencies associated with duplicative survey efforts were reduced. Another goal was to reduce survey design costs by implementing a uniform framework for DHHS-sponsored surveys that have overlapping analytic focus with respect to questionnaire content, data editing, imputation, estimation, database structure, and development of analytic files.

By moving to this integrated, annual household data collection effort, DHHS expanded and enhanced its analytic capabilities. The DHHS Survey Integration Plan:

- Retained the design of the core NHIS household interview. This core provides cross-sectional population statistics on health status and health care use, with sufficient sample size to allow for analyses based on detailed breakdowns by age, race, sex, income, and other sociodemographic characteristics. The core also allows the use of data on a broad range of topics currently covered by NHIS;
- Retained the analytic capacity to obtain annual and quarterly population estimates of health care use and the prevalence of health conditions, both for the Nation and for policy-relevant population subgroups;
- Provides the ability to model individual and family-level health status, access to care and use, expenditures, and insurance behavior over the year and examine the distribution of these measures across individuals. The longitudinal feature of MEPS (collecting data

over multiple years) further enhances the capacity to model behavior over time;

- Provides the ability to relate data from a detailed sample (e.g., MEPS) to a larger sample (e.g., NHIS) to enhance the utility of MEPS for national health account estimation and microsimulation modeling, including disaggregation by age group or geographic area.
- Provides the potential to yield both national and State-level estimates for marginal costs using the enhanced sample design of the NHIS, which includes 358 primary sampling units;
- Provides, as a result of the longitudinal aspect of the MEPS integrated data collection effort, an increase in statistical power to examine change or make comparisons over time; the capacity to examine changes over time as well as changes in the relationships among measures of health status, access to care, health care use, expenditures, health insurance coverage, employment, functional limitations and disabilities, and demographic characteristics.

Enhancements to MEPS Household Component

The original NMES-3 sample design called for an independent screening interview to identify a nationally representative sample and facilitate oversampling of policy-relevant population subgroups. Data collection and training costs associated with this independent screening interview were projected to exceed $8 million. As part of the DHHS Survey Integration Plan, this separate screening interview was eliminated. Instead, NHIS was specified as the sampling frame for MEPS. NHIS is an ongoing annual household survey of approximately 42,000 households (109,000 individuals) conducted by the National Center for Health Statistics (NCHS) to obtain national estimates on health care use, health conditions, health status, insurance coverage, and access for the U.S. civilian noninstitutionalized population. In addition to the cost savings achieved by substituting NHIS as the MEPS sample frame, the design modification resulted in an enhanced analytic capacity of the resultant survey data. In addition, use of the 1995 NHIS data in concert with the 1996 MEPS data provides additional capacity for longitudinal analyses not available in the original (NMES-3) design. Furthermore, the greater number and dispersion of the sample primary sampling units that comprise the MEPS national sample resulted in improvements in precision over the original design specifications.

Design and Estimation Strategies and Innovations in the MEPS for the Measurement of Health Care Quality

Efforts are underway in the Department of Health and Human Services towards the development of a national health care quality reporting system. The purpose of the reporting system is to provide an annual profile of the nation's quality of care and to help measure improvements over time. Quality is often defined as meeting customers' expectations. Consequently, the quality reporting system will need to include a comprehensive set of indicators that characterize several dimensions of patient satisfaction and consumer satisfaction with providers, health plans and access to care. This section focuses on the statistical and methodological design strategies and innovations in the MEPS achieved through coordinated efforts between survey sponsors and experts in quality measurement both within DHHS and the research community at large.

Coordinated Efforts of the AHRQ-MEPS Steering Group to Enhance Survey Design, Analytic Utility and Data Access

The Medical Expenditure Panel Survey (MEPS) is the only longitudinal, nationally representative survey designed to provide in-depth information on the health care use, expenses, payments and insurance coverage. AHRQ's reauthorizing legislation and data requirements for the National Quality Report (NQR) and the National Disparities Report recently necessitated the implementation of a series of "fast-track" enhancements to the MEPS to permit improved health care quality measurement and studies of access to care at the national level. An AHRQ-MEPS Steering Committee was established to provide recommendations to the Director of AHRQ regarding the most appropriate enhancements to the MEPS content to permit analyses of the relationships between health care quality, outcomes, access, use and cost at the national level; to provide information on the quality of care and patient outcomes for frequently occurring clinical conditions; and to implement design changes to improve the precision of survey estimates through cost effective sample design modifications. From its inception in the Spring of 2000, the Committee members included a wide range of science partners in informing recommended enhancements, and also served to align the MEPS and its products more directly with all the goals of the Agency. All Committee recommendations were implemented rapidly without jeopardizing the effective operation of the MEPS survey. Without their work, it would not have been possible for the Agency to provide information on the relationship between health care quality, outcomes, access, use and cost to department stakeholders including academicians, insurers, employers, the Assistant Secretary for Planning and Evaluation (ASPE), the Office of Management and Budget (OMB), the National Center for Health Statistics, CDC, and the Centers for Medicare and Medicaid Studies (CMS). The scope and depth of the resultant enhanced MEPS data collection effort reflects the needs of government agencies, legislative bodies, and health professionals for comprehensive national estimates necessary for the formulation and analysis of national health policies.

The Committee efforts substantially increased the number and diversity of research users - in and out of AHRQ - in the specification of the MEPS enhancements related to the content, design and direction of the survey. The MEPS data made available for analysis through this Committee's efforts are currently being used to inform questions about the health care quality of the nation. The MEPS enhancements will permit more detailed studies of concern to the Department and the public: the extent to which Americans, and especially children, have access to care; their use of clinical preventive services; their satisfaction with health plans; and their health care quality.

Design and Content Modifications to the MEPS to Support Quality of Care Analyses at the National Level

The MEPS healthcare quality enhancements called for a significant household survey sample expansion of individuals with certain illnesses of national interest in terms of patient satisfaction with care received, the quality of the care and the burden of disease. The intent of this enhancement was to permit more focused analyses of the quality of care received for these special populations. In order to move forward with sample design analyses and MEPS questionnaire design modifications according to schedule, it was necessary to finalize the set of medical conditions that would be given special emphasis with respect to health care quality measurement and patient satisfaction.

A set of formal criteria were established to guide the decision making process regarding the selection of the set of medical conditions that were to be given special attention for implementing the planned MEPS healthcare quality enhancements. More specifically, the selection decision was based on an evaluation of conditions using the following criteria:

- Sufficient prevalence to support reliable estimates,
- Availability of diagnostic questions used in other national surveys,
- Accuracy of household reported conditions,
- Availability of evidence-based quality measures, and
- Level of medical expenditures for treatment of the condition.

Based on the review of the criteria under consideration, it was recommended that the following medical conditions be given special attention for implementing MEPS healthcare quality enhancements based on their capacity to meet most or all of the specified targets: Diabetes, Asthma, Hypertension, Ischemic Heart Disease, Arthritis, Stroke and COPD. It should be noted that the selection of diabetes and ischemic heart disease as targeted conditions also cover two clinical areas that are the focus of the forthcoming DHHS Report on Health Care Disparities. A summary of the availability of relevant diagnostic questions, the capacity of households to accurately report these conditions, the availability of evidence based quality measures and the level of medical expenditures for treatment of the conditions under consideration are available from AHRQ.

To further improve the precision of the survey estimates beyond the gains from the increase in geographic areas from 100 PSUs to 195 PSUs, in particular for individuals with at least one of the medical conditions given special attention for implementing MEPS healthcare quality enhancements, a decision was made to increase the 2002 MEPS sample to a total sample of 15,000 households. In addition, the following two sample allocation methods were under consideration for implementing the desired sample increase: 1) the adoption of a uniform sample size increase versus 2) a targeted oversample of individuals with specific conditions. As a consequence of the subsampling method within households adopted in the National Health Interview Survey to obtain medical condition data (the selection of only one adult and, when available, one child to answer the questions related to medical conditions), it was recognized that the implementation of a targeted oversample of individuals with specific conditions would be significantly limited by the constraints of the NHIS design. Consequently, the sample design recommendation was to implement a sample size increase in MEPS that would enhance the representation and precision of the targeted conditions without a targeted oversample. This sample design modification has the following attractions :

- For fixed sample size, it achieves greater precision in national estimates of general population characteristics relative to a targeted oversample
- It required only minimal modifications to the prior MEPS sample selection procedures;

There are minimal additional complexities in the development of MEPS estimation weights.
In addition to the improvements in precision for individuals with the targeted conditions, the adoption of this sample enhancement in MEPS for 2002 also facilitated gains in precision for minorities and ethnic groups which support the Department's Initiative to Eliminate Racial and Ethnic Disparities, for adults with functional limitations and for children with special health care needs.

Inclusion of Additional Questions in a MEPS Self Administered Questionnaire (SAQ) to Measure Quality of Care and Patient Satisfaction:
The selection of a core set of questions that measure quality of care and patient satisfaction was governed by the need to adopt measures that were carefully tested and validated, to insure the collection of meaningful and reliable information. Consequently, a subset of questions that were developed for the Consumer Assessments of Health Plans Study (CAHPS) were selected for inclusion in a self-administered questionnaire (SAQ) in the MEPS to measure several dimensions of healthcare quality and patient satisfaction. In addition, the Self Administered Questionnaire included the complete set of questions from the SF-12 (Medical Outcomes Study, Short Form) to improve the survey's capacity to measure health status. It also included the set of questions that comprise the EuroQuol 5D (EQ-5D), including the visual analogue scale, to facilitate international comparisons on health status and quality measurement.

Data Center
Many MEPS databases include considerably more data that can be made available to the general public because of the constraints of confidentiality guidelines. In order to facilitate the use of such data, while maintaining the confidentiality promised to respondents, AHRQ's Center for Cost and Financing Studies (CCFS) has developed a Data Center, which is a physical space at AHRQ in Rockville, Maryland where researchers with approved projects can be allowed access to data files not available for public dissemination. These data, which are classified as "restricted", contain information that are not released to the public. These data sets may contain geographic variables at a lower level than released for public use, more detailed condition information, or may consist of unedited data base segments not yet prepared for public release. These restricted data sets do not contain information that would directly identify a respondent (name, social security number, street address).

In order to protect the confidentiality of respondents, the physical environment in the CCFS Data Center is monitored. Researchers are allowed access only to the information required to complete their project. Materials cannot be removed from the Data Center until they have been reviewed by specific CCFS staff for disclosure avoidance. This disclosure review is conducted by a CCFS employee with knowledge of the project and is also reviewed by the Data Center Manager. Only summary output (tables, regression equations, parameter estimates) may be removed from the Data Center. Micro data files can not be removed from the Data Center.

Coordinated Efforts of the Interagency Committee on Employment-Related Health Insurance Surveys to Enhance Survey Capacity

The Interagency Committee on Employment-Related Health Insurance Surveys includes the following federal organizations as participants: AHRQ, the Bureau of Labor Statistics (DOL/BLS), Centers for Medicare and Medicaid Services (CMS), NCHS, the DHHS Office of the Assistant Secretary for Planning and Evaluation (ASPE), the Bureau of Economic Analysis, OMB, the Department of the Treasury, and the Bureau of the Census. The purpose of the committee is to communicate and coordinate federal efforts to collect information on establishment-based health insurance. Furthermore, a stated goal is to understand the purpose of each survey, the uses of survey data, the needs of data users, and the gaps in information collected. The Committee's immediate focus was on the BLS sponsored National Compensation Survey (NCS) and AHRQ's Medical Expenditure Panel Survey - Insurance Component (MEPS-IC), with the objective of:

- Investigating the aims of each survey, types of information collected, estimates produced, uses of data for estimation and research
- Assessing similarities and differences in uses of surveys and data collected
- Assessing gaps in data collection and data needs.

The MEPS Insurance Component (IC) consists of two subcomponents, the household sample and the list sample. The household sample collects detailed information on the health insurance held by and offered to respondents to the MEPS Household Component. These data, when linked back to the original household respondent, allow for the analysis of individual behavior and choices made with respect to health care use and spending. The list sample consists of a sample of business establishments and governments throughout the United States. From this survey, national, regional, and State-level estimates (for almost all States each year) can be made of the amount, types, and costs of health insurance available to Americans through their workplace. The Committee's efforts in reviewing the focus of the MEPS-IC and the NCS helped ensure the analytical objectives of the respective surveys were mutually reinforcing and complementary, rather than overlapping.

Based on the coordinated efforts of this Interagency Committee, the Bureau of Economic Analysis uses data from the MEPS Insurance Component in the computation of the health cost component for employer sponsored health insurance coverage for estimates of the US Gross Domestic Product (GDP) and is studying the potential use of MEPS IC data for their State-level measures. Many other Federal offices, such as the Treasury Department, the Joint Committee on Taxation, the Centers for Medicare and Medicaid Services, and the Pension and Welfare Benefits Administration, are frequent users of MEPS IC data and often make special request for specific estimates.

Many of the MEPS IC estimates are at the State-level - making them particularly valuable to both Federal and State agencies. Special data request have been provided to representative agencies from most States. In support of the HRSA State Planning Grant program (that helps State agencies analyze and address the issue of the uninsured), the MEPS IC survey has produced many additional tables of estimates. Some States (Massachusetts, Arkansas, and

Wisconsin) have provided funding for additional MEPS IC sample for their States in order to improve their State estimates for specific years. In the past two years, HRSA has also funded additional MEPS IC sample in many of their grantee States to increase the number of States for which estimates can be made in a given year.

Summary

Over the past several years, the Medical Expenditure Panel Survey (MEPS) data have quickly become a linchpin for the nation's economic models and their projections of health care expenditures and utilization. The enhanced level of detail and analytical content enables public and private sector economic models to develop national and regional estimates of the impact of changes in financing, coverage, and reimbursement policy, as well as estimates of who benefits and who bears the cost of a change in policy. No other national population based survey provides the foundation for estimating the impact of changes on different economic groups or special populations of interest, such as the poor, elderly, veterans, the uninsured, or racial/ethnic groups. This paper has highlighted several examples of effective co-ordinated efforts between survey sponsors, stakeholders and data users, to demonstrate the notable enhancements in design, access and analytic utility for the MEPS that have been adopted to help inform health policy and facilitate health care quality measurement.

Acknowledgment

The views expressed in this paper are those of the author and no official endorsement by the Department of Health and Human Services or the Agency for Healthcare Research and Quality is intended or should be inferred. The author wish to thank Joel W. Cohen and Trena Ezzati-Rice for their careful review of the manuscript and helpful comments.

References

Agency for Healthcare Research and Quality (1996). Technical Overview of the Consumer Assessment of Health Plans.

Arnett RA, Hunter E, Cohen S, et al. The Department of Health and Human Services' Survey Integration Plan. In: Proceedings of the American Statistical Association (ASA). Section on Government Statistics. Chicago: 1996 Aug.

Cohen, J. W. (1997). "Design and Methods of the Medical Expenditure Panel Survey Household Component .Rockville (MD): Agency for Health Care Policy and Research; 1997. MEPS Methodology Report, No. 1. AHCPR Pub. No. 97-0026.

Cohen SB. The redesign of the Medical Expenditure Panel Survey, a component of the DHHS Survey Integration Plan. Proceedings of the Council of Professional Associations on Federal Statistics (COPAFS) Seminar on Statistical Methodology in the Public Service, Bethesda (MD); 1996 Nov.

Cohen, Steven B. (2000) "Methodological Issues for the Design of Consumer and Patient Satisfaction Surveys." Forthcoming in the *2000 Proceedings of the American Statistical Association, Section on Health Policy Statistics.*

Cohen, S.B. (2000) "Sample Design of the 1997 Medical Expenditure Panel Survey Household Component". MEPS Methodology Report 11. AHRQ Pub. No. 01-0001.

Hunter E, Arnett R, Cohen S, et al. HHS Survey Integration Plan: Background materials. Agency for Health Care Policy and Research, Rockville (MD), and National Center for Health Statistics, Hyattsville (MD); 1995.

Westat, Inc. (2000). "Survey Design Evaluations to Inform the MEPS Health Care Quality Enhancements". Working papers.

Coordinated efforts involving the National Center for Health Statistics and its survey co-sponsors, stakeholders, and data users[3]

Jane F. Gentleman, National Center for Health Statistics[4]

This paper describes coordinated activities within the National Center for Health Statistics (NCHS), which is part of the Centers for Disease Control and Prevention (CDC), and between NCHS and its survey co-sponsors, stakeholders, and survey data users. Some of these activities are the results of survey integration efforts that began in the previous decade within the Department of Health and Human Services, where survey integration may be thought of as the conscious design and carrying out of surveys so as to achieve synergy between surveys that improves the effectiveness of the surveys. This paper focuses mostly on surveys conducted by NCHS' Division of Health Interview Statistics.

The National Center for Health Statistics has four "data divisions," defined according to the type of data collected. Vital statistics–administrative data on births, deaths, and other life-related events–are collected by the Division of Vital Statistics from all of the states, and processed and merged into national data bases maintained at NCHS. One product is the National Death Index, a cumulative compilation of information about all deaths in the United States. Collecting national vital statistics requires ongoing consultation and cooperation among the states and NCHS. An example of such cooperation is the development of standard birth and death certificates that improve comparability of the data from different states and facilitate combining and analyzing those data. The standards are reviewed and revised approximately every 10 years, with participation in that process by data users, including recognized experts in epidemiology and public health. For further information on the national vital statistics system, see Freedman and Weed (2002) and references cited therein.

The National Health Care Survey, conducted by the Division of Health Care Statistics, is really a family of sample surveys that gather data on the use of health services and on the characteristics of patients, providers, and facilities involved in health care transactions. These surveys cover hospitals, nursing homes, doctors' offices, emergency rooms, ambulatory care units, etc. One challenge is the goal of creating components of the National Health Care Survey that are mutually exclusive and exhaustive in their coverage of the health care delivery systems. In reality, the boundaries between the different types of health care systems are sometimes blurred, and single individuals commonly utilize two or more of these systems in a given time period. To adapt to rapid changes in health care delivery systems, NCHS is updating its health care survey sampling frames and survey designs, which has involved extensive consultation with experts and data users. For further information on the National Health Care Survey, see Demlo and Gentleman (2002) and references cited therein.

---

[3] Presented at the Federal Committee on Statistical Methodology's Statistical Policy Seminar on Challenges to the Federal Statistical System in Fostering Access to Statistics, Enhancing the Design, Access and Analytical Utility of Federal Surveys Through Coordinated Efforts Between Sponsors, Stakeholders and Data Users, Bethesda, Maryland, 2002.

[4] Director, Division of Health Interview Statistics, National Center for Health Statistics, 3311 Toledo Road, Hyattsville, MD 20782.

The Division of Health Examination Statistics conducts the National Health and Nutrition Examination Survey (NHANES), an ongoing series of surveys that originated in 1960. A random sample of subjects answer questions about their health, and they undergo extensive physical examinations in NHANES' specially-outfitted trailers. These Mobile Examination Centers visit communities around the country each year. NHANES managers periodically issue calls for proposed topical material to be covered by the survey. At any given time, intense collaboration occurs among NCHS and some 15-20 collaborators who are co-sponsoring the survey. NHANES also organizes conferences regularly to facilitate communication among survey managers, co-sponsors, and data users. For further information on NHANES, see Berman et al. (2002) and references cited therein.

Interview surveys conducted by the Division of Health Interview Statistics include the National Health Interview Survey (NHIS), the National Immunization Survey (NIS), the State and Local Area Integrated Telephone Survey (SLAITS), and the Joint Canada/United States Health Survey (JCUHS). NHIS is the principal source of information on the health of the civilian, noninstitutionalized household population of the United States. It is an in-person interview survey, covering everyone living in about 41,000 households (about 107,000 persons) each year. NIS is a telephone survey that collects data on immunizations received by children 19-35 months of age from all 50 states and in 28 metropolitan areas. It is co-sponsored by the National Immunization Program in Atlanta and NCHS. SLAITS is a telephone survey mechanism that utilizes the same sampling frame as NIS to conduct topical surveys, either national or state-based. JCUHS is a one-time (2002-2003) bi-national telephone survey covering the United States and Canada at the same time with virtually the same questions. The remaining discussion in this paper will focus on activities involving these DHIS surveys. For further information on NHIS, NIS, and SLAITS, respectively, see Demlo and Gentleman (2002), Zell et al. (2000), and Blumberg et al. (2002), and references cited therein. For further information on JCUHS, see Gentleman (2003).

Some coordinated activities between surveys/agencies

The National Health Interview Survey (NHIS) and the National Health and Nutrition Examination Survey (NHANES)

Many of the questions on the NHANES questionnaire are also on the NHIS questionnaire. This permits comparative analyses of results from the two surveys for purposes of assessing data quality and for cross-walking between the two surveys. For example, comparisons among NHANES physical examination data, NHANES interview data, and NHIS data are useful because interview data are self-reported or reported by proxy, and are thus prone to more reporting error than are objective physical examination data. Also, NHANES physical examinations can reveal undiagnosed conditions, yielding overall estimates of condition prevalence that should be higher than estimates based on interview data.

The National Health Interview Survey (NHIS) and the National Immunization Survey (NIS)

The child immunization section of the NHIS questionnaire until very recently contained a subset of questions that asked parents to provide the types and dates of their children's immunizations

and to give NCHS permission to contact the immunization provider(s) by mail to request further information. Having similar questions on both NHIS and NIS permits calibration of NIS estimates to adjust for the fact that NIS, as a telephone survey, cannot cover households without telephones.

The State and Local Integrated Telephone Survey (SLAITS) and the National Immunization Survey (NIS)

Fielding NIS requires screening a very large sample of households in order to identify a sufficient number of households with children of an appropriate age for NIS. For example, in 1999, more than 2 million phone numbers were called by NIS in the search for households with age-eligible children, resulting in the identification of about 36,000 such households. SLAITS capitalizes on that effort by utilizing not just some of the families screened into the NIS sample, but also some of the families screened out of NIS, depending on the requirements of the particular SLAITS survey being conducted. Because NIS targets children, SLAITS surveys are often about the health of children. For example, SLAITS' National Survey of Early Childhood Health (NSECH), conducted in 2000 by NCHS and co-sponsored by The Gerber Foundation, the American Academy of Pediatrics, and the UCLA Center for Healthier Children, Families, and Communities, addresses infants' and toddlers' health-related needs, pediatric health care experiences, and child-rearing practices. For further information about NSECH, see Blumberg et al. (2002).

The National Health Interview Survey (NHIS) and the Medical Expenditure Panel Survey (MEPS)

Half of the interviewed households from NHIS are reserved for subsequent follow-up by MEPS, which is conducted by the Agency for Healthcare Research and Quality. MEPS collects additional data from some of the NHIS respondents about health care use, health care expenses, and health insurance coverage. Linked NHIS-MEPS microdata, some of which are publicly available on the NCHS Web site, provide short-term longitudinal data for an extensive array of variables.

The National Health Interview Survey (NHIS) and the National Death Index (NDI)

Periodically, NCHS staff link NHIS data to the NDI, thus ultimately obtaining information about the underlying and contributing causes of death ("multiple causes of death") of NHIS participants. The linked microdata, which provide longitudinal information that is valuable for outcome analysis, are publicly available on the NCHS Web site.

The National Health Interview Survey (NHIS) and its supplement co-sponsors

Currently, a median time of 57 minutes is required to administer the NHIS to a family. In designing each year's NHIS questionnaire, about 20 minutes of this time is reserved for one or more sets of supplementary questions co-sponsored by agencies external to NCHS. The process of selecting, scheduling, designing, testing, administering, processing, and analyzing data from a one-year supplement involves several years of collaboration between NCHS staff and the external co-sponsor. Examples of supplements since 1990 are the Cancer Control supplement, co-sponsored by the National Cancer Institute, National Institutes of Health (NIH) and CDC; questions that track progress of the objectives of DHHS' Healthy People 2000 and Healthy

People 2010 programs; the Child Mental Health supplement, co-sponsored by the National Institute of Mental Health, NIH; Alternative Medicine, co-sponsored by the Center for Complementary and Alternative Medicine, NIH; and a short battery of questions about cell phone use, sponsored by NCHS.

The National Health Interview Survey (NHIS) and telephone surveys
The 2003 NHIS will contain questions about cell phone use, in addition to its ongoing core questions about the presence of ordinary telephones in the household. This NHIS supplement will provide designers and managers of telephone surveys with needed information to adapt to and adjust for the rapid proliferation of cell phones in the United States. Since many telephone surveys use households with land line telephones as their randomly-selected source of respondents, it is important for designers of telephone surveys to learn about the use of land line telephones versus wireless telephones by household residents.

NCHS and Statistics Canada
Since 1999, NCHS and Statistics Canada's Health Statistics Division have held an annual Interchange to share information about their many activities of common interest. At one of those meetings, a discussion of the difficulties of comparing estimates from the two countries' respective national health surveys (the NHIS in the United States, and the National Population Health Survey and the Canadian Community Health Survey in Canada) led to a plan to conduct a one-time, joint telephone survey covering both countries at the same time, and using essentially the same questions in both countries. Consequently, the Joint Canada/United States Health Survey began collecting data in late 2002. Respondents in Canada were interviewed in their choice of English or French; respondents in the United States could use either English or Spanish. The two co-sponsoring national statistics agencies will also collaborate in analyzing the data. This bi-national collaborative effort is consistent with the World Health Organization's goal to have a common health survey that will enhance the ability to compare health status across many countries.

The National Health Interview Survey (NHIS) and its responses to DHHS needs and regulations
NCHS surveys adhere to Office of Management and Budget (OMB) requirements for collection and presentation of information about race and ethnicity. For example, the NHIS question about a participant's race permits specification of more than one race, which is now an OMB requirement, and when administering that question, the NHIS interviewer displays a list of races categorized according to OMB specifications. Another example of NHIS supporting DHHS needs and regulations is the presence on every NHIS questionnaire in recent years of supplementary questions for measuring progress toward reaching objectives of DHHS' Healthy People program.

Some interactions between NCHS and data users

NCHS constantly interacts with users of its survey data. Some examples include the following:

- Release of microdata to the public
- Release and dissemination of analytical results
- Organized systems of responses to requests for information and data

- Maintenance of the NCHS Web site and of listserves
- Holding of workshops on specific surveys
- Sponsorship of the NCHS Data Users Conference
- The NCHS Research Data Center
- Sponsorship of expert panels

The examples above are but a few of the many NCHS activities involving interaction, cooperation, consultation, and coordination within NCHS and between NCHS and its survey co-sponsors, stakeholders, and data users. For extensive information about NCHS and its surveys, and access to selected NCHS microdata files, see the NCHS Web site at http://www.cdc.gov/nchs/.

References

Berman, Lewis; Ostchega, Yechiam; Reed-Gillette, Debra S, and Porter, Kathryn (2002). Chapter 33, The National Vital Statistics System. In: Public Health Informatics and Information Systems, Editors: Patrick W. O'Carroll, William A. Yasnoff, M. Elizabeth Ward, Laura H. Ripp, and Ernest L. Martin. Pages 710-740. Springer Verlag.

Blumberg, Stephen J.; Olson, Lorayn; Osborn, Larry; Srinath, K. P.; and Harrison, Holly (2002). Design and Operation of the National Survey of Early Childhood Health, 2000, National Center for Health Statistics. Vital and Health Statistics 1(40).

Demlo, Linda K. and Gentleman, Jane F. (2002). Chapter 15, Morbidity Data. In Public Health Informatics and Information Systems, Editors: Patrick W. O'Carroll, William A. Yasnoff, M. Elizabeth Ward, Laura H. Ripp, and Ernest L. Martin. Pages 286-315. Springer Verlag.

Freedman, Mary Anne and Weed, James A. (2002). Chapter 14, The National Vital Statistics System. In Public Health Informatics and Information Systems, Editors: Patrick W. O'Carroll, William A. Yasnoff, M. Elizabeth Ward, Laura H. Ripp, and Ernest L. Martin. Pages 269-285. Springer Verlag.

Gentleman, Jane F.; Simile, Catherine; Miller, Kristen; Bailie, Lorna; Lavigne, Mylène; and Madans, Jennifer (2003). Examining Comparability: A Joint Bi-national Population Health Survey in the United States and Canada. To appear in Proceedings of the International Conference on Improving Surveys, Copenhagen, 2002.

Zell, Elizabeth R.; Ezzati-Rice, Trena M.; Battaglia, Michael P.; and Wright, Robert A. (2000). National immunization survey: The methodology of a vaccination surveillance system. Public Health Reports 115, pages 65-77.

Session 4

E-Government and New Dissemination Paradigms

Introductory Remarks

Lawrence A. Greenfeld, Director
Bureau of Justice Statistics

Good afternoon and welcome to Session 4 entitled E-Government and New Dissemination Paradigms. We have some excellent speakers who are going to talk about Stats Canada and what they are doing to improve the distribution of information over the Internet and some speakers on FEDSTATS.

There are a variety of challenges confronting US Federal statistical agencies as greater centralization of computing authority and control occurs. Many of us must now work with parent agencies with CIO's and reason with them about mechanisms to consider which focus on how data can be insulated and protected, content controlled and managed by the stats office, and presentation of data offered in a manner consistent with the needs to assure privacy to respondents and to guard against pre-release. This is not easy as CIO's are now charged with taking control of entire Departments computing resources and often the budgets associated with both hardware and software acquisition. For a small stats agency in particular, protecting our core values about data and its handling for public use and our desire to insure the proper usability of what we produce is an emerging challenge. I have little doubt that soon all stats agencies will begin to face such issues.

Although this section is primarily focused on the users of statistical data, it is important to think about the use of the Internet as a data collection tool, particularly for adminstrative data from agencies. BJS has been gradually migrating certain collections from mail-out to web-based. This has created a set of interesting challenges with respect to respondent-identification and the ability to edit previously submitted data.

At BJS, we make all of our pubs and datasets for public use. We have nearly 4,000 staff-produced spreadsheets of data which are cross-referenced to relevant reports and datasets. Every graphic on our website easily converts to a spreadsheet for download with just a couple of clicks. In addition, we have a wide variety of datasets with which customers can directly interact to produce tabulations and cross-tabulations. Any number published by BJS should be capable of being reproduced by the public. What a dramatic change from the days when customers were bound by what was in books and limited to the use of whatever data was printed on a page of a Federal document. Having been in my field now for over 30 years, the extent to which we have liberated and democratized statistical information and the data used in computations in the last few years is absolutely awesome. It is our job to insure that those managing computing resources do not interfere with this kind of progress simply to promote uniformity within Departments. Maintaining the vitality, creativity, and exuberance in stats agencies about sharing their policy-relevant and publicly-funded information collections is our most important challenge and responsibility.

I am very appreciative for the work of Cathy Dippo from Bureau of Labor Statistics who organized this session and to our speakers and discussant.

We will begin the session with David Roy from Statsitics Canada who will present his thought about “How the Internet is transforming Client and Respondent Relationships at Statistics Canada.”

How the Internet is Transforming Client and Respondent Relationships at Statistics Canada

David Roy, Director Marketing
Statistics Canada

Introduction

I'd like to begin by thanking Cathryn Dippo of the Bureau of Labour Statistics for inviting Statistics Canada to take part in a discussion of E- Government and New Dissemination Paradigms.

Like other national statistics offices, (NSO), Statistics Canada's use of the on-line channel began well before the creation of a Canadian E-Government initiative. The Internet is a natural fit for the business of a national statistical office and our user communities were among its early adopters - so we were well advanced when Canada's E-Government initiative began in 1999.
Also, for many years there has been a sharing of information on dissemination and marketing strategies among NSOs. Most recently there was an excellent meeting held in early September in Annapolis involving sixteen countries, that was organized by John Kavaliunas and Colleen Flannery of the USCB. Statistics Canada has benefited greatly from these meetings and to some extent I think there is a great commonality in the dissemination strategies – emerging paradigms - of many of the participating countries because we have been sharing information on best practices for many years.

In my presentation I'll begin by giving some context to our activities by briefly describing the Canadian E-Government initiative. Then I'll give a summary of some of our activities – in dissemination and other key services and how our client relationships are being transformed -and finally I'll provide a couple of information sources on E-Government that you might find helpful.

Canada's Government On-Line Initiative

What I'm going to be talking about is Statistics Canada's activities that are related to a program called Government On-Line, (GOL). This Federal Government-wide initiative includes the delivery of all appropriate information and services on-line as well as a Service Improvement Initiative. The latter is essentially the application of marketing principles to government activities – understanding client needs, developing appropriate products and service standards and monitoring performance and client satisfaction. In Canada the GOL initiative has been strongly client focused. That's one new paradigm in itself.

The GOL initiative was launched in late 1999 with the goal of having 'all' information and services accessible online by 2004. This target has recently been extended to 2005 - in part because investment funds have not been as available since last September – and now the target only applies to services for which there is 'sufficient demand' to warrant the development of an online delivery option.

Of course, E-Government has a broader context than the service delivery focus of GOL, and it incorporates a more fundamental re-examination of our government and democratic processes. Statistics Canada participates in such an initiative and I'll say a few words about it at the end of the presentation.

The Government On-Line, (GOL), initiative was a high priority of the government from the outset and among the benefits frequently mentioned by the Prime Minister and Cabinet Ministers were:

- Playing a leadership role in creating the infrastructure and practices to encourage a wider use of the Internet among businesses,

- More efficient service delivery - a high priority of citizens because of its potential to lead to tax reductions,

- Higher approval ratings of the Federal Government by Canadians in public opinion and satisfaction surveys, and

- National unity through the perceived high value of Federal Services

Several parallel initiatives were conducted to increase connectivity of schools and communities, establish an appropriate technical infrastructure, make available cultural content and provide an environment conducive to e-transactions.

Government On-Line Objectives

Here you see some of the same ideas expressed in the objectives that were set for the GOL initiative

- Stronger relationships with clients and better service
  - Interact with more clients where they live and work

- A catalyst for electronic commerce
  - Help meet the Prime Minister's challenge to capture a 5% share of world of e-commerce by the year 2003

.

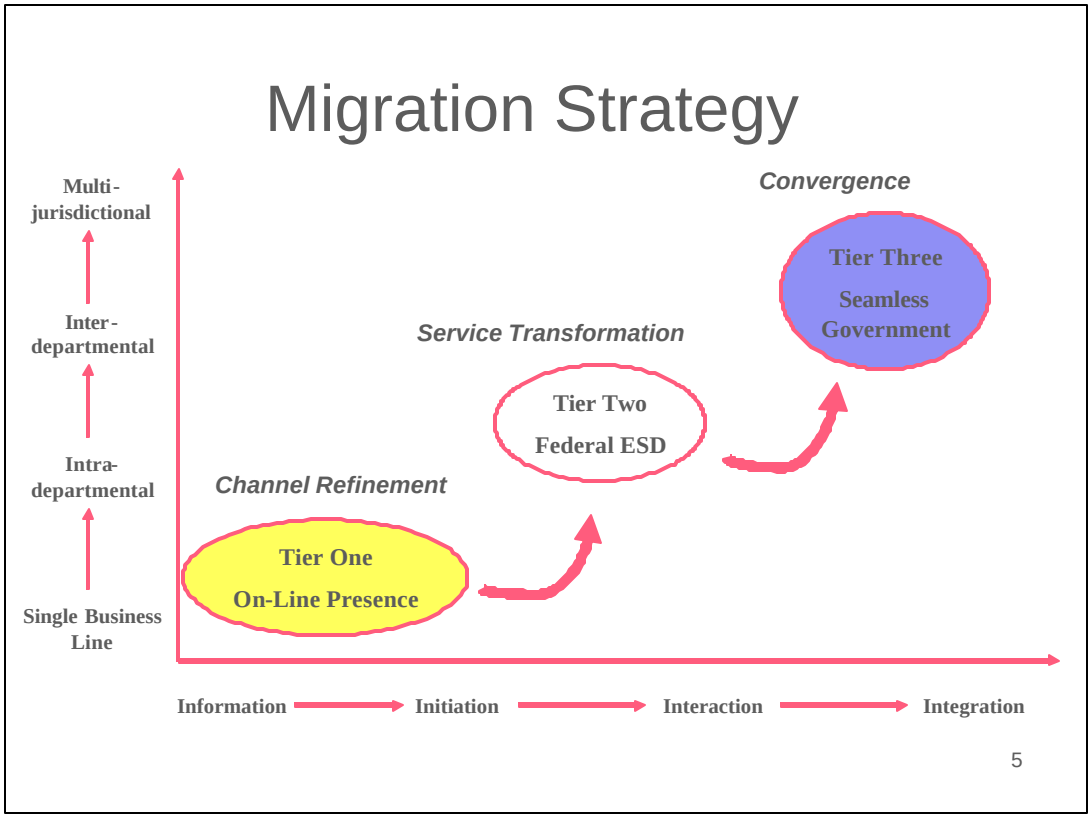
The focus on improved relationships was motivated by some early research conducted by our Treasury Board showing that Canadians' satisfaction ratings for most public sector organizations were well below the ratings of private sector services. There was concern that online government services would be judged by the service standards and client service orientation of private sector organizations in the delivery of E-services and so a very strong client orientation for GOL was adopted.

Other research among business showed that Canadian businesses rated the Internet far lower as a priority than US businesses and there was a concern that Canada would not get the share of global e-commerce that would ensure the competitiveness of our economy in world markets.

One of the most significant findings of this research for Statistics Canada was that Canadians placed 'completing government surveys and questionnaires' as their second most important use of the online channel after tax filing.

Phases of Government On-Line Implementation

This graphic illustrates a planned phased approach to the GOL initiative which would take advantage of lessons learned along the way and apply them to subsequent activities.



The horizontal axis denotes the type of on-line interaction and the vertical axis denotes integration among service delivery agents.

Tier One was meant to establish the federal government's on-line presence by putting key departmental and program information on-line and making it accessible either directly through a department site or through a revamped Government of Canada Portal. The target for this phase was December 31, 2000 and generally it was met. Statistics Canada had achieved this target about two years before that date.

Tier Two represents a significant step-up from Tier One. This second tier is essentially the delivery of end-to-end secure 'transactions' for all key programs and services by December 2005. For Statistics Canada, transactions also include data collection activities. The words

'service transformation' characterize this stage – the fundamental redesign of service delivery from a client needs perspective to capitalize on the inherent benefits of the Internet.
Tier Three involves inter-jurisdictional service delivery and a variety of pilot projects are already underway to foster partnerships and the cross-jurisdictional integration of services from different levels of government - another new paradigm. I'll be mentioning some pilot projects that Statistics Canada has been involved in.

Departmental 'Key Services'

Each Department/Agency was required to developed a GOL plan for each of its 'Key Services' and for Statistics Canada these are the three key Service we identified.

Collection: Collecting data from individual citizens, households, institutions and businesses as part of census and survey programs undertaken by Statistics Canada.

Communications & Dissemination: Serving information users via the news media, with standard products, the Internet, custom services and our distributor network with outputs of statistical programs.

Stakeholder Relationships: Managing relationships with key interest groups and constituencies with whom Statistics Canada has strategic alliances, e.g. associations, provincial agencies, education, data researchers.

You'll notice that we did not identify programs such as 'Census' or 'National Accounts' as key services. The functional approach we chose provides both a highly simplified way of describing all of the Agency's client relationships and also an effective way to plan and implement our online activities in an integrated way.

While only one of these key services has information dissemination as its principal focus, the other two – collection and stakeholder relationships – have strong dissemination components as well.

As a starting point we developed a strategy paper for each key service on the opportunities that the on-line channel presented for each service's constituency. These formed the basis for a corporate plan that we produced for the Treasury Board and which we continue to update. The corporate plan is a template based document which allows Treasury Board to compile an overall government plan.

The balance of my presentation will be about the 'service transformations' occurring in each of these key services and how they are fundamentally changing the relationships between Statistics Canada and its clients. In the process, a number of new paradigms should become obvious.

The 'Communications and Dissemination' Key Service

Our Communications and Dissemination key service has already achieved the GOL Tier Two 2005 goal of service transformation – a fundamental re-engineering of our dissemination services from a client perspective. The key elements of this transformation include:

A Corporate Data Warehouse: At the heart of our dissemination strategy is a corporate data warehouse - CANSIM II - which includes virtually all of Statistics Canada's published information and is the source from which much of the other content of our web site is dynamically updated. Since its launch two years ago the number of time series has grown from 800,000 to approximately 11 million.

All Publications Available On-Line: With a small number of exceptions all tabular and analytical publications, methodology papers, user guides and research papers are available on-line – primarily in PDF.

Official Release On-Line: The DAILY, our official release publication for data and products, has over 7,000 subscribers and in the near future subscriptions to it will be available for 28 'themes' – health, employment etc, - as a first step towards more specific personalization.

Daily Updates: Over 450 National/Provincial tables in the Canadian Statistics module are updated on a daily basis and most are linked to Statistical Data Documentation, (Meta Data).

Community Profiles: Profiles of 6,000 Canadian communities now include Census and Health information and other social data will be added.

On-Line Catalogue: There is a comprehensive Online Catalogue and products descriptions are linked to our Integrated Meta Data Base which describes the statistical survey where the information originates and the underlying concepts.

E-Commerce: The site has included E-Commerce since 1997 and total revenues in 2001-02 were approximately equal to the cost of maintaining the site.

Integration of Service Delivery Channels: A 'Contact Us' button is included on almost all site pages which provides users with a range of access options including toll free telephone and email. The latter are received by our Advisory Services group and answered directly or routed to the appropriate subject matter or other contact for direct response. Last year over 30,000 email messages were answered and are themselves an excellent source of client research on information needs and navigational issues. Standards of service for all service channels including custom services are published on the site.

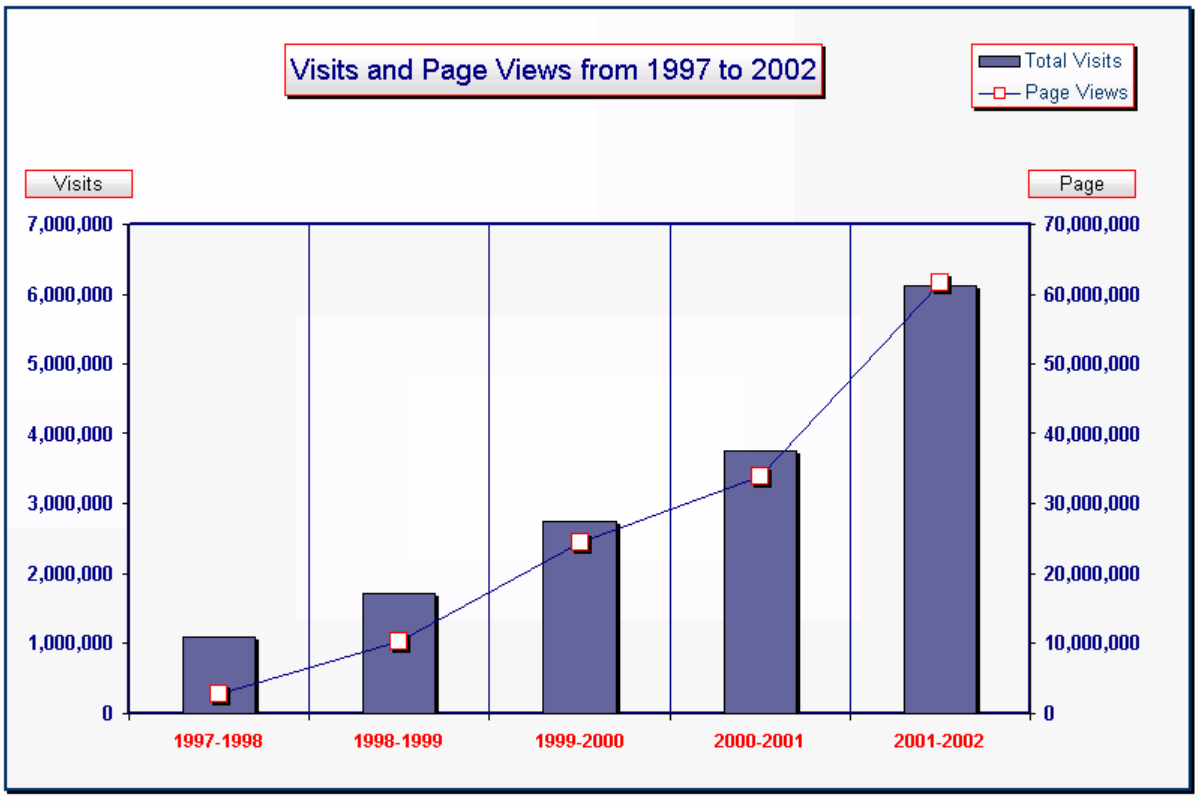
Common Look and Feel: All Federal Government sites follow a set of strict guidelines that give them a common look and feel. This benefits users who develop a familiarity with the type of information located in each area of pages on Federal Government sites, (common toolbars, navigation features, etc), and contributes to ease of use for visitors. Although many sites initially

resented the limitations these standards placed on creativity and their ability to have a unique look, most would agree there is still sufficient latitude for individual 'branding' and the users benefit from the common design elements of government sites and within sites.

Client Focused Site Development: The development of our site has been guided by research with visitors since its inception. We have conducted a number of online surveys with site visitors, focus testing, observational research and testing of particular products by closed user groups.

Growth in Internet Traffic

Site traffic has grown steadily – by over 50 percent in 2001-02 – with over 6 million visits last year. The following chart illustrates the pattern of growth in visits and page views we've experienced. In part the growth can be explained by the general increase in Internet use among information users but there are a number of other reasons. We've promoted the republishing of content from our site with the condition that those doing so provide a link back to www.statcan.ca . Today there are over 10,000 pages from 3,000 sites indexed in AltaVista that link to our site. We've also invested heavily in registering our pages with the most widely used search engines so we come up high in search results. And we do a significant number of other awareness creating activities as well.

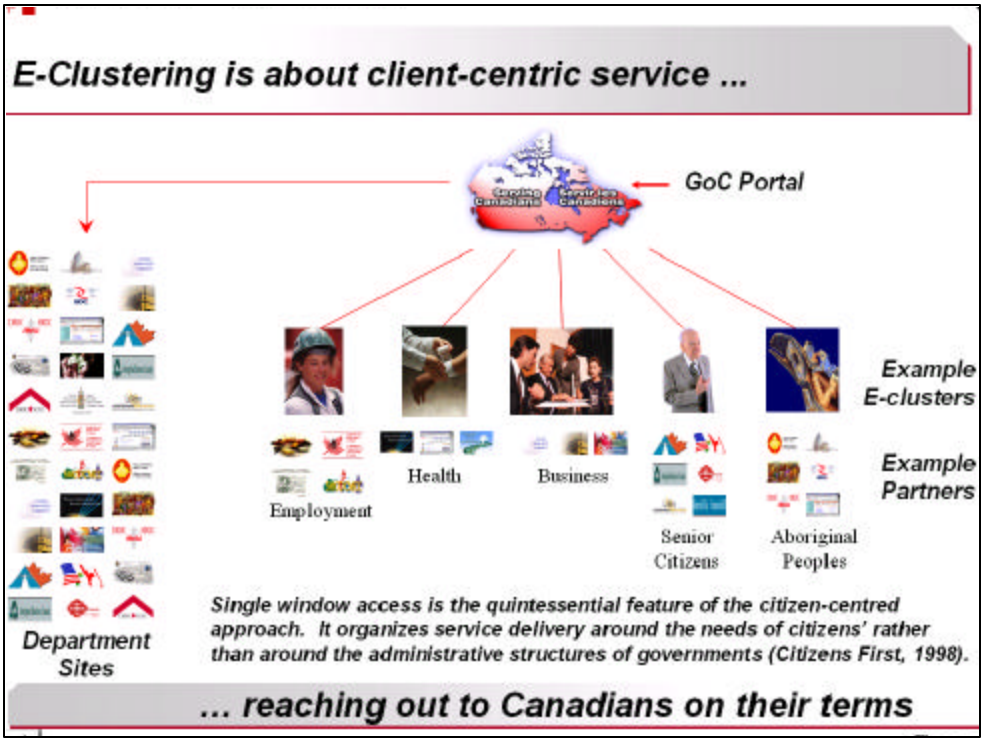


'E-Clusters' – A citizen Centered Approach to Government Services

E-Clusters are one of the core elements of the Government On-Line initiative as they allow citizens to find information and services without having to understand the structure of government. E-Clusters are single entry points to information and services on a common theme which are provided by a number of Departments and Agencies and they are accessed through the Home Page of the Canada site, www.canada.ca .

Statistics Canada participated in the development of the E-Cluster concept, particularly in the market research to determine the categories of information and services sought by three major client groups; Canadian Citizens, Businesses and International Visitors. The Canada site with these three 'gateways' was launched in January 2001 with 35 E-Clusters.

The following graphic illustrates the concept.



Statistics Canada will play the lead role in developing two E-Clusters:

- **Economy** – which involves three partner departments and the Bank of Canada provides information on Canada's economy in relation to other countries and is designed for citizens rather than specialists in this area , and

- **Business Information and Statistics** – which involved ten partner- departments and is aimed at small and medium size enterprises to improve the success rate of new start-ups and enhance the international competitiveness of Canadian Business.

E-Clusters in Action

The home page of the Canada site is found at www.canada.gc.ca which provides access to thirty five E-Clusters grouped in three categories or Gateways:

- Services for Canadians
- Services for Non-Canadians
- Services for Canadian Business

Today approximately 6 percent of our site traffic comes through these portals. This compares with 34 percent from search engines.

If you click on Services for Canadians it will bring you to a listing of topics organized by Subjects and Audiences. Because of the range of information Statistics Canada provides, we expect that eventually almost all E-Clusters will have links to our site which will provide many more pathways to our content.

If you click on Economy from the Subjects list you reach the home page which Statistics Canada created in partnership with four other Agencies, Industry Canada, Foreign Affairs and International Trade, Agriculture Canada, Finance Canada and the Bank of Canada.

If you visit this site you'll notice that the information created for this site is designed to inform the average Canadian about the performance of the economy. More typically our users are economists and policy planners but this site is targeted to a broader audience and provides a great deal of information on economic concepts as well as a time line of key economic events.

The home page of the Economy E-Cluster includes a number of key economic indicators. These indicators are updated dynamically from Statistics Canada's corporate data warehouse, CANSIM, whenever it is updated. This is the first table that is dynamically updated outside Statistics Canada's web site but many more are anticipated.

If you click on Current Economy you'll get an idea of the range of information available from the partner Agencies. Among other information it includes:

- A Quarterly newsletter from Finance Canada, 'The Economy in Brief'.
- Monthly Analysis from Industry Canada which provides more detail on trends within industries,
- The Statistics Canada Daily links to the home page of www.statcan.ca
- Other headings such as Families and Workers have links to Canadian Statistics tables on www.statcan.ca.

We will not promote the site until we have completed the first visitors research study. We have, however, registered it with most of the major search engines and it is listed first when the search term 'Canadian Economy' is used in Google.

The site is expected to get a great deal of visibility when major economic announcements are made such as a Federal budget or at pivotal points in the performance of the economy, e.g. entering a recession or a recovery and will provide information on these topics for the average citizen.

A Business Portal within Our Site

Research with visitors to www.statcan.ca over time has shown some under-representation of business information users. Focus testing has revealed a preference by businesses for a focal point or portal providing links to information of interest to them on the Statistics Canada site. The GOL initiative presented Statistics Canada with the opportunity to create a Business Data portal which can be accessed from the left side tool bar on our home page or from the Business Gateway on the Canada site.

Further focus testing during the development of the Business Data page revealed that users wanted both a thematic access to information, Browse our Comprehensive List of Business Topics, as well as organization of content around key business activities, e.g. Obtain Trade Data for Canada and Abroad.

The page was launched in October 2001 and has surpassed the traffic forecast. We have conducted some research with site visitors and they have given the concept favourable ratings but want more content added to the site, particularly organization of information by industry so they can compare their firm's performance to their industry and geographic comparisons.

The page also includes the top 10 business information products as well as the same key indicators that appear on the homepage of www.statcan.ca. Visitors also indicate they would like to see indicators more directly related to business activity in this area.

We will continue to develop Business Data with additional content and will consider adding links to sites of other Federal Agencies with relevant content and possibly to provincial sites.

The Data Collection Key Service

Our Data Collection key service is at a much more preliminary stage of development than Communications and Dissemination. While there have been some early business survey experiments they were not truly online activities and required downloading of an application or questionnaire, completion off-line, encryption and then transmission. In general take-up rates were low.

There was also a small test conducted in two municipalities within the 2001 Census and take up was also low for similar reasons.

Statistics Canada's approach can best be described as cautious because of the many unknowns associated with electronic data reporting. Certainly, in the initial stages, it will be an additional channel creating the uncertain impact of mixed methodologies on data quality.

We have obtained funding from our Treasury Board to create an online response option for 60 business surveys and one household survey by 2005. The surveys selected are mainly monthly

and quarterly surveys with relatively few questions and respondent communities that are highly connected to the internet therefore offering the greatest potential to maximize take-up rates.
The Electronic Data Reporting project will also create a Personalized Reporting site for a small number of very large businesses to provide them with information on the surveys they will be asked to complete, assist them with managing online reporting and provide a focal point for respondent support.

The 2006 Census will draw on lessons learned from these initiatives and will be implementing an online response option throughout Canada. Census management have set an operational target of 25 percent response for online response.

Respondent Research

We recently conducted a study of households and business, which had just completed a Statistics Canada survey, to better understand respondents' readiness and willingness to use online response. Combined, about 85% of respondents had Internet access at the location where they completed their survey. About 80% of those who had Internet access said they 'definitely' or 'probably' would have used an online option to complete their most recent survey if it had been available to them.

They would only have used online, however, if it had been more convenient, more efficient and they were assured that there could be no unauthorized access to their information. Security of their information was the most important decision factor.

Online response is not a question of 'if' but 'when'. Certainly businesses who are using the Internet to manage supply chains for reasons of efficiency and who are able to do e-filing of tax returns will have growing expectations that survey questionnaires can be completed online as well. Households will value both the convenience dimension of online as well as the improved security online should eventually offer.

Earlier studies of factors that would motivate respondents to participate in surveys - particularly businesses - included access to the survey results. This expectation is expected to increase with the use of online reporting. Providing a business with a profile of how the firm compares to its industry and with access to other relevant data useful to its decision-making will not only motivate participation in surveys but should also improve the quality of response. Other timeliness and quality improvements are possible if respondents can link survey templates on personalized web pages with their own electronic information systems.

These features of online data reporting should present the opportunity for Statistics Canada to transform survey participation from an onerous activity - based on legal obligation - to one that is advantageous for respondents. Our goal must be to find that new 'value proposition'.

The Stakeholder Relations Key Service

The Stakeholder Relations key service could be included as part of our other two key services, Communication and Dissemination or Data Collection. However, we decided it would be useful to identify a category of activities that we conduct in order to improve relationships with the interest groups and constituencies with whom Statistics Canada has strategic relationships, e.g. business associations, provincial agencies, the education and researcher sectors. The following are brief descriptions of some typical initiatives.

Education Community Liaison Program: Statistics Canada has made a strong commitment to the use of Canadian information and data in Canadian classrooms and academic research. The Educational Community Liaison Program includes the development of a Learning Resources module on www.statcan.ca and the creation of an Education Account Executive position in each of our regional offices. These resources work with teachers and schools, educational publishers, faculties of education that train teachers, and with school boards and ministries of education to encourage the use of statistics Canada data in teaching activities. The majority of this information, including teacher developed lesson plans and curriculum guides, is provided via the Learning Resources module on our site.

The Data Liberation Initiative: The Data Liberation Initiative was created to provide access to all of Statistics Canada's published electronic databases and public use micro data files for research and teaching purposes in Canadian universities. All have now joined the program at a fee which coves its cost. Electronic files are distributed to data librarians via the Internet and a very active user community has evolved sharing information on the holdings, again via list serves and other Internet communications.

Pilot Inter-jurisdictional Projects: Several pilot projects were funded through the GOL initiative which have been completed and are now being evaluated to assess the potential to apply lessons learned in other program areas.

- In conjunction with Health Canada, online training materials were developed for local health professionals to support the use of data for local decision making;

- Synthetic micro data files of education data were made available to researchers via the Internet which allowed them to specify tabulation requests from unpublished data, to submit them and have confidentiality screened results returned online in order to minimize the normal time requirement; and

- A secure communications channel was established to collect justice information and enable pre-release reviews by the justice community including local police departments.

The findings of the pilot studies will be available in the Fall of 2002.

Increasingly the Internet will be used to manage relationships with key stakeholders.

Census Consultations: For the 2006 Census we have planned a two-stage process to simultaneously discuss Census content and outputs with data users. In the first phase we will provide information materials through traditional channels and offer a range of options to

provide in depth recommendations and feedback as a second phase. There will be at least one pilot test of obtaining this input online through a 2006 Census consultation web site.

On-Line Advisory Committees: Statistics Canada has 22 subject matter Advisory Committees and the National Statistic Council which guide our programs. Later this Fall Statistics Canada's GOL working group we will contact the secretaries and chairpersons of these committees to identify a small number to test online consultative processes using extranets and closed user groups. Again, the GOL initiative has developed some standardized approaches and tools for these types of consultative activities and we will use these in the test.

Respondent Relations and Research: A critical element of the success of the electronic data reporting project, (EDR), will be the provision of information to prospective respondents related to their key concerns such as security and confidentiality, and the convenience and efficiency of the process. As well respondents must have an online single point of access to support, links to survey results and other data related to their interests. Research on respondent relations in support of EDR will be conducted in conjunction with the 11 surveys which will begin to offer an online response option later this Fall.

Dynamic Updating of Other Sites: there has already been a large increase in the number of organizations wanting to republish Statistics Canada data on their sites and it will continue to grow. Tables on our site are dynamically updated whenever the CANSIM II database is updated. The Key Indicators table on the Economy E-Cluster is the first instance of this process being used for a table on another site. This process will be actively promoted as it ensures that wherever STC data appear they are consistent and will be accompanied by a link back to www.statcan.ca.

Recruitment: An Employment Opportunities module has been added to www.statcan.ca to provide information on the full range of recruitment initiatives which generate the majority of our new professional, technical and social science support staff. This module will evolve to provide more of the primary screening of applicants to streamline the process.

New Data Dissemination Paradigms

What are the new paradigms for National Statistical Offices in an E-Government world?

One Stop Data Shopping: First our web sites must be comprehensive repositories - enabling information users to access all of our published data online – and our research shows that effective search is the critical factor in successful access to content and finding the information sought is the key determinant of visitor satisfaction with their site experience. As well all information must be linked to the underlying meta information for users to fully understand the concepts and the processes through which it was created to use it effectively.

Dynamic 'Database Publishing': Because of the huge amounts of information available from our sites their overall integrity must be ensured by updating processes that, to the extent possible, minimize human intervention. Otherwise the cost of maintaining a comprehensive site is prohibitive. Today most NSO sites are driven by linked databases, (multi-dimensional tables, meta information, analytical text, catalogues, etc), which allow data to be presented in a variety

of formats. When a database is updated it automatically triggers the updating of information throughout the site so there is consistency.

Personalization: To build effective relationships with site visitors, they need to be able to identify which topics are of interest to them and be notified of the availability of newly released information related to their interests or have it automatically sent to them. Demand for personalized services will grow quickly.

Single Points of Entry: Information users expect Government Portals or Gateways to provide access to information and services from many sources without having to understand the structure of government. The E-Clusters do this effectively across Departments, and within departmental sites users expect to be able to search thematically and to have other integrative mechanisms such as our Community Profiles and our Business Data modules to integrate content across statistical prgrams.

Branding: Our data will be republished, so we must provide the tools for other sites to provide appropriate sourcing information and to create links back to our sites which are more comprehensive and current. If we are not identified as the source of our data, respondents will not see the value in participating in our surveys. We also need to do more public opinion research to understand more about how households and business perceive our brand – to know more about our 'brand equity' to help us develop more effective communications programs.

Respondents are Clients Too: We need to use information outputs to create a new 'value proposition' for survey respondents to motivate them to provide high quality input to our surveys. We must apply the same marketing principles to electronic data reporting we have applied to our dissemination activities so we re-engineer them from a client perspective.

Online Partnerships: Build online partnerships with key stakeholders - groups that play roles that sustain our core mission. Closed user groups, Extranets and online consultative processes help to build relationships.

Apply Marketing Principles: E-Government is a client focused process. Know your clients, listen to their messages and act on them.

Don't Re-Invent the Wheel: And finally, build relationships with your international colleagues and share best practices. There are likely many people who are also working on 'your great idea'. Develop a network – you may even get to travel.

E-Government Information Sources

Although the title of this session and my paper refer to E-Government, most of what I have talked about is really the use of the Internet to deliver our organization's information and services – what we call Government On-Line. This is occurring in all of the developed countries in the world and I'm pleased that we're sharing our experiences much as we have with output databases.

E-Government is a much broader concept that is also being studied in democratic countries around the world. New communications and information technologies make many of our existing institutions and their focus irrelevant as connectivity erases organizational boundaries and even national boundaries. It also permits a much broader participation in policy development processes and increased transparency and accountability in government.

Statistics Canada has participated in funding an initiative called Crossing Boundaries which explores these opportunities, in part because of the key role played by information in the policy development process and because there is a growing perception that information is an essential public resource in this new paradigm.

We have had one presentation for our senior management community on the first report, 'Realigning Governance: From E-Government to E-Democracy. If this is a topic of interest, you can register to receive their newsletter URL and any of their reports at www.crossingboundaries.ca .

Finally, I want to mention a report that was prepared by Andersen Consulting called the Accenture Report. It is their third annual assessment of E-Government in 23 countries.
Their assessment model includes ratings for 'service maturity' and 'customer relations management' which are combined to give an overall rating for each country. Service maturity measures the breadth of services available online plus degree of completeness. Customer relationship management measures the level of service sophistication. An electronic copy of the report can be obtained at www.accenture.com

Our Treasury Board has adopted this model to assess the performance of Canadian Federal departments and agencies in E-Government.

Thank you for the opportunity to participate in the conference. Please contact me by email if you would like further information on any of the topics in this presentation.

FedStats—Statistical Information Dissemination in the 21st Century—
The Next Generation

Valerie Gregg And Marshall DeBerry
FedStats Interagency Task Force FedStats Program Manager

Preface

Citizen Access to Federal Statistics Scenario 2020[5]

Individuals want access to federal statistical data. They wish to learn, for example, the demographics of different areas (e.g., information about schools, cost of living, recreation), what is going on in business and agriculture, what is driving prices in a particular area, or what to expect with regard to inflation and interest rates.

How far have we come today toward realizing this vision? FedStats provides a single portal for federally collected data sets and for documents based on that data. Data sources and documents are organized topically and geographically across all the federal statistics agencies. In many cases, the available data are constrained, owing to confidentiality protection, but summary information and reports may be available. Still, one cannot make such queries as, How many people will be displaced if an evacuation at the 100-year flood line for Manhattan, Kansas, is required? Or, what would be the economic impact of locating a particular new business in my town?

Imagine asking FedStats the latter question in 2020. This might trigger a series of questions back to the user not only to acquire more details about that business but to learn more about that user: his or her quantitative/scientific literacy and visual/verbal/textual/cognitive abilities. Then, the relevant data, complemented by additional data sources where needed, would be "crunched" with the aid of models and simulations. A response containing the requested information both fully and in user-friendly form would quickly be returned to the individual making the query.

To realize this requires IT innovation on several fronts, such as representation of information, archiving and searching, modeling and simulation, and information integration. Subtle but important issues, such as the underlying integrity of responses, will also become key. For example, when people of varying degrees of quantitative sophistication ask the same basic question, answers must be consistent.

Taking the scenario one step further: imagine being able to get a second opinion. The local chamber of commerce has contracted with a small economic modeling company to give you access to a model that uses a different set of assumptions. Running this model using a portal to the company offered by the chamber, the model accesses the same underlying census and economic data that were used in the government's model. The

[5] This scenario is paraphrased from Appendix A, "E-Government Scenarios" of the National Academies of Science, National Research Council's Computer Science and Telecommunications Board (CSTB) May 2002 report entitled *Information Technology Research, Innovation and E-Government.* The full report is available on-line at http://books.nap.edu/html/itr_e_gov/.

modeling company's software is able to access the underlying government databases directly, using an application programming interface offered by the government to allow non-government computer programs to analyze the data in new or different ways.

This paper, prepared for the Federal Committee on Statistical Methodology's Statistical Policy Seminar Challenges to the Federal Statistical System in Fostering Access to Statistics "FedStats—Statistical Information Dissemination in the 21st Century—The Next Generation" will provide one perspective on bringing this scenario to fruition.

Introduction

FedStats is a major success story and an exemplar for interagency, multi-sector partnerships. The award-winning website not only exceeds the initial objective as defined in 1995 by the Interagency Council on Statistical Policy (ICSP), it is now rapidly becoming a demonstration environment for new technologies that will enable the entire Federal statistical community, as well as individual agencies, to become a leader in "Electronic-Government", or "E-Gov" implementation.

This paper provides an historical perspective on how FedStats evolved, how FedStats will continue to evolve within the E-Gov context, and the role FedStats will play in near, mid- and long-term statistical information dissemination in the 21st century. FedStats will help lead statistical agency dissemination efforts towards realizing the 2020 Scenario described in the preface.

Background

The United States Federal statistical system is decentralized, with individual agencies having statutory responsibility and authority for statistical activities. Hence, it is difficult for the general public, and even frequent data users such as social science researchers, to know about and to access the entire wealth of information produced by the Federal statistical system. To address these organizational barriers to accessing Federal data, the ICSP (consisting of the agency heads of the 14 largest U.S. statistical agencies), under the leadership of the Chief Statistician of the United States, Katherine K. Wallman, launched *FedStats* in May 1997. Prior to the public launch, the FedStats Interagency Task Force had been working together since the fall of 1995 to design and develop a "One-Stop Shopping" or "Virtual Statistical Agency" for Federal Statistics Website.

This interagency web site http://www.fedstats.gov/ now serves as the Internet gateway to the full range of official Federal statistical information available to the public from more than 100 U.S. Federal agencies. FedStats provides a centralized set of links to the Internet sites and the subject-matter data that individual agencies maintain and update. The site's primary objective is to help users find the information they need without having to know and understand in advance how the decentralized U.S. Federal statistical system is organized or which agency or agencies may produce the data they are seeking.

From June 1997 through August of 2002 there have been nearly 8 million user visits to the FedStats site, which represents nearly 25.5 million pages served to visitors. User traffic has increased by approximately 60 percent from 2001 to 2002. The user profile represents a wide spectrum of visitors, ranging from private citizens, academic users, the media, policy makers, and visitors from countries outside the United States. Frequently visited sections of the site include the "Topic Links A to Z" section and the MapStats section, which provides a simple "drill down" capability to retrieve statistical information at various levels of United States geography.

The Task Force reports to the ICSP on an annual basis, providing an annual assessment of the previous year, a set of recommended projects for the coming year and a set of resource requirements. Starting in Fiscal Year 1998, the U.S. Bureau of the Census, via interagency agreements with each of the ICSP agencies, is reimbursed annually for supporting the technical design, development, and maintenance of FedStats. The agreement covers the costs of the FedStats Chief Architect, an additional technical FTE and hardware and software. Until this year, the total cost was $285,000/year (5 largest agencies paying $30,000 each and the 9 smaller agencies paying $15,000 each).

The Interagency Task Force continues to upgrade and expand FedStats coverage and access to Federal statistical sources. Additionally, the Federal statistical community is exploring new information technologies and undertaking research projects in collaboration with the National Science Foundation's (NSF) Digital Government (DG) Research Program to achieve a much broader vision for the future (discussed in more detail in a further section). New technologies and methods being developed as a result of more than 14 NSF DG research grants are helping to guide design and development of the Next Generation of FedStats. (For more information on the DG-FedStats research projects see Appendix II; for more information on the DG Research Program see *http://www.diggov.org*)

Current Features and Capabilities

Over the past five years, FedStats has become "The gateway to statistics from over 100 U.S. Federal agencies". The current features and capabilities include the following:

Links to statistics

- **Topic links A to Z—**Direct access to statistical data on topics of your choice.
- **MapStats**—Statistical profiles of States, counties, Congressional Districts, and Federal judicial districts (drop down list of states)
- **Statistics by geography from U.S. agencies**—International comparisons, national, State, county, and local.
- **Statistical reference shelf**—Published collections of statistics available online including the Statistical Abstract of the United States.
- **Search**—across agency websites.

Links to statistical agencies

- **Agencies listed alphabetically**—with descriptions of the statistics they provide and links to their websites, contact information, and key statistics.
- **Agencies by subject**—select a subject (drop down list of key subjects)
- **Press Releases**—The latest news and announcements from individual agencies.
- **Kids' pages**—on agency websites.
- **Data access tools**—Selected agency online databases.

Other features

- **Additional Links**—to other statistical sites and general government locator sites.
- **About FedStats**
- **Feedback**
- **Federal statistical policy**—Budget documents, working papers, and Federal Register notices.
- **Site privacy policy**
- **Site document accessibility**

Many of these features and capabilities offered at the FedStats including the design of the homepage have evolved over time as a result of usability testing and research on information seeking behaviors. For example, Topic links A to Z, three different experimental versions in addition to the active version on the FedStats website were tested to help determine the best way to present an index of topics. The results of the usability testing helped guide the current design of the index.[6]

Site Architecture

The FedStats site is designed to be robust and flexible in terms of data access and display. Web pages are designed to meet the Federal government requirements for access by the disabled (Section 508 of the Rehabilitation Act) as well as being accessible to the wide variety of web browsers available on personal computers and mobile devices, such as cell phones. Computer hardware that uses the Unix operating system is used for the public portion of the site, and development work is done on computers that use the Linux operating system. Open Source software has been used extensively on the site because it is robust, scalable and a very usable utility for web development. Open Source software is software that is available for use without the payment of royalties or fees to an organization, and may be inspected and further modified as needed by other programmers. A variety of Open Source software tools are used extensively in developing the FedStats site, and including the Linux operating system, the Apache web server, the MySQL database server, and Perl and PHP software code for the development of web pages.

[6] Hert, C.A., Jacob, E.; Dawson, P. (2000). A Usability Assessment of Online Indexing Structures in the Networked Environment. Journal of the American Society for Information Science 51(11): 971-988. The technical report is available at http://istweb.syr.edu/~hert/BLSphase2.html

Ongoing Projects

FedStats Section 508 Accessibility Workshop

Section 508 of the Rehabilitation Act requires Federal agencies to meet specific requirements in making their websites accessible to people with disabilities. Several of the requirements are particularly problematic for the Federal statistical community as they affect tables, statistical graphics, and formulas. However, little attention has been paid to the accessibility of these elements in a statistical context. Given the enormous volume of tables, formulas, and statistical graphics on Federal statistical agency sites, FedStats Interagency Task Force decided to sponsor a 508/Accessibility Workshop on June 24, 2002, to focus on ways that statistical agencies can meet the new accessibility requirements and make their Web content accessible to people with disabilities. The workshop brought together about 150 participants including Webmasters and content managers from statistical as well as other federal agencies, researchers, vendors (assistive technology, Web editors and validators, and authoring tools), standards organizations, and the disability community. Forty Federal agencies were represented. Presentations and related materials from the workshop are available at http://workshops.fedstats.gov.

As a result of the workshop, Interagency Task Force plans to release three papers in the newly established FedStats Working Paper series. The first paper will summarize the workshop proceedings--highlighting the areas in which additional research and work needs to be done. The second paper will offer a recommend implementation of the Section 508 guidelines for tables as a short-term solution to the problems many agencies are facing. And the third paper will propose ways in which the current standards could be changed to better facilitate the accessibility and usability of complex statistical tables.

MapStats for Kids

In August of 2001, the FedStats Taskforce received a $90,000 cash award through a competitive selection process from the e-Government Committee of the Federal CIO Council for the development of a *MapStats For Kids* section of the site. The *MapStats for Kids* project is focused on making Federal statistical information interesting and meaningful to younger citizens and thereby foster the development of statistical literacy. Statistical literacy can be viewed as the ability to interpret, critically evaluate, and communicate about statistical information, conveyed either through numbers or graphics. The GeoVISTA Center and Geography department of Penn State University was selected to work on developing a prototype for a *MapStats for Kids* section of the site based on their past work in geospatial displays of quantitative information. A target audience of fourth to eighth graders was selected as being age-appropriate in the development of the prototype, and the software tool Macromedia *Flash* was chosen to create interactive web applications that would engage the target audience. By presenting young citizens with statistical data and information in an engaging manner, these visitors to the site would be stimulated to further explore and ask questions about the various data series collected and disseminated by the various Federal statistical agencies.

To date, several prototypes have been developed which work towards developing three sets of skills that are central to statistical data analysis: logico-mathematical skills, representational skills, and spatial skills. Logico-mathematical skills can be related to the concept of geo-coding, that, understands the unique representation of units within a hierarchical framework, such as countries, states and counties. Representational skills can be represented by the concepts of understanding symbols on a map—blue for water, black for roads—and the context in which they represent. For example, a black line may represent a road, but due to its small representational size on a map, younger children may view it as not representative of their real-world experiences of what constitutes a road. Spatial skills can be thought of as representational objects, such as the outlines of state boundaries or three dimensional shaded relief projections on a map, and mentally "mapping" them into a context that conveys the underlying meaning. All of these skills are important in the process of manipulating and understanding statistical data. For example, young children may be presented with the current rankings of sports teams located throughout the United States, and using these three skill sets could gain a better understanding of the concepts of averages, regional variations, and the concept of place among various geographic boundaries. As the project progresses, the FedStats Taskforce will continue to evaluate and suggest different strategies that can be utilized in developing these skill set areas, with the goal to have a fully functional *MapStats for Kids* section on the site with the resultant software code available for use by other interested agencies.[7] (For more information about this project see: http://www.geovista.psu.edu/grants/MapStatsKids/index.html)

Outreach and Promotion

The FedStats Interagency Task Force recognizes the need to systematically undertake outreach and promotion activities. While some efforts have included working with the Interagency Public Information Officers, others have included printing brochures and flash cards for distribution by individual member agencies at their respective outreach events. Still other efforts have included contacting members of the news media to feature new FedStats capabilities.

FedStats is represented on the Cross Agency e-Gov Solutions Working Group that is a part of the Government Services Administration' Office of Citizen Services and Communications which is responsible for the First Gov web portal. As a member of the working group, FedStats seeks to share best practices with other portal projects across the spectrum of Federal agencies.

FedStats has garnered the interest of non-Statistical agencies like the Department of Housing and Urban Development and the U. S. Geologic Survey and has collaborated on several projects related to geospatial representations of agency information with FedStats data and applications. Both of these agencies have discussed becoming official members of the Interagency Task Force, and have in the past, contributed towards the design, development and implementation of the MapStats project.

[7] Paraphrased from the MapStats for Kids - Phase I Report; PI Alan M. MacEachren et. al; GeoVista Center and Geography, Pennsylvania State University, July 30, 2002, page 3.

Another manner in which FedStats promotes itself is to enter competitions for recognition, some of which award funds to the winners. For example, FedStats was awarded $90,000 by the CIO Council for development of the "MapStats for Kids" project.

New Project

Improving Automated Access to Statistical Databases

Most federal statistical agencies provide user access to electronic databases and data files through their Internet websites. This is a valuable service that users of statistical data rely on and use routinely. There are, however, many users for whom the web browser interface to federal statistics does not fully support their data access needs. These "power" users are those who: (1) regularly download many databases and data files; (2) regularly download data from several agencies; (3) need downloads of entire databases; or (4) need to maintain timely subject-area databases using the most current statistical releases from one or more agencies. Ironically, this user community includes many federal agencies that use federal statistics as input to their own programs (e.g. economic analysis).

Existing technologies are available to provide power users with automated, computer-to-computer, data exchange through the Internet, but there are several roadblocks to their implementation that the Interagency Task Force is in a unique position to resolve. Among these obstacles is the lack of a standard protocol for automated data exchange. The Interagency Task Force is forming a working group to begin addressing this problem and plans to draft a protocol for exchange of non-confidential data for prototyping and testing. This protocol will be based, in part, on the method used to maintain the White House Federal Statistics Briefing Room. An additional obstacle, when a standard protocol is available, is the need for a registry of statistical agencies that support the protocol and the data they make available through it--a role parallel to the role that http://www.fedstats.gov/ now plays for statistical agency websites.

FedStats Within The E-Government (E-Gov) Context

During the past several years, as new information technologies have proliferated and been applied to government operations and services, the public's expectations for ease of access and use of government information and services has increased. "E-Gov" initiatives have assumed a much higher profile within the Federal Government. While agencies have increasing E-Gov demands, there are little or no new resources to implement E-Gov applications. However, for a rather small investment, the Federal statistical community is well positioned to continue building valuable E-Gov services by leveraging the various research and development collaborations being undertaken by the FedStats agencies and their public and private partners. These types of collaborations save individual agency from having to do E-Gov all by themselves. FedStats has often been cited as an exemplar for providing valuable E-Gov information services to the public.

In 2001, the Interagency Task Force conducted an intensive strategic assessment and planning process, taking into account various E-Gov Directives and initiatives issued during the Clinton Administration. The outcome was a newer, more comprehensive strategic plan with a mission,

vision, and strategic goals that would enable FedStats to move well beyond a simple, yet highly acclaimed, award-winning portal web site towards the Next Generation FedStats. A year later the mission, vision and goals remain entirely consistent with the more detailed E-Government vision outlined by the Bush administration.

"Expanding E-Government" Initiative

Mark E. Forman, the Office of Management and Budget's Associate Director for Information Technology and E-Government issued on February 27, 2002, his E-gov strategy report entitled Implementing the President 's Management Agenda for E-Government—Simplified Delivery of Services to Citizens. Information on this E-government effort may be found on the Internet at, http://www.firstgov.gov.

In the report, several key goals and strategies that are most relevant to the FedStats mission include (emphasis added)--

> "Among the primary goals in the President's "Expanding E-Government" initiative are to **make it easy for citizens to obtain service and interact with the federal government**; improve government efficiency and effectiveness; and to improve government 's responsiveness to citizens."
>
> "Effective E-Gov strategies will result in significant improvements by, among other things "simplifying delivery of services to citizens; **making it possible for citizens, businesses, other levels of government and federal employees to easily find information and get service from the federal government**; and by simplifying agencies' business processes and reducing costs through integrating and eliminating redundant systems."

And, on providing opportunities to transform delivery of government services, the report provides the following guidance:

> "**Build easy to find, easy to use, one-stop points-of-service that make it easy for citizens to access high-quality government services**."

The report concludes that the E-Gov pay-off will not result from automating current processes, but rather through the:

> **"…transformation of how the government interacts with its citizens and customers**. Only through changing how we do business internally —that is, streamlining work processes to take advantage of modern IT systems —will citizens experience the transformation envisioned."

FedStats is entirely consistent with Forman's E-government strategy and is clearly evident in the FedStats mission, vision and strategic goals:

Mission Statement

To provide effective, efficient, and timely access to, and use of, the full range of Federal statistical information needed for informed decision-making.

Vision

Informed decision-making starts with the information and knowledge available through FedStats.

Strategic Goals

- To provide Federal statistical information/knowledge effectively, efficiently, and in a timely manner.
- To enhance the effective use of statistical information.
- To provide an organizational framework and resource base in order to achieve the FedStats' mission.
- To foster broad collaboration that can strengthen the statistical system.
- To be widely recognized as an essential resource and knowledge base for informed decision-making.

To effectively accomplish the mission, vision, and goals, the Interagency Task Force and the FedStats website will have to continually evolve. While the Interagency Task Force remains a collection of involved and committed agency representatives meeting on a monthly basis, the actual infrastructure is becoming more substantial and agile because of several factors noted below.

The Interagency Task Force recognized that ICSP agencies needed assistance in leveraging and/or making operational, in a more systematic and beneficial manner, best practices and approaches for statistical information dissemination, methods, and new technologies. These might be developed within the FedStats environment, or might be those innovations being developed in individual agencies and/or by academic researchers collaborating with statistical agencies via NSF's DG Research Program.

In September 2002, the ICSP agreed with the Interagency Task Force's recommendation to hire a full-time program manager for FedStats and to fund the position by increasing individual agency contributions. to cover the costs of a full-time FedStats Program Manager. The total FedStats budget in FY 2003 will be $470,000 (5 largest agencies contributing $50,000 each and the 9 smaller agencies contributing $25,000 each).

Next Generation FedStats

FedStats will continue to be a premier E-government portal. So what is the Next Generation of FedStats and how might it differ from the current portal?

The Next Generation FedStats will be a national distributed statistical digital library with tools for information finding, for information extraction and reuse, information visualization, and for transforming knowledge into intelligence while maintaining the privacy and confidentiality of respondents. To achieve this vision, FedStats will require common user interfaces, data access and searching tools usable by persons with different levels of computer and statistical literacy, which enables appropriate uses of the data with analysis within and between databases.

The current decentralized, independent sources of statistical information have few commonalties in terms of concepts and definitions; system architectures, software, and hardware; measurement methods; interfaces; or dissemination and presentation modalities. Interoperability is a major hurdle in a variety of areas. Data integration issues abound. Significant challenges in high-end computing and computation and large-scale networking exist for the making the Next Generation FedStats vision a reality.

Computer and information scientists will solve some of these challenges, while others will require a more multidisciplinary, multi-sector approach. For example, involving mathematical statisticians with expertise in creating estimates from complex sample surveys, building small area estimation models, and estimating measures of error for the resulting estimates that incorporate all sources of error, including those due to sampling and nonsampling errors.

If the metadata needed to interpret and use statistical information are to be made available and integrated with the data, the processes and procedures for collecting and compiling statistical information must also be the focus of information technologies research and development efforts.

As one of the first set of Federal agency partners with the NSF in its Digital Government program over four years ago, the statistical agencies have improved upon their historical tradition of being in the forefront in exploring new and novel ways to better handle the ever-increasing volume of data that flow from the varied statistical programs of the U.S. government. In turn, the NSF and the research community have recognized that the Federal statistical agencies have a unique challenge in ensuring that statistical information is collected and provided to the public in as robust and reliable manner as possible, while ensuring that cost-efficiencies are achieved.

Digital Government Research Projects

Over the past five years, the statistical community has taken the "longer-view" on how to improve the Federal Statistical community's data and information dissemination programs. The NSF's Digital Government Research Program is providing government agencies with unique opportunities to better understand what new information technologies are being developed in university research labs and to participate in test bed applications development along side the researchers funded being funded by NSF. Many opportunities exist for leveraging these research efforts (and those yet to be defined) that could lead to radical improvements in agency business practices as well as improving government information services.

In February 1999, the National Academies of Science's Committee on Computing and Communications Research to Enable Better Use of Information Technology in Government, chartered by the National Research Council's Computer Science and Telecommunications Board (CSTB) and the Committee on National Statistics (CNSTAT) held the second of two workshops as part of a larger study being undertaken at the request of the DG research program. The workshop focused on the Federal statistics application area.

> "Underlying the presentations and discussions at the workshop was a desire to tap IT innovations in order to realize a vision for the federal statistical agencies. A prominent theme in the discussions was how to address the decentralized nature of the US national statistical system through virtual mechanisms. The look-up facilities provided by the FedStats Web site are a first step toward addressing this challenge. Other related challenges cited by workshop participants include finding ways for users to conduct queries across data sets from multiple surveys, including queries across data developed by more than one agency—a hard problem given that each survey has its own set of objectives and definitions associated with the information it provides."[8]
>
> Further, the workshop identified a broad range of IT issues for engaging the information technology research and federal statistics communities in research activities of mutual interest. These include human computer interaction, database system, data mining, metadata, information integration, and information security. Two other challenges of particular interest include survey instruments and the need to limit disclosure of confidential information. [9]

In the convening years, the NSF's DG program has funded more than fourteen FedStats research projects. These projects examine such topics as privacy and confidentiality issues in microdata files, new ways to display information contained in statistical tables, tools and methods for automatically building metadata, testbeds for distributed architectures that enable data integration, data collection technologies such as those involved in the use of handheld devices and wireless data transmission, data visualization and validation technologies, etc. Now the challenge is transferring knowledge and/or technologies from the research labs to productions systems. The FedStats environment now that there will be a more permanent infrastructure can help with the transition of results to the FedStats website and or to individual statistical agencies. (See Appendix II for more details)

NSF's digital government grantees have received over $10 million to focus on FedStats-related research. The ICSP agencies have already augmented the NSF awards by approximately an additional $2 million.

---

[8] *Summary of a Workshop on Information Technology Research for Federal Statistics*; Computer Science and Telecommunications Board and the Committee on National Statistics; National Research Council. The full report can be found at http://books.nap.edu/html/itr_federal_stats/

[9] *Summary of a Workshop on Information Technology Research for Federal Statistics*; Computer Science and Telecommunications Board and the Committee on National Statistics; National Research Council. The full report can be found at http://books.nap.edu/html/itr_federal_stats/

In the DG Research Program's recent announcement, proposals are being accepted for two classes—

> 1) Multi-disciplinary and multi-sector partnerships of researchers in information technologies and government agencies at all levels in order to foster collaboration among societal sectors, and
>
> 2) Research on the relationships between the design and use of information technologies on: i) forms, processes, and outcomes of democracy, ii) government organizational forms, learning, and adaptation, iii) new forms of government-government collaboration, iv) citizen/government interaction, and v) other social and political science research related to IT and government.

This second class of proposals, in addition to the first class, which FedStats has leveraged quite well, will enable scientists to better identify and understand the government and citizen user needs for the Next Generation of FedStats. This is an untapped opportunity ripe for further exploration by the Federal Statistical community.

FedStats Interagency Research and Development (R&D) Working Group

In addition to the FedStats Interagency Task Force, in 1997 the ICSP authorized a FedStats interagency R &D working group. As a first step, the working group identified common challenges facing many statistical agencies that could potentially be overcome by applying cutting-edge information technologies. The FedStats R&D working group coordinates the Federal agency responsibilities and activities (along with the academic researchers) as outlined in each DG research proposal. The FedStats R&D working group also is fostering new and/or modified FedStats R&D partnerships that will continue to develop research proposals for submission to the wide array of NSF and other Federal agency research programs.

Conclusions

In seven years much has been accomplished for laying the frameworks for statistical data dissemination in the 21st century, both within individual agencies and by interagency efforts such as FedStats. However, to realize the Next Generation of FedStats much remains to be done.

As noted in the May 2002 CSTB report *Information Technology Research, Innovation and E-Government* [10]

> "A number of these portals represent a fairly mature realization of present-day information-access technology, but considerable scope for improvement remains."

[10] National Academies of Science, National Research Council's Telecommunications Board (CSTB) May 2002 report entitled *Information Technology Research, Innovation and E-Government*. The full report is available on-line at http://books.nap.edu/html/itr_e_gov/.

> "At present, much of the thinking about e-government focuses on what can be delivered with today's technology…But it is also essential that, in looking ahead, planners contemplate how both technology and user expectations will evolve."

The ICSP, the FedStats Interagency Task Force, and the FedStats R& D Working Group are looking ahead, trying to bring the Next Generation FedStats to fruition. Partnering with the NSF academic community is one way in which strategic understanding of new technologies can most effectively be put to use in bringing the best tools, technologies, and policies into practice. Technology transition will remain a challenge, but the FedStats environment, with a solid infrastructure in place, is well positioned to make the vision "Citizen Access to Federal Statistics" in the 21st century a reality. [11]

[11] National Academies of Science, National Research Council's Telecommunications Board (CSTB) May 2002 report entitled *Information Technology Research, Innovation and E-Government*. The full report is available on-line at http://books.nap.edu/html/itr_e_gov/.

Appendix I

Task Force Agency Liaisons

Valerie Gregg, Co-Chair, Census/NSF
Marshall DeBerry, FedStats Program Manager, BJS
Cathryn S. Dippo, BLS and Chair of the FedStats R & D Working Group
Michael Moore, BEA
John Bosley, Rick Devens, BLS
Marianne Zawitz, BJS
Jeff Butler, BTS
EPA
Rachael Taylor, David Raszewski, Census
John Weiner, Colleen Blessing, William Jeffers, EIA
Jim Horsfield, ERS
George Patton, NASS
Bruce Taylor, NCES
Rob Weinzimer, NCHS
John Gawalt, NSF
William Wong, SOI
Laurie Brown, SSA
David Chase, John Sperling, HUD
Bill Tolar, USGS

Appendix II

FedStats/NSF Digital Government Research Projects

A list of the research project titles, the principle investigators (PI), and their academic institutions follows. In addition to these fourteen research grants, the DG Research Program has awarded another 9 FedStats-related grants including one workshop, four planning, and four small grants for exploratory research.

1. A Web-Based Query System for Disclosure-Limited Statistical Analysis of Confidential Data; PI Alan Karr; National Institute of Statistical Sciences

> Working with several Federal statistical agencies, this grant will address an important topic for Federal statistical agencies. As part of their missions, these agencies collect a great deal of microdata (data related to an individual or particular business); this data must remain confidential. Thus, only aggregated microdata is provided publicly. However, the aggregation process reduces much of the value of the microdata for deriving knowledge to be used in research, policy and commercial purposes, so there is a balancing need to provide as much data as possible. What is proposed here is a large-scale system which tracks the history of provision of derived data and which "understands" and can quantify the potential for working backward from the derived data.

2. Data Confidentiality, Data Quality, and Data Integration for Federal Databases: Foundations to Software Prototypes; PI Alan Karr; National Institute of Statistical Sciences

> This award will support research in data confidentiality, data quality, and data integration. Prototypes will be built which can scale to operate on large sets of federally held data. Researchers will partner with several large Federal Government statistical agencies. This topic is of particular importance given the balance these agencies must strive for, in terms of their dual missions to collect and keep private confidential data, while at the same time making that data accessible for research and policy issues. This grant will support a multi-disciplinary multi-institution team, with participants from five universities, one non-profit, and one national laboratory. The disciplines represented include computer science, statistical science, and systems engineering.

3. Adaptive Interfaces for Collection Survey Data From Users; PI Michael Schober; New School University

> The objective of this research is to determine how best to design computer systems for collecting data from (rather than providing data to) users. Government agencies might use such systems to gather the factual data used to calculate the unemployment rate or the Consumer Price Index. Three sets of laboratory experiments focus on actual and simulated desktop (i.e., keyboard and mouse entry) and speech survey interviewing systems. The first

set of studies examines response accuracy and user satisfaction with systems that monitor users' speed of responding and speech patterns in order to diagnose when users misinterpret concepts in the survey questions and could use additional clarification. The second set of studies examines user response accuracy and satisfaction with interfaces that do (or do not) tailor this clarification through dialogue. The third set of studies contrasts interfaces that require users to educate themselves about how the questions should be interpreted with interfaces that engage users in dialogue to figure out the correct answer. The project uses the methods of experimental psychology to provide guidelines for future development of interfaces that collect information from users. This research could significantly improve the accuracy of data collected online by government agencies and others.

4. Citizen Access to Government Statistical Data; PI Gary Marchionini, University of North Carolina

This proposal will conduct research to improve the location/retrieval, reading, navigation and manipulation of tabular statistical data from Federal agencies. These data cover many different domains (e.g., health, labor, transportation), of interest to professionals in the field and to citizens. This work will be accomplished through collaboration with the Bureau of Labor Statistics, the Energy Information Agency, and the National Center for Health Statistics and the Bureau of Census.

5. Collaborative Research: Integration of Data and Interfaces To Enhance Human Understanding of Government Statistics—Toward the National Statistical Knowledge Network; Co-PI Gary Marchionini, University of North Carolina and Catherine Plaisant, University of Maryland

This award will support collaborative research with several Federal statistical agencies to develop better statistical data models, to explore the use of XML, to develop better map-querying tools and to integrate other available tools for manipulating, browsing, and visualizing tabular data. The goal is to develop better human/computer interfaces for expert users to novices, to increase general statistical literacy, and to provide seamless access to data held by multiple Federal agencies and agencies at other levels of government, in particular state and local data.

6. Quality Graphics for Federal Statistical Summaries; PI Alan MacEachren; Penn State University

This award will support collaborative research with several Federal statistical agencies to develop better statistical data models, to explore the use of XML, to develop better map-querying tools and to integrate other available tools for manipulating, browsing, and visualizing tabular data. The goal is to develop better human/computer interfaces for expert users to novices, to increase general statistical literacy, and to provide seamless access to data held by multiple Federal agencies and agencies at other levels of government, in particular state and local data.

7. Quality Graphics for Federal Statistical Summaries; PI Dan Carr; George Mason University—See MacEachren

8. Quality Graphics for Federal Statistical Summaries; PI David Scott; Rice University—See MacEachren

9. Collecting and Using Geospatial Data in the Field: An Extensible Framework and Testbed; PI, Sarah Nusser, Iowa State University

> This work will conceive, develop, and test an extensible framework to support the collection and use of geospatial data in the field. Partner Federal agencies include the Bureau of the Census, the US Geological Survey, and several agencies of the US Department of Agriculture. The proposed activities are designed to meet five key objectives:
>
> 1. Develop a model documenting and formalizing the infrastructure, tools, and key capabilities required to support a flexible and extensible field data collection system. 2. Conduct research on computer science tools and associated information technologies required to fully integrate digital geospatial data into the collection process. 3. Conduct research on infrastructure components that are needed to implement the system in a manner that limits the complexity of the system from the vantage point of the user in the field. 4. Investigate emerging field data collection technologies to determine how the usage of geospatial data is transformed by these new interfaces. 5. Explore the framework model and research developments in an application environment by developing prototype components and testbeds that correspond to agency data collection settings.
>
> Six developments will be needed to address the research objectives: 1. A user-driven framework model, 2. A conceptual framework for conflation of heterogeneous geospatial data for field use, 3. A multi-agent system to support tools required using and collecting geospatial data in the field, 4. Interoperable searching and discovery mechanism for prepared, existing, and potentially unknown sources of data, 5. Object-oriented warehouse designs for the field data collection environment, and 6. Evaluations of emerging field technologies and their impact on user activities.

10. Digital Government Research Center: Energy Data Collection; PI Yigal Arens; Information Sciences Institute, University of Southern California

> This proposal will create an Energy Data Collection to support real-time integrated viewing, interaction, and manipulation of the Department of Energy's gasoline-related data collection, through a partnership with the Energy Information Agency. The proposed research will cover automated ontology development and distributed information integration across data held by multiple Federal agencies.

11. I2T: An Information Integration Testbed for Digital Government; PI, Chaitan Baru; San Diego Super Computing Center

This project will address one of the major problems in government information systems, the inability to integrated information from various heterogeneous data sources. Usually these data are collected and managed by different agencies at different levels of government, providing more impediments to integration. Partners from the Bureau of the Census, National Archives and Records Administration, US Geological Survey, the State of Pennsylvania, and the San Diego Association of Governments will work with researchers from the San Diego Supercomputer Center, the University of California at San Diego, the University of Michigan, and the University of Pennsylvania. Building upon the initial work of the Mediation of Information using XML (MIX) project, this grant has four major technical thrusts: 1. Allow for an extension of MIX's wrapper technology to the domain of geospatial information, 2. Develop data transfer protocols for lightweight network-based agents, 3. Investigate new interfaces to the data, and 4. Build wrapper toolkits for geospatial and statistical survey metadata.

12. Survey Authoring and Administration Testbed; PI, Robert Balzer; Information Sciences Institute, University of Southern California

This grant will address an important problem area for the US Bureau of the Census, and through them, the various Federal agencies who commission the Bureau to conduct statistical survey, i.e., the specification and creation of complex survey instruments. At present Census is using a very old proprietary system, which occupies nearly 100 Census staff. The PI will use commercially-available software as an infrastructure upon which will be created a research prototype of a modern Web/relational database system, using modern software engineering techniques, to allow graphically-specific surveys with built-in error-checking and administration.

13. Digital Government Research Center (DGRC): Bringing Complex Data to Users; Co-PI Judith Klavans, Columbia University and Ed Hovy, Information Sciences Institute, University of Southern California

In partnership with the Federal Energy Information Agency on the topic of trade data, Columbia University and the Information Sciences Institute of the University of Southern California will work in three areas of relevance to the Agency mission: 1. Main memory query processing, which provides extremely fast querying of multiple statistical data sets, an area of concern to all statistical agencies which must provide aggregated data which maintains the confidentiality of the citizens and businesses which contributed the data; 2. Multilingual question and answering, which will explore the possibility of providing automated translation and querying from English to Spanish and Chinese, and perhaps one other language. As the US population becomes increasingly multi-lingual, natural language processing as a service of gov't web sites will become more and more expected. 3. Usability testing of components developed in this and in another grant to this team under the Digital Government program.

14. Digital Government: Improving Statistical Literacy Through FedStats; PI Bill Smith; American Statistical Association

This grant will support a planning process to develop concepts for research in user interfaces and forms of on-line learning and analysis to improve statistical literacy for the citizen. The proposer will work with an existing group of collaborating Federal statistical agencies, know as "FedStats". FedStats has an award-winning web site, and collaborates with several other Digital Government award recipients.

Session 5

Improving Data Quality

Ensuring Information Quality: Challenges And Opportunities

Katherine K. Wallman
Office of Management and Budget

It's a pleasure to join with my colleagues to discuss ongoing efforts of OMB and the Federal agencies to improve the quality of information that agencies disseminate to the public. As most if not all of you know, a recent law added further impetus to, and substantially broadened, the scope of our long-term efforts in this arena.

Background

The particular efforts we are discussing today began late in 2000, when Congresswoman Jo Ann Emerson sponsored an amendment to OMB's appropriations bill. This provision required OMB to develop government-wide standards "for ensuring and maximizing" the quality of information disseminated by Federal agencies. It was enacted as Section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001. (This law should not be confused with an earlier "information access" law – one that I know is very familiar to you – that was sponsored by Senator Richard Shelby of Alabama and amended the Freedom of Information Act to provide greater public access to research data generated under Federal research grants.)

At OMB, we call this more recent legislation the "Information Quality Act." There were no hearings or extensive legislative history, and no fanfare when it passed. Yet, in our view, this law provides a very important opportunity to raise the quality of government information. Together, the information access and information quality laws will be mutually reinforcing in promoting responsible public access to technical information produced and used by Federal agencies.

The Information Quality Law establishes a performance-oriented information quality system across the government. It could help build quality into the system from the beginning and lead to evolutionary progress. We expect there will be a network effect – with cross-fertilization between agencies, among agency programs, and between government and citizens.

Interagency dialogue is flourishing as agencies developed and are now beginning to implement their particular guidelines, in concert with OMB's government-wide guidelines. No better example could be cited than the collaborative efforts of the statistical agencies in addressing this new challenge.

With that background, let me spend the remainder of my time briefly walking you through three phases of our recent information quality efforts: (1) OMB's general guidelines; (2) the agency guidelines; and, on the horizon (3) implementation of the guidelines.

Phase One: OMB's Government-Wide Guidelines

OMB issued its government-wide guidelines in interim final form on September 28, 2001, and in final form on February 22, 2002 (67 FR 8452). To implement the statute, OMB imposed three core responsibilities on the Federal agencies.

First, agencies must embrace a basic standard of quality as a performance goal, as embodied in the OMB government-wide guidelines, and develop pre-dissemination review procedures. In information collections proposed by agencies under the Paperwork Reduction Act, the agency and OMB can consider whether the quality of subsequent disseminations would meet the applicable performance standards.

Second, agencies must report annually to OMB on "the number and nature of complaints" and "how such complaints were handled by the agency.

Finally, agencies must establish a petition process allowing affected parties to request that the agency correct information that does not comply with the OMB or agency guidelines. OMB made clear that the burden of proof is squarely on the affected parties; they must demonstrate that a specific dissemination does not meet the applicable quality standards. The opportunity for complaints and appeals went into effect on October 1.

The scope of the Information Quality Act is very broad. It spans information related to regulatory, statistical, research, and benefits programs. It covers all Federal agencies subject to the Paperwork Reduction Act, including the independent regulatory commissions. OMB's guidelines define "information" as "any communication or representation of knowledge such as facts or data" in any medium. (Indeed, this is why OMB calls this law the Information Quality Act, and not the Data Quality Act. It covers more than just quantitative data.)

OMB's guidelines explain that "quality" encompasses "utility" (usefulness to its intended users), "integrity" (security), and "objectivity." "Objectivity" focuses on whether the disseminated information is accurate, reliable and unbiased as a matter of presentation and substance.

At the same time, OMB provided a variety of exemptions from the guidelines to protect privacy and commercial secrets, and to facilitate press releases, third party submissions in public filings, archival records, personal articles by agency employees, testimony, and subpoenas and adjudicative determinations. OMB also provided agencies discretion to reject complaints that are groundless or made in bad faith, or boil down to a difference of opinion.

OMB recognized that information quality can be costly and encouraged agencies to consider the social value of better information in different contexts. Ordinary information is distinguished from "influential" information -- that is, scientific, financial and statistical information having a clear and substantial impact on important public policies or important private sector decisions. "Influential" information is subject to higher standards of quality. With several important exceptions and qualifications, influential information should be reproducible by qualified third parties.

Phase Two: Agency-Specific Guidelines and OMB Review

In moving to the development of agency-specific guidelines, it is important to note that the statistical agencies were decidedly "out in front" on this challenge. In as very real sense, they were perhaps most ready to meet the challenge, for information quality standards historically have been central to their work. What was especially remarkable, however, was the fact that the statistical agencies, under the umbrella of the Interagency Council on Statistical Policy, voluntarily came together at the earliest stages of this process to develop a common template for their agency guidelines, and subsequently published a common Federal Register notice to draw the public's attention to their individual statistical agency guidelines. (My co-panelist Nancy Kirkendall will be discussing that initiative in more detail.)

At a broader level, to facilitate development of the agency guidelines, OMB – with support from the agencies -- arranged for three workshops that were conducted by the National Academies last Spring. These workshops were widely attended by hundreds of agency staff and interested members of the public. They facilitated the early exchange of ideas and fostered the development of the agency guidelines.

OMB's review of the agencies' guidelines began when proposed drafts were released for public comment in May. Based on a preliminary review, OIRA Administrator John Graham sent a June 10 memorandum to the President=s Management Council suggesting for the agencies' consideration particularly noteworthy provisions gleaned from various drafts. He also provided guidance for greater uniformity in some provisions.

Similarly, on September 5, while OIRA was completing its review of agencies' draft final guidelines, Administrator Graham sent a short follow-up memorandum to the President=s Management Council encouraging greater uniformity on a few process issues.

By October 1, OMB had completed its review of the information quality guidelines for more than 65 Federal departments and agencies (including over 45 guidelines developed for specific components of Federal departments.

Phase Three: Implementation

Having developed information quality guidelines, the agencies now must turn to the equally challenging task of implementing them. Agencies must ensure that the new procedures and criteria are integrated into their day-to-day activities. On October 4, Administrator Graham sent a third memorandum to the President's Management Council outlining OMB's current plans for providing continuing guidance to agencies on applying OMB's information quality guidelines, as well as for monitoring the agencies' implementation.

In the October 4 memo, OMB established two basic oversight measures:

- First, we offered some preliminary suggestions to the agencies on information to include in their annual reports – most notably descriptions of the kinds of complaints they receive

and their resolution – so we and the public can understand the effectiveness of the administrative correction process.

- Second, to help OMB gauge the public interest in information quality issues and agencies= responses, we requested that each agency provide us with copies of complaints and related information involving several key issues:

    1. major policy questions of strong interest to two or more Federal agencies;

    2. “influential” disseminations alleged to be in violation of OMB's government-wide guidelines;

    3. novel procedural, technical or policy issues; or

    4. disseminations occurring in a public comment process where the complainant shows a reasonable likelihood of suffering actual harm if the agency does not promptly consider the complaint and doing so would not unduly delay the agency’s proceeding.

(Agencies that post their complaints and responses on their websites will not need to forward these materials to OMB.)

Conclusion

In sum, we have an ambitious legislative mandate, and many of you are helping us implement this responsibility effectively.

The Census Bureau Quality Program and
Section 515 Information Quality Guidelines

Cynthia Z.F. Clark and Jay Keller

Census Bureau Quality Program

Prior to OMB's Information Quality Guidelines directive, and our participation in the joint statistical agency activities described by Nancy Kirkendall, the Census Bureau had established a Quality Program designed to relate the different quality efforts underway throughout the Census Bureau.

The program, which is under the stewardship of the Census Bureau's Methodology and Standards Directorate, partners with program areas and is designed to build excellence through innovative techniques, technologies, evaluations and improvements in our business processes.

- Specific objectives of the program are to ensure that Census Bureau products meet quality standards and that we provide sufficient information on quality so that users can determine the appropriateness of these data for the intended purposes.

- The strategies for achieving the goals are to:
  - design processes,
  - establish quality principles, standards, guidelines, and best practices,
  - develop tools and checklists,
  - and design web sites to facilitate communication.

The Quality Management Repository (QMR) was established as a portal intranet site in the summer of 2001:

- to share,
- manage,
- and disseminate information addressing principles, practices and related quality issues to Census Bureau employees.

QMR users can find and view information by "product" and "process." The process documents are organized around the standard workflow of surveys and censuses, with the Census Bureau using the following categories:

- Content
- Planning
- Design
- Data Collection
- Data Processing
- Data Quality, Analysis, and Evaluation
- Dissemination
- Data Products and Services

The QMR view of documents organized by product includes menu selections for principles, standards, guidelines, current practices, and training. These documents provide direct support to project managers in developing, tracking, and updating their quality management plans.

Census Bureau Guidelines

At the time of the OMB directive and the initial work of the joint statistical agency group, the Census Bureau had begun work populating our Quality Management Repository with principles, standards, guidelines, and best practice documents.

- Criteria was established for each type of document as well as a template.
- Documents were to be issued by the Census Bureau Methodology and Standards Council after receiving review from the program divisions and the associate directors.
- These documents were developed as issues arose by convening cross-directorate teams.

- Additionally, an effort was made to inventory and review documents previously issued that provided direction or guidance.
    - In most cases the previous direction provided was for individual directorate programs with the exception of the well known Technical Paper 32 (and a follow-up memorandum) that provided direction for the Discussion and Presentation of Errors.
    - By contrast the current approach was to develop documents that were corporate or bureau wide.

In developing the Census Bureau Section 515 Information Quality Guidelines, we took an organizational approach—as, it turned out, did many statistical and other federal agencies—inspired by the Social Security Administration model. Our guidelines discuss:

- the role of the Census Bureau,
- efforts to ensure utility (and relevance) in our products,
- objectivity guidelines (including the use of reliable data sources, sound analytic techniques, required reviews before the release of data, and informing users of data quality and methodology),
- guidelines on transparency and reproducibility,
- data integrity,
- the Census Bureau's performance principles in the eight categories of statistical activities identified by the statistical agencies,
- and administrative correction mechanisms.

The Census Bureau quality processes are very similar to but not exactly the same as the joint statistical agency activities (Chart 1). In the process of preparing the Census Bureau Section 515 Information Quality Guidelines, the Census Bureau desired to ensure consistency between the activities identified by the statistical agencies and the previously established (but not yet populated) Quality Framework. To do this, we chartered eight working groups of internal experts from throughout the organization to develop principles for each of the joint agency activities, drawing upon previous documents and known practices at the Census Bureau. These principles were envisioned as broad underlying policies, approaches and direction that govern the design of the

activity in question with emphasis on those that relate to quality. They appear both in our Section 515 Information Quality Guidelines and in the Census Bureau Quality Management Repository. These written principles now provide an encompassing framework for future development of relevant standards, guidelines, and best practices.

Efforts at the Department of Commerce

At the same time the Census Bureau was participating in the joint statistical agency activities to develop the Federal Register Notice and the categories of statistical agency activities, we were also part of a Department of Commerce effort to develop umbrella DOC guidelines.

The DOC effort was headed by the Chief Information Officer and the Office of General Counsel at Commerce. Teams were formed, made up of representatives of Commerce operating units, to develop the overall DOC guidelines and instructions for operating units to follow in developing their individual guidelines. For some aspects of OMB's information quality guidelines requirements, such as in the areas of computer system integrity, financial information, and organizational and administrative information, the DOC guidelines ultimately served as a model for its operating units to use. However, operating units were responsible for developing their own guidelines, particularly in the area of "objectivity," and especially components of the objectivity requirements including transparency (of methods) and reproducibility (of results).

Corrections Mechanism

Commerce also developed a prototype corrections mechanism process. An early issue for us was the Department's initial objective for the centralizing of requests for correction—perhaps at the Department of Commerce, or at minimum at each operating unit. Our internal objective was to maintain the decentralization of processes already in place for corrections of our current programs:

- Count Question Resolution,
- Local Update of Census Addresses,
- Governmental Unit Boundaries and Street and Address Range Information,
- Small Area Income and Poverty Estimates,
- Population Estimates,
- Foreign Trade Statistics.

Because these programs had their own complaint procedures, which in some cases were longstanding and highly publicized, we secured approval from the Department of Commerce to keep these programs in place, and to advertise methods for the public to request correction of these programs through their individual mechanisms on our Information Quality Guidelines website. We also established a corrections mechanism for "All Other" complaints—any requests for correction that do not fall into the preexisting programs.

To fulfill Department of Commerce requirements that the tracking of corrections requests be automated and that tracking occurs during and after the resolution process, our Computer Assisted Survey Research Office designed automated procedures using Microsoft Access, and worked with our various program areas to ensure that they either had their own automated tracking system or

could incorporate the use of our newly designed system by October 1. Our current plans are to develop monthly summaries of corrections requests across our seven preexisting mechanisms and our "all other" procedure, and use these to provide quarterly (or more frequent) reports to the Department of Commerce and the annual report to OMB.

Continuing the Quality Program at the Census Bureau

Besides developing our agency's Section 515 Information Quality Guidelines, and the performance principles associated with the eight statistical agency activites, we have continued to develop quality principles, standards, guidelines, and best practices to populate the Quality Management Repository. Standards in this framework are survey or statistical methodology procedures required for all Census Bureau program areas. We developed two standards that are particularly relevant to our Section 515 Information Quality Guidelines:

- Standard for Correcting Information that does not Comply with Census Bureau Section 515 Information Quality Guidelines (Dissemination; Data Products and Services – issued 05/16/02)
- Standard for Review of Census Bureau Documents and Presentations (Data Products and Services – issued 08/09/02)

Other standards in the Quality Framework include:

- Standard: Source and Accuracy Statements for Census and Survey Data Tabulations and Model-Based Estimates (Dissemination – reissued 09/24/02)
- Standard: Minimal Information to Accompany any report of Census Bureau Data (Dissemination – soon to be issued)
- Standard: Definitions for Survey and Census Metadata (Planning – soon to be issued)

Guidelines in the framework highlight survey or statistical methodology procedures recommended for all Census Bureau program areas. They are being developed using a checklist approach that would guide the employee in ensuring that all relevant aspects are considered in planning and executing a statistical program activity. Our guidelines currently include:

- Quality Checklist for Census Bureau Products (Planning – issued 05/07/01)
- Coding Verification (Data Processing – issued 06/13/02)
- Sample Selection Verification (Design – issued 10/29/01)

We have several efforts currently underway. They include the development of:

- Standards for Pretesting Questionnaires for Census Bureau Demographic, Decennial, and Economic Census, Surveys, and Tests.
- Standards for Discussion and Presentation of Errors in Data (a revision of Technical Paper 32),
- Guidelines for Quality Assurance for CAPI Interviewing,
- Guidelines for Quality Assurance for Commercial Printing,
- Guidelines for Quality Assurance for Record Linkage,
- Guidelines for Quality Assurance Procedures for Research and Evaluation Reports.

The Quality Program will convene working groups to develop standards and guidelines as issues arise. Additionally, previous guidance is being reviewed to determine whether these documents need to be revised and reissued in the Quality Framework.

Chart 1
Census Bureau Quality Processes and Statistical Agency Activities

Quality Framework Processes	Statistical Agency Activities
Content	Development of Concepts and Methods
Planning	Planning and Design
Design	
Data Collection	Collection of Data
Data Processing	Processing and Editing of Data
Data Quality, Analysis and Evaluation	Analysis of Data
	Production of Estimates or Projections
Dissemination	Establishment of Review Procedures
	Dissemination of Data
Data Products and Services	

Information Quality Guidelines At NCES

Marilyn McMillen Seastrom
National Center for Education Statistics

Purpose of Statistical Standards

The National Center for Education Statistics (NCES), the principal statistical agency within the U.S. Department of Education, released the 2002 revised version of the *NCES Statistical Standards* on October 1, 2002: http://nces.ed.gov/statprog/stat_standards.asp

Our primary goal is to provide high quality, reliable, useful, and informative statistical information to public policy decision makers and to the general public. Thus, much of the standards and guidelines are geared towards fulfilling that goal. In particular, the standards and guidelines are intended for use by NCES staff and contractors to guide them in their data collection, analysis, and dissemination activities. These standards and guidelines are also intended to present a clear statement for data users regarding how data should be collected in NCES surveys, and the limits of acceptable applications and use. Beyond these immediate uses, we hope that other organizations involved in similar public endeavors will find the contents of some of these standards and guidelines useful in their work as well. All users of these standards and guidelines should be cognizant of the fact that the contents of the NCES standards are continually being reviewed for technological and statistical advances.

Background of Statistical Standards

Data quality is the cornerstone of all official statistics programs. To this end, there are a number of international and national groups that have devoted considerable time and effort to delineating important concepts and principles for official statistics. On the international front, the United Nations (UN) and the Economic Commission For Europe (ECE) have both adopted a set of "Fundamental Principles of Official Statistics." Included among the 10 principles are calls for statistical agencies to use professional standards that are based on scientific principles to guide the methods and procedures for the collection, processing, storage, and presentation of statistical data. The principles also call for the inclusion of relevant information on the sources, methods, and procedures of the statistics. In a similar vein, one of the main objectives identified by the Statistics Directorate of the Organization for Economic Co-operation and Development (OECD) includes the development of international statistical standards, systems, and collaborations. Similarly, the International Monetary Fund's (IMF) data dissemination standard includes the integrity and quality of data, coverage, periodicity and timeliness, public access to data, and full documentation of the data collection.

In the United States, there are two national committees that have each been working for a quarter of a century to improve statistical methods and data quality—the Federal Committee on Statistical Methodology (FCSM) and the Committee on National Statistics (CNSTAT). The Office of Management and the Budget (OMB) convenes the Federal Committee to provide a forum for communicating and disseminating information about statistical practices among all

Federal statistical agencies. The FCSM also recommends the introduction of new methodologies in Federal statistical programs to improve data quality.

The National Research Council of the National Academy of Sciences convenes CNSTAT, a committee of prominent researchers from universities and private research organizations, to study statistical topics to improve the effectiveness of the Federal statistical system. CNSTAT monitors the statistical policy and coordinating activities of the Federal government, reviews the statistical programs of federal agencies and suggests improvements, reviews data-handling and privacy and confidentiality policies and provides recommendations for best practices, studies data gaps and recommends additions as necessary, and reviews extant methodologies and suggests improved statistical methods.

CNSTAT published a monograph on the "Principles and Practices for a Federal Agency" to assist Federal statistical agencies. The main principles include relevance of data, credibility among data users, confidentiality of data, and trust among data providers. Many of the practices identified parallel the "Fundamental Principles of Official Statistics" promulgated by the UN and the ECE. For example, statistical agencies should have a commitment to high quality and professional standards. In discussing openness about the data, CNSTAT stresses the importance of providing a full description of the data, the methods used, and assumptions made. The description should include reliable indicators of the kinds and amount of error in the data. CNSTAT also stressed the importance of wide dissemination of data presented in a user-friendly format. The CNSTAT guide was one of the tools used by NCES staff in planning their current revision of the agency's statistical standards.

Development of Statistical Standards at NCES

NCES first adopted written statistical standards in the spring of 1987. These standards were the result of a multi-year evaluation and planning process that included a recommendation for the development of statistical standards from the Committee on National Statistics at the National Academy of Science. With that recommendation, a statistical standards program was initiated at NCES in 1985. Using the Energy Information Administration's Standards Manual and the Census Bureau's technical paper on "Standards for Discussion and Presentation of Errors in Survey and Census Data," NCES staff, in consultation with outside experts developed the 1987 version of NCES statistical standards.

With the adoption of this first set of standards, the Agency Director called for a formal evaluation to start the following fall, to insure that the standards were fully implemented and to identify any difficulties with the standards. In 1989, the Center undertook a full-scale revision of the 1987 standards. The revisions were developed by NCES staff, and reflected their first-hand experiences in using the 1987 standards. After multiple reviews of interim drafts by NCES staff and the NCES Advisory Council of Education Statistics, NCES Senior Staff accepted the revised standards in the spring of 1992.

At the June 1992 release of the *NCES Statistical Standards* report, the Acting Commissioner summarized the standards in the following statement:

> They: (1) codify how we expect to behave professionally, (2) indicate the basis on which we expect to be judged by our peers in the statistical community, (3) represent the quality we expect in any of our efforts or those of our contractors and grantees, (4) provide a means to assure consistency among the studies the Center conducts, and (5) document for users, the methods and principles the Center employs in the collection of data.

The Acting Commissioner also reiterated the Center's commitment to periodic evaluations of the implementation of the standards and to a periodic review of the standards' operational feasibility.

The current revision process began in the summer of 1999 with a review of existing standards from a number of national and international statistical policy agencies and committees and from other international and national statistical agencies. At the same time the 1992 *NCES Statistical Standards* were made available on the Web, and NCES staff were given a 30-day period to submit comments concerning potential revisions and additions to the NCES standards. Following these activities an agency-wide Steering Committee was formed to work on the standards revision process. The Steering Committee formed 15 Working Groups that comprised more than one-half of the NCES staff to work on the set of topics identified in the 1999 reviews.

Each Working Group drafted their assigned standards; each of which underwent a multi-step review process. Following a 30-day NCES staff comment period, the working group members made revisions, the Steering Committee reviewed the drafts and submitted them to Senior Staff. The drafts were then reviewed by Senior Staff, modified as necessary, and then shared with a group of 40 to 50 representatives of the contractors who work with NCES on data collection, analysis, and dissemination. Additional revisions were incorporated following the input from this broad group. NCES also commissioned the National Institute of Statistical Sciences to convene an independent review panel of statistical experts to review and comment on the draft standards prior to final acceptance by the Steering Committee and Senior Management. The standards on this Web site are the result of the efforts of the many persons who participated in this multi-stage review process, but ultimately NCES takes responsibilities for any lack of clarity or completeness.

During the recent NCES standards revision, the Office of Management and Budget (OMB) issued government-wide guidelines for ensuring and maximizing the quality of information disseminated by Federal agencies. The OMB guidelines direct all agencies covered by the Paperwork Reduction Act (44 U.S.C. chapter 35) to develop and implement procedures for reviewing and substantiating the quality of information disseminated by the agency. In order to meet these goals, each agency is required to develop and promulgate quality guidelines.

In response to the OMB guidelines, the federal statistical agencies collaborated to identify a set of activities that are essential to maintaining the quality and credibility of statistical data. The NCES draft revised standards are organized around the shared framework for federal statistical agencies. NCES remains committed to the principles outlined by the 1992 NCES Acting Commissioner; what is more, these principles are reaffirmed in the OMB call for data quality guidelines.

OMB Quality Guidelines

Background

Section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001 (Public Law 106-554), directed the U.S. Office of Management and Budget (OMB) to issue government-wide guidelines that "provide policy and procedural guidance to Federal agencies for ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by Federal agencies." Information, as defined by OMB, includes any communication or representation of knowledge, such as facts or data, in any medium or form, including textual, numerical, graphic, cartographic, narrative or audiovisual forms. Dissemination refers to any agency initiated or sponsored distribution of information to the public (OMB, *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies*, February 22, 2002, 67 FR 8452-8460).

> NCES provides the public with a wide variety of information about the condition of american education. Information quality is important to nces because educators, researchers, policymakers, and the public use NCES products for a variety of purposes. Thus it is important that information products that NCES disseminates are accurate and reliable. Most of the information products are available both as printed and electronic documents. They are announced on the NCES website (nces.ed.gov), and most electronic versions can be accessed and downloaded directly from the website.

Purpose and Scope

NCES guidelines have been identified as Standards for the last 15 years, thus we will retain that label. The purpose of these Standards is to describe NCES policy and procedures for reviewing and substantiating the quality of information before it is disseminated. These Standards are consistent with those issued by OMB and the Department of Education. These Standards represent a performance goal for NCES and are intended to improve the quality of the information NCES shares with the public.

In addition to the NCES Standards, the Department of Education and OMB have more general Information Quality Guidelines that apply to NCES. What is more, NCES will follow the request for corrections and appeal process described in the Department Information Quality Guidelines. www.ed.gov/offices/OCIO/info_quality/info_guide.html

The Standards are applicable to any information that NCES disseminates after October 1, 2002. In addition, some previously released information products continue to be used for decision-making or are relied upon by the Department of Education and the public as official, authoritative, government data; these data are, in effect, constantly being re-disseminated and thus are subject to these Standards and to the Department and OMB Information Quality Guidelines. Previously released information products that do not meet these criteria are considered archived information and thus are not subject to the Guidelines.

In addition to archived reports, these Standards do not cover all other information held or disseminated by NCES. The Department of Education Information Quality Guidelines include a list of excluded items, although that list also applies to NCES, the items that are particularly relevant to NCES are included here. For example, the guidelines generally do not cover: internal information such as employee records; internal procedural, operational, or policy manuals prepared for the management and operations of the Department of Education (and NCES) that are not primarily intended for public dissemination; information collected or developed by NCES that is not disseminated to the public, including documents intended only for inter-agency or intra-agency communications; opinions that are clearly identified as such, and that do not represent facts or NCES views; correspondence with individuals; comments received from the public in response to *Federal Register* notices, electronic links to information on other Web sites; and research findings published by NCES data cooperatives or grantees, unless NCES represents or uses the information as the official position of the Department, or in support of the official position of the Department, or has authority to review and approve the information before release.

> For information covered by information quality guidelines, the NCES standards provide a basic standard of quality that can be defined based on the three elements of quality as defined by OMB: utility, objectivity, and integrity. These elements are intended to ensure that information disseminated by the nces is useful, accurate, reliable, unbiased, and secure.

Framework

Utility *refers to the usefulness of the information to its intended users. The usefulness of information disseminated by NCES should be considered from the perspective of NCES, educators, education researchers, policymakers, and the public. Utility is achieved by staying informed of information needs and developing new products and services where appropriate.*

NCES wants to ensure that information it disseminates meets the needs of the intended users. NCES relies upon internal reviews and analyses, along with feedback from advisory committees, educators, education researchers, policymakers, and the public to ensure that information disseminated by NCES meets the needs of intended users. In addition, all information products should be grammatically correct and clearly written in plain English. The target audience should be clearly identified, and the product should be understandable to that audience.

Consistent with OMB guidance, the goal is to maximize the usefulness of information and minimize the cost to the government and the public. When disseminating its information products, NCES will utilize all feasible and available dissemination channels so that the public, education researchers, and policymakers can locate NCES information in an equitable and timely fashion.

The information disseminated by NCES includes administrative and statistical data. NCES collects and disseminates administrative data from universe collections of elementary and secondary and postsecondary institutions. These universe collections are based on reports aggregated from records from schools, school districts, and states. NCES also collects and disseminates data from a number of sample survey data collections that are designed to fill the

information needs for statistical data. NCES supports both ongoing sample survey data collections and special purpose surveys that are designed to fill data gaps or information needs that are identified through internal review, legislative mandates, or input from data users outside the Department. All statistical reports and related products are reviewed to ensure their usefulness to the intended users. Where appropriate, contact information is available on each publication to facilitate feedback and questions by users.

The specific NCES standards that contribute directly to the utility and the dissemination of information include those on the Initial Planning of Surveys (1-1), Publication and Product Planning (1-2), and the Release and Dissemination of Reports and Data Products (7-3).

Objectivity *refers to whether information is accurate, reliable, unbiased, and is presented in an accurate, clear, and unbiased manner. It involves both the content of the information and the presentation of the information. This includes complete, accurate, and easily understood documentation of the source of the information, with a description of the sources of any errors that may affect the quality of the data, when appropriate. Objectivity is achieved by using reliable information sources and appropriate techniques to prepare information products.*

NCES strives to present information to the public in an accurate, clear, complete, and unbiased manner. Prior to dissemination to the public, all products are reviewed for objectivity using sound statistical methods and the principles of transparency and reproducibility, as delineated in the OMB Information Quality Guidelines. In addition, all products undergo editorial and technical peer review to assist NCES in meeting this goal.

NCES is committed to the principles for objectivity in administrative and statistical data that are outlined in the Department of Education's Guidelines. To that end, we have specific standards that relate to each of the Department's principles:

1. In formulating a data collection plan goals of the study should be clearly described—Initial Planning of Surveys (1-1), Design of Surveys (2-1), Developing a Request for Proposal (RFP) for Surveys (2-3).
2. The subjects to be studied and the data to be collected should be clearly defined, using broadly understood concepts and definitions—Initial Planning of Surveys (1-1), Codes and Abbreviations (1-4), Defining Race and Ethnicity Data (1-5), Design of Surveys (2-1), Developing a Request for Proposal (RFP) for Surveys (2-3), Maintaining Data Series (2-5).
3. The data collection techniques should be well thought out, clearly articulated, and designed to use state of the art methodologies in the data collection—Initial Planning of Surveys (1-1), Design of Surveys (2-1), Survey Response Rate Parameters (2-2), Developing a Request for Proposal (RFP) for Surveys (2-3), Pretesting Survey Systems (2-4), Educational Testing (2-6), Coverage for Frames and Samples (3-1), Achieving Acceptable Response Rates (3-2), Monitoring and Documenting Survey Contracts (3-3).
4. In designing the work, every effort should be made to minimize the amount of time required for survey participants—Achieving Acceptable Response Rates (3-2).

5. The source of data should be reliable. In the case of sample survey data, the sample should be drawn from a complete list of items to be tested or evaluated, the appropriate respondents must be identified, correctly sampled, and queried with survey instruments that have been properly developed and tested—Initial Planning of Surveys (1-1), Design of Surveys (2-1), Pretesting Survey Systems (2-4), Coverage for Frames and Samples (3-1).
6. Response rates should be monitored during data collection. When necessary, appropriate steps should be taken to ensure the respondents are a representative sample—Computation of Response Rates (1-3), Survey Response Rate Parameters (2-2), Achieving Acceptable Response Rates (3-2), Monitoring and Documenting Survey Contracts (3-3), Nonresponse Bias Analysis (4-4).
7. Care should be taken to ensure the confidentiality of personally identifiable data, as required by law, during data collection, processing, and analysis of the resulting data—Maintaining Confidentiality (4-2).
8. Upon completion of the work, the data should be processed in a manner sufficient to ensure that the data are cleaned and edited to help ensure that the data are accurate and reliable— Initial Planning of Surveys (1-1), Design of Surveys (2-1), Monitoring and Documenting Survey Contracts (3-3), Data Editing and Imputation of Item Nonresponse (4-1), Evaluation of Surveys (4-3).
9. The data collection should be properly documented and stored, and the documentation should include an evaluation of the quality of the data with a description of any limitations of the data—Monitoring and Documenting Survey Contracts (3-3), Documenting a Survey System (3-4), Machine Readable Products (7-1).
10. Data should be capable of being reproduced or replicated based on information included in the documentation including, for example:
    a) The source(s) of the information;
    b) The date the information was current;
    c) Any known limitations on the information;
    d) The reason why the information is provided;
    e) Descriptions of any statistical techniques or mathematical operations applied to source data; and
    f) Identification of other sources of potentially corroborating or conflicting information.

The relevant standards include—Monitoring and Documenting Survey Contracts (3-3), Documenting a Survey System (3-4), Machine Readable Products (7-1), Survey Documentation in Reports (7-2).

11. If secondary analysis of data is employed, the source should be acknowledged, the reliability of the data should be confirmed and documented, and any shortcomings or

explicit errors should be acknowledged (e.g., the representativeness of the data, measurement error, data preparation error, processing error, sampling errors, and nonresponse errors)—Survey Documentation in Reports (7-2).

12. The analysis should be selected and implemented to ensure that the data are correctly analyzed using modern statistical techniques suitable for hypothesis testing. Techniques may vary from simple tabulations and descriptive analysis to multivariate analysis of complex interrelationships. Care should be taken to ensure that the techniques are appropriate for the data and the questions under inquiry—Statistical Analysis, Inference, and Comparisons (5-1), Variance Estimation (5-2), Rounding (5-3), Tabular and Graphic Presentations of Data (5-4).
13. Reports should also include the reason the information is provided, its potential uses, and cautions as to inappropriate extractions or conclusions, and the identification of other sources of corroborating or conflicting information—Survey Documentation in Reports (7-2).
14. Descriptions of the data and all analytical work should be reported in sufficient detail to ensure that the findings could be reproduced using the same data and methods of analysis; this includes the preservation of the data set used to produce the work—Monitoring and Documenting Survey Contracts (3-3), Documenting a Survey System (3-4), Evaluation of Surveys (4-3), Machine Readable Products (7-1), Survey Documentation in Reports (7-2).
15. All reports, data, and documentation should undergo editorial and technical review to ensure accuracy and clarity prior to dissemination. Qualified technical staff and peers outside the Department should do the technical review—Review of Reports and Data Products (6-1).
16. To ensure the utility of the work, all work must be conducted and released in a timely manner—Publication and Product Planning (1-2), Release and Dissemination of Reports and Data Products (7-3).
17. There should be established procedures to correct any identified errors. These procedures may include the publication of errata sheets, revised publications, or Web postings—Review of Reports and Data Products (6-1), Release and Dissemination of Reports and Data Products (7-3).

Integrity *refers to the security or protection of information from unauthorized access or revision. Integrity ensures that the information is not compromised through corruption or falsification.*

NCES has in place appropriate security provisions for the protection of confidential information that is contained in all identified systems of records. In accordance with statutory and administrative provisions governing the protection of information, NCES protects administrative records and sample survey data that include personally identifiable information, especially survey data that are collected under a pledge of confidentiality. Applicable provisions governing the protection of information include the following:

- Privacy Act;
- Computer Security Act of 1987;

- Freedom of Information Act;
- OMB Circulars A-123, A-127, and A-130;
- Federal Policy for the Protection of Human Subjects;
- Government Information Security Reform Act; and
- National Education Statistics Act, as amended by the USA Patriot Act of 2001.

The relevant standard is Maintaining Confidentiality (4-2).

Influential Information

The OMB guidelines for implementing section 515 recognize that some government information needs to meet higher quality standards than a basic standard of quality. The level of effort required to ensure the quality of information is tied to the uses of the information. Information that is defined as "influential" requires a higher level of effort to ensure its' quality and reproducibility. Scientific, financial, and statistical information is considered influential if the Department can reasonably determine that the information is likely to have a clear and substantial impact on important public policies or private sector decisions if disseminated.

Influential information must be accompanied by supporting documentation that allows an external user to clearly understand the steps involved in producing the information and, to be able to reproduce the information. Any influential original data files must describe the design, collection, and processing of the data in sufficient detail that an interested third party could understand the specifics of the original data and, if necessary, independently replicate the data collection. In the case of influential analytic results, the mathematical and statistical processes used to produce the report must be described in sufficient detail to allow an independent analyst to substantially reproduce the findings using the original data and identical methods.

When full public access to NCES data and methods is not possible due to other compelling interests, NCES will apply especially rigorous robustness checks to analytic results and will document the checks that were undertaken. In those cases where protecting the confidentiality of individually identifiable data precludes the full release of a data file, persons seeking access to such data and methods are required to follow applicable NCES requirements and procedures for seeking such access. In all cases, the interest in transparency of the agency's data shall not override other compelling interests such as privacy, intellectual property, and other confidentiality protections (16 CFR 4.9-4.11 and OMB Guidelines, par V.b.3.ii.B.j.).

Inasmuch as it is not always possible to predict in advance all of the uses of the information included in NCES data collections, all information collected and disseminated by NCES is held to the standards of quality, reproducibility, and documentation that are required for influential information.

Information Correction Requests and Appeals

Effective October 1, 2002 the Department of Education and NCES will allow any affected person to request the correction of information the Department disseminates that does not comply with applicable OMB, Department of Education, and NCES information quality guidelines. An affected person is an individual or an entity that may use, benefit or be harmed by the disseminated information at issue.

All NCES information products include the names of knowledgeable staff that can assist users in understanding the information presented, and in determining whether there is an error that warrants action using the correction process described in this section. Users of NCES information should consult with the contact person listed in the product before filing a formal request for correction.

Information Correction Requests

In the Department of Education's correction request process, the burden of proof rests with the requester. An affected person who believes that information the Department disseminates does not adhere to the information quality guidelines of OMB or the Department, or an office of the Department that has issued program-specific guidelines, and who would like to request correction of specific information, needs to provide the following information:

- Identification of the requester (*i.e.*, name, mailing address, telephone number, and organizational affiliation, if any);
- A detailed description of the information that the requester believes does not comply with the Department's, OMB's, or NCES guidelines, including the exact name of the data collection or report, the disseminating office and author, if known, and a description of the specific item in question;
- Potential impacts on the requester from the information identified for correction (*i.e.*, describe the requestor's interest in the information and how the requestor is affected by the information in question); and
- An explanation of the reason(s) that the information should be corrected (*i.e.*, describe clearly and specifically the elements of the information quality guidelines that were not followed).

This information should be provided to the Deputy Chief Information Officer for Information Management at the following address

Director, Information Management
Office of the Chief Information Officer
US Department of Education
RE: Information Quality Request
Room 4060, ROB-3
400 Maryland Avenue, SW
Washington, DC 20202

Alternatively, requesters may submit e-mail requests to the following address: "ocio.infoqualityrequest@ed.gov." Requesters should indicate that they are submitting an Information Quality Request in the subject line of the e-mail.

Review

The Director, Information Management, CIO (DIM/CIO) will review the request and determine whether it contains all the information required for a complaint. If the request is unclear or incomplete, the Department will seek clarification from the requester.

If the request is clear and complete, the DIM/CIO will forward it to the appropriate program office(s) for a response to the requester. The responsible office(s) will determine whether a correction is warranted, and if so, what corrective action it will take. Any corrective action will be determined based on the nature and timeliness of the information involved, as well as the significance of the error on the use of the information, the magnitude of the error, and the cost of undertaking a correction.

Comments about information on which the Department has sought public comment, such as rulemaking or studies cited in a rulemaking, will be responded to through the public comment process, or through an individual response if there was no published process for responding to all comments. The Department may choose to provide an earlier response, if doing so is appropriate, and will not delay issuance of the final action in the matter.

The Department is not required to change the content or status of information simply based on the receipt of a request for correction. The Department may reject a request that appears to be made in bad faith or without justification, and is only required to undertake the degree of correction that is appropriate for the nature and timeliness of the information involved. In addition, the Department need not respond substantively to requests that concern information not covered by the information quality guidelines.

Response

The Department will respond to all requests for correction within 60 calendar days of the DIM/CIO's receipt of the request, including requests that the Department elects not to process further. For requests that merit review -

- If the request is clear and complete, the Department's response will explain the findings of the review, or will inform the requester if more time is needed to complete the review, the reason(s) for the additional time, and an estimate of the time it will take to respond. The appropriate program office will be responsible for determining what action is necessary and, if an error was made, it will determine the appropriate level of correction.
- If the request is incomplete or unclear, the DIM/CIO, will seek clarification from the requester. In the case of an unclear or incomplete request, the requester may submit additional clarifying information if he or she so chooses. However, the deadline for the Department's review and response will be based upon the date the clarifying information is received.

Once a decision is made, the response will explain to the requester that he or she has a right to appeal the decision. Copies of all Department correspondence related to Information Quality Requests will be maintained by the DIM/CIO.

Appeals

If a requester is not satisfied with the Department's decision on the request (including the corrective action, if any), he or she may appeal to the Department's Chief Information Officer within thirty (30) calendar days of receipt of the Department's decision. This administrative appeal must include a copy of the initial request, a copy of the Department's decision, and a letter

explaining why he or she believes the Department's decision was inadequate, incomplete, or in error.

This appeal information should be provided to the Department's Chief Information Officer (OCIO) at the following address:

The Chief Information Officer
US Department of Education
RE: Information Quality Appeal
Room 4082, ROB-3
400 Maryland Avenue, SW
Washington, DC 20202

Alternatively, requesters may submit an appeal by e-mail to the following address: "ocio.infoqualityappeal@ed.gov."

Requesters should indicate that they are submitting an Information Quality Appeal in the subject line of the e-mail. Such e-mail requests must include all of the information specified for an appeal submitted by regular mail.

The Department will ensure that all appeals are subjected to an impartial review that is conducted by parties other than those who prepared the Department's decision. The Department will respond to all appeals within 60 calendar days of the CIO's receipt of the appeal, or will inform the requester if more time is needed to complete the review of the appeal, and the reason(s) for the additional time.

Session 6

Preserving the Past, Linking to the Future

Evolution in Access Services for Electronic Records in the U.S. National Archives [12]

Margaret O. Adams
National Archives and Records Administration

The National Archives' program for electronic records has had a user-orientation throughout its history. Its creation was, in part, a response to the concerns of some of the nation's economists and historians. They and National Archives and Records Service (NARS) archivists understood by the early 1960s that the computer-readable data created in the administration of federal government programs represented irreplaceable primary documentary material for both short and long-term policy and social scientific analysis, as well as for historical research.

To document the need for concerted effort to assure preservation and access to valuable federal data, a Committee on Preservation and Use of Economic Data, sponsored by the Social Science Research Council undertook to study providing access to federal statistical records. Supporting the study, the Office of Statistical Standards, Bureau of the Budget, with help from NARS, inventoried machine-readable data in some Federal agencies.

The Committee's 1965 report, informally known by the name of its chairman, Yale University economist Richard Ruggles, urged the Bureau of the Budget to create a new federal agency, a Federal Data Center, and used the 50-page inventory of machine-readable data held by federal agencies to bolster its proposal. It envisioned an agency that would provide systematic and comprehensive coverage of the material of its areas of competence, analogous to the Library of Congress. The report also suggested that the proposed new center could serve the same function for machine readable statistical data "as the [National] Archives now does in the area of basic [paper or microfilm] records and documents . . ." and would need the type of "interagency authority that the National Archives had."

In other words, the proposed new center was to be modeled partially on the Library of Congress and partially on the National Archives, as the committee members understood the respective roles of those institutions. The primary functions for the proposed center were support and services for machine readable data "so that within the proper safeguards concerning the disclosure of information, both federal agencies and users outside of the government would have access to basic data." After reviewing the report the Bureau of the Budget appointed its own task force to consider "measures which should be taken to improve the storage of and access to U.S. Government statistics." Its recommendations supported and broadened those in the Ruggles report. Nonetheless, controversy over privacy issues and fears about the "big brother" aspects of a national databank doomed the proposals of both reports, as did recognition by some in the U.S. Congress that NARS already had statutory authority to accession records regardless of media and

[12] Paper prepared for presentation at the FCSM/COPAFS Seminar, Bethesda, MD, November 6, 2002. It is based upon a lengthier chapter on this topic by the author in a forthcoming monograph to be published by Scarecrow Press. The presentation paper includes no citations; all are available from the author, upon request. The views and opinions in this paper are the author's and do not necessarily represent the official policy of the National Archives and Records Administration.

that NARS had experience preserving confidential, security classified, or otherwise restricted government records.

As Thomas Brown has described in his presentation here today, about the time the Bureau of the Budget issued its recommendations for a national data center, then Archivist of the U.S., Robert H. Bahmer, established an internal NARS Committee on the Disposition of Machine-Readable Records. Its 1968 report echoed many of the themes in the Ruggles and in the bureau's reports, but diverged from their primary recommendation on the creation of a new federal data center. By doing so, the NARS report laid the foundation for the emergence of NARS' program for machine-readable records.

The sentiments expressed in all the reports directly influenced the evolution of reference services in the data archives program NARS created later in 1968. As if to emphasize that a data archives program had to be responsive to social scientists, the NARS report described the needs of economists for machine-readable federal statistical data, both historical and contemporary, as "voracious," concluding that "to establish the nature and degree of economic trends, old raw data is as valuable as new."

The first activity of the NARS Data Archives Staff was a survey of the magnetic tape libraries in the Federal government. This was in keeping with archival practices and necessary for identifying computer-readable files of possible long-term value. And, it responded to another of the recommendations in the Ruggles Report. During the survey, NARS staff found what the economists had suggested: "every agency had its own group of academics and researchers who knew all about their own records but were not knowledgeable about any …[others]. …[N]obody knew where the records really were, and only vague clues were available from some of the published statistical tables…."

The machine-readable archives program began "to furnish reference services on its holdings" as soon as it had accessioned records, which, as Brown mentioned, occurred in April 1970. An undated paper by Gerald Rosenkrantz, who became Director of the Data Archives Staff in September 1970, makes clear that the expectation for reference services for accessioned machine-readable files was that NARS would provide researchers copies of individual [full] files on a cost-recovery basis. This was the service the social scientists wanted. It meant that NARS data processing needs for a reference services program were limited to tape or file copying. Once NARS became aware that some federal agencies were creating computer-readable "document location indexes" there was additional anticipation of a future need to be able mechanically to search such files.

The work plan for FY 1973 mentioned that "the reference workload is accelerating as the branch becomes better known" and that the branch was negotiating the transfer of several files with "public demand." The Chief reported that in FY 1972, the Branch copied approximately 250 reels of tape [files] for researchers, and expected the volume to grow to about 800 in FY 1973. The work plan for FY 1974 reveals a growing staff, with four new people to be funded from a contract with the National Technical Information Service (NTIS), with whom NARS established a partnership for continuing to inventory magnetic tape libraries in federal agencies. The plan

also noted that the transfer of aviation data from the Civil Aeronautics Board (CAB) made NARS the supplier of historical and contemporary statistics for the airline industry.

In a January 1974 published interview with our discussant today, Connie Citro, who was then the editor of the *Review of Public Data Use*, Rosenkrantz candidly described NARS' machine-readable records accessioning and reference program. He distinguished between NARS and the earlier proposed federal databank, making clear that an archives has no right to translate or change any data [records] that it receives. He noted that NARS was handling "the complete public release of records for two small regulatory agencies, the CAB, and the Securities and Exchange Commission (SEC)." Neither agency had a revolving fund into which they could deposit revenues to offset the costs of providing copies of their records, so these agencies were pleased that NARS did and could offer this service. In return NARS received the records early in their life-cycle, when potential accessioning problems would be minimized.

Elaborating, Rosenkrantz unabashedly revealed some of the motivation of the NARS program. "We decided to concentrate on regulatory agencies and some of the statistical bureaus, …[because they had files in high public demand]. … We have operated on what might be called an opportunistic basis…, but the long-range goals have never really changed. We need a reference operation with competent people. You can theorize all you want, but you won't learn any better than if you actually have files which users want…. You won't learn [to solve] technical problems…unless you have operating experience. You can't sit on …tapes [that are] highly classified and then expect to read and service them properly [in] 25 years…if you've never done anything until then." With the interview, the *Review of Public Data Use* printed a partial list of data holdings of the National Archives: 14 series in 9 Record Groups. (Record groups correspond, in general, to a federal bureau, agency, or department.) The RPDU list served as an informal catalog until NARS' published in 1975 a *Catalog of Machine-Readable Records in the National Archives of the United States*. It described 75 series in 15 Record Groups. A second edition in 1977 described 120 series in 18 Record Groups.

As Rosenkrantz anticipated, providing reference services for federal records of high public interest, -- responding to researcher inquiries about the records, providing tape copies of files (or extracts from files), and describing the records -- provided valuable hands-on experiences for NARS' staff. In FY 1979, they completed 1350 responses and copied 943 files of accessioned and temporary machine-readable files. This level of activity suggests the experience gained from serving a category of researchers new to NARS: quantitatively-oriented, computer-using, academic social scientists and private sector analysts. From all reports NARS staff met their expectations.

Brown has detailed the collapse of momentum in NARS' Machine-Readable program in the 1980s. Suffice it to say that severe staff reductions negatively impacted all parts of the program, including its reference services. But the early 1980s also marked the transfer to NARS of data files with records for individual casualties of the Korean and Vietnam wars. Transfer of those records altered forever the mix of researchers who sought reference services from NARS' electronic records program, and presaged rising expectations for record-level access to archival electronic records that figures prominently to this day.

No third edition of the *Catalog* ever was published, and while the catalog database for accessioned electronic records ceased to be actively maintained, it still lives. The staff continued outreach to researchers by publishing the first *National Archives Computer Data Bulletin* in Spring, 1981. It highlighted some new accessions including operational records from the Vietnam war, and accretions to statistical series previously described. The second, and final …*Bulletin* was not issued until Spring, 1985. By then the Branch had curtailed many services but basic file copying continued, though not always with the timely turnaround that researchers sought.

Remarkably, during the 1980s the scaled-down branch also rose to the challenge of the new demand for record-level access to the casualty records. Patterning on services the Department of Defense had offered prior to transferring the casualty databases to archival custody, NARS staff produced extract "state lists" in printout form from the databases. In the printouts, literal meanings substituted for coded data, making the records humanly readable. The electronic files from which the casualty lists were printed to paper served in 1998 as the source that enabled electronic records staff to post state-level casualty extract lists on the NARA homepage, a first realization of electronic access to NARA's electronic records. The public response to this online access has been overwhelmingly positive, has spurred new kinds of inquiries, and raised new service expectations.

Towards the end of the 1980s, the electronic records program began to regain momentum and in FY 1989, staff completed 2003 responses to inquiries and copied 1231 files for researchers. For reference services, one of the first projects in the rebuilding phase was to reestablish descriptive efforts by reconstituting a Title List of holdings.

Electronic records reference services evolved during the 1990s, as we expect they will into the indefinite future, by utilizing new technologies. Technology, and a dedicated though small staff, have been key to coping with an increasing volume of inquiries and to rising expectations for types of services. Those increases, in turn, reflect growth in the scope and variety of the electronic records federal agencies have transferred to NARA, as well as, by the end of the decade, the ubiquity of powerful home computers and the Internet. By the end of the 1990s, accessioned electronic records files numbered in the neighborhood of a 150,000, including a substantial representation from federal statistical agencies.

Innovations included reference services by email beginning in March 1991; offering copies of files of electronic records on CD-R and/or diskette in FY 1997; and towards the end of the decade, mounting on the NARA homepage all the informal reference reports prepared over the years, as well as a public extract of the title list. While the latter has its uses, it now identifies only about ten percent of the accessioned holdings. Every new service or information offering has caused a spike in demand for current and also for new kinds of access. Offering file-level access, that is, copies of electronic records files that researchers can keep or redistribute, and use in an unlimited manner, with their own computing hardware and software, continues to be popular. This form of access meets the needs of analysts but is of limited usefulness for the researcher seeking specific information preserved in the records but who has neither the ability, interest, nor institutional support for undertaking data analysis.

The electronic records reference services program was insulated from the direct impact of the *Armstrong et al v. Executive Office of the President et al* case that dominated life in NARA's electronic records program for several years in the 1990s, but the overall challenges and demands stemming from the litigation clearly took a toll. Routine preservation work suffered while resources were drained to meet court-imposed preservation and related requirements. Development of online record-level access to any of NARA's accessioned electronic records was postponed. Plans to experiment with FTP as a mode for providing copies of electronic records files went to a back-burner.

In FY 1999, the electronic records reference staff completed 4226 responses to inquiries and copied 2133 electronic records files for researchers. The responses covered records in 58 record groups and in donated historical materials; the electronic records files copied for researchers that year came from 25 record groups and from donated historical materials. The file most frequently copied (approximately ten times a year), is one of the 137 files from the Ownership Reporting System (insider-trading data) series, Records of the Securities and Exchange Commission. The insider-trading records are perennially in demand.

On an annual basis, about half of the reference demand is information "from" records, and essentially represents requests for "record-level" access to electronic records. Of this demand, more than half tends to relate to records in the military record groups in which series of casualty and prisoner of war records are preserved. The remainder of demand divides between inquiries seeking information "about" records, which can be a prelude to seeking information "from" records or to placing an order for records reproductions, and the category called "other. Requests related to records from the federal statistical and/or regulatory agencies are dominant in the "about" records category and attest to the continuing interest in ordering copies of archival electronic data files of this type, even as expectations for record-level access to other types of electronic records are rising.

Some very brief comments on "who" the researchers are who have used NARA's electronic records in recent years. They are, after all, the "future" of ages past; they are the benefactors of NARA's 30-year program to preserve and provide access to electronic records. They are everyman and everywoman, from the highest levels of government to the solitary citizen. They use archival electronic records usually in ways unrelated to the purposes for which the records were initially created, collected, compiled, etc. for purposes as disparate as the most sophisticated policy analysis to locating information concerning the fate of loved ones, and everything in between. Their individual stories are fascinating, but since telling even a few of them would take far longer than we have today, let me, share just one. Several years ago, electronic records reference staff worked with a reporter who was assisting the family of a U.S. military casualty of the Vietnam War, whom the reporter and family suspected might be that war's "unknown soldier." Using some in-house automated capabilities, they searched for, identified, and retrieved the casualty and air sortie records for the pilot and the mission in which he perished. As the reporter later noted, "the information we obtained from those electronic records helped us defend and maintain the integrity of the story. And that same data was used by the family as they fought with the Department of Department of Defense to get the Tomb of the Unknowns opened. Eventually DoD was persuaded by the overwhelming evidence and opened

the Tomb. DNA testing was done. And . . . Michael Blassie was buried near his boyhood home in St. Louis under a stone bearing his own name."

At the end of the 20th century, accessioned electronic records were not yet directly transferable, searchable or retrievable by the public across the Internet. To address the expectation for online access to electronic records, beginning in FY 1999, NARA has invested in two Information Technology projects. One has developed the capability to receive electronic files electronically, utilizing a standard known as "file transfer protocol," or, FTP and we expect to begin testing outbound FTP capabilities soon. The second project is aimed at offering online record-level access to NARA's electronic records holdings and is known as the Access to Archival Databases (or, AAD) resource. It offers the promise of online public access to a selection of accessioned electronic records in structured formats that are in high demand and allows searching and retrieving of specific records from within structured databases. We hope to begin offering public access to this resource next month. I have distributed a list of the series of archival electronic records that will be included in the first rollout of AAD and a general description of the resource.

Preserving the Past, Linking to the Future Discussion

Constance F. Citro
Committee on National Statistics
National Research Council of The National Academies

I am delighted to be here to discuss three excellent and thought-provoking papers. As a history buff and one whose professional career began in the late 1960s—about the time the National Archives began to establish an electronic data records access and use program—I was entranced to read the companion histories of the Center for Electronic Records (in Tom Brown's paper) and the Archives' electronic data access services (in Peggy Adams' paper). I was also captivated by the ideas for future that Ken Thibodeau presented in his paper.

I have only a few comments on the papers as such. For Brown's paper, it would help the reader if he were to add organization charts that trace the name changes and locus of the electronic records program within the Archives; similarly, if he were to add figures for staff size and budget for the entire Archives to enable the reader to grasp the relative size of the electronic records program over the decades. The charts in Adams' paper about electronic data access requests from users are helpful. They would be enhanced by comparison charts for access requests for other types of Archives records and, perhaps, for other electronic archives as well (e.g., the Interuniversity Consortium for Political and Social Research, ICPSR). I would also suggest that Adams add an explicit discussion of the confidentiality protections that Archives affords its electronic records. My main query about Thibodeau's paper has to do with the status of the Electronic Records Archives Program—is it an idea, an initiative, a program? I am delighted to learn that it has just now been given official status within the Archives. Finally, all of the papers should include a list of acronyms for the reader who is not familiar with Archives terminology.

The bulk of my remarks concerns themes and lessons that I think these three papers offer for the broader federal statistical system. I make three main points:

1. Archiving public electronic data is essential.
2. The history of the electronic records program at the Archives is both deeply inspiring and profoundly depressing; it parallels ups and downs experienced by federal statistical agencies.
3. The federal statistical system is currently in perilous straits. To help minimize the very real likelihood and consequent adverse effects of declining budgets, credibility, and independence, agencies in the system should: (a) reach out to other statistical agencies; (b) reach out to other relevant communities of expertise, such as computer science; (c) build documentation, evaluation, and preservation up front in major data collection programs; and (d) reach out to users, encouraging them to be proactive in supporting the system.

Archiving is Essential

You cannot use what you do not preserve. The statistical system should be glad that the Archives has an active electronic data access and use program and is well versed in techniques of record preservation across time and changes in media. However, Archives cannot, and does not desire to, hold more than a fraction of federal statistical data sets. Agencies need to be proactive in working out archiving plans for their data. Part of an agency's archiving plan should include consultation with Archives about which data sets to transfer to Archives and when. Another part of such a plan should be ways to provide access, use, and preservation services for data that Archives will not hold. For example, from its inception, the Bureau of Justice Statistics has deposited all of its electronic data sets with ICPSR. There should be no repetition of past incidents when valuable data sets were allowed to molder and almost be lost to posterity (examples are the data files for the "other"—i.e., not March—months of Current Population Survey supplements, for which Judith Rowe at Princeton arranged a rescue).

Inspiring and Depressing History

The history of programs for accessioning, preserving, and providing access to electronic data at the Archives is inspiring because it shows, over and over again, the dedication and perseverance of professional civil servants who have kept a needed program alive in the face of almost overwhelming forces against it. Such dedication and expertise of professional staff is evident throughout the entire federal statistical system.

The history of electronic records services at the Archives is also depressing because, so often, exogenous forces battered and threatened the program. Over four decades the program experienced—and barely survived—threats due to downsizing of government, pressure to contract for agency services with the private sector, centralization of information technology (IT) functions, vacancies in top positions, and unfunded mandates. Sometimes, such changes were implemented with careful planning; more often, they were implemented mindlessly with little thought about the particular needs of the small but vital program of electronic records access at the Archives.

Federal Statistical System in Peril

At this time it is my belief that the federal statistical system is in perilous straits, facing a confluence of exogenous threats. There is continued pressure to downsize government—without consideration that statistical agencies are already facing staff shortages due to retirements and recruiting difficulties. There is renewed pressure to contract out government functions—without consideration that statistical agencies must have sufficient in-house staff to ensure data quality and usability. There is pressure to centralize information technology—without consideration of the need to protect the confidentiality of respondents and the credibility of federal statistics. There is pressure to centralize media relations and contacts with outsiders—without consideration of the need for statistical agencies to maintain independence. There are unfunded mandates and vacancies in key agency positions. There are fewer champions of statistics in the Congress. There are overt threats to statistical agency independence, such as the provision in the 2001 Patriot Act for access to confidential data from the National Center for Education Statistics.

Finally, there are strong and growing pressures to reduce budgets (or, at best, hold them steady) for agencies, like statistical agencies, whose role is vital for the maintenance of our free, democratic and capitalist society, but whose value is not fully appreciated and is not directly tied to the war on terrorism.

Responding to these threats to the federal statistical system will be challenging, particularly in view of how decentralized the system is. I offer four suggestions to statistical agencies:

First, reach out to other agencies in the system. Such reaching out is inherent in the mission of the National Archives. Mechanisms to foster cooperation among statistical agencies exist as well, but they need to be strengthened. When evaluating individual initiatives for cooperation, each agency needs to put aside turf concerns as much as possible in order to strengthen the system as a whole. These perilous times do not allow the luxury of turf battles. No agency is immune from threat; therefore, every agency should welcome cooperative efforts that enhance the system's overall capabilities even if no individual agency gains all it originally wanted.

Second, reach out to other relevant communities of expertise. The most heartening part of Thibodeau's paper on the development effort for the Electronic Records Archives Program is the relationships the Archives has built with the supercomputing world in academia and e-government initiatives at such agencies as the Patent and Trademark Office and the National Science Foundation. Archives knew it could never command the resources to develop the computer systems it needed for electronic data, but it could—and did—leverage its scant resources to foster and benefit from the initiatives of others.

In a small example of the kind of reaching out that would benefit the federal statistical system, the Committee on National Statistics last spring held a workshop on survey automation techniques, funded by the U.S. Census Bureau. The workshop brought together survey researchers with computer software engineers and developers. The discussions identified fruitful ways in which private sector software documentation, development, and testing tools could be used to facilitate the job of statistical agency staff who are turning complex survey instruments into computer-assisted interviewing software code. Such outreach to the computer science community should continue and grow—it can help the statistical system develop better data systems with less investment of scarce in-house time and resources.

Third, build documentation, evaluation, and archiving up front into the development of statistical data systems. The Archives has plans for government agencies to use e-government software that enables agency staff do their work electronically and at the same time create a well-documented and organized set of electronic records that are readily preserved for future use. Statistical agencies should similarly strive to develop software systems that facilitate good documentation, ready availability of data samples for timely evaluation, and, ultimately, the ability to preserve important data sets for the future. The Census Bureau is currently developing a Master Trace Sample (MTS) of sampled addresses from the 2000 Census Master Address File with information from every step of data collection, processing, and tabulation. The purpose of the MTS is to facilitate not only in-depth evaluation of 2000 census processes and their effects, but also to provide a simulation database for testing proposed methods for 2010. For 2010 the Bureau's goal should be to build MTS capabilities into its data management and processing

systems from the outset, so that evaluation can be more timely and the ability of the sample to support future census planning can be enhanced.

Fourth, reach out to users. Federal statistical agencies already do a good job of communicating with users about data products and services. They need to further inform users of the threats they are facing, and users, in turn, need to step up to the plate. Instead of assuming that the case for a strong federal statistical data system is self-evident to right-thinking people, users need to be proactive in their support for the system with key decision makers.

In conclusion, I compliment the three paper authors and commend the lessons in their papers to the broader statistical community. It is very rewarding to study history; it is even more rewarding to learn from the past to improve the present and the prospects for the future.

Session 7

Benefits and Stewardship of Linked Survey and Administrative Data

Data Stewardship and Accountability at the U. S. Census Bureau[13]

Nancy A. Potok and Gerald W. Gates
U.S. Census Bureau

Introduction

Statistical agencies have long recognized the fundamental tension between their mandate to provide high-quality data that informs sound research and public policy development and their requirement to protect the privacy and confidentiality of their respondents. These dynamics often operate at odds with one another, as demands for richer data products face off against increasing public concerns about privacy, the increased availability of personal information on the internet, and newer, cheaper desktop data processing capability. However, a statistical agency's reputation for respecting privacy and confidentiality is critical to maintaining high response rates and, thus, the quality of its data.[14] The U.S. Census Bureau's mission to be the "preeminent collector and provider of data on people and the economy of the United States," requires that this tension be balanced successfully.

The Census Bureau's legal mandate, Title 13 of the United States Code, authorizes the collection of data, but it also establishes strict requirements for maintaining the confidentiality of data collected from its respondents. Indeed, the Census Bureau may not publish data about a particular establishment or individual that allows them to be identified. Even when the Census Bureau requires expert consultation from outside the agency, such experts are not permitted access to the data unless they are brought on as "Special Sworn Status" individuals[15] – effectively temporary staff – who are sworn to uphold the Census Bureau's confidentiality standards. Criminal penalties, specifically up to $250,000 in fines and 5 years imprisonment, further help to create an environment intolerant to such disclosures. Given the agency's strong legal mandate and ethical commitment to privacy and data confidentiality, how does it ensure that collected data result in useful, relevant and timely products?

[13] This paper has undergone a review more limited in scope than that given to official Census Bureau publications. It is released to inform interested parties about the Census Bureau's data stewardship approach to balancing confidentiality protections while providing quality data and to encourage discussion of these important issues.

[14] See Pat Doyle, Julia I. Lane, Jules J.M. Theeuwes, Laura V. Zayatz, Eds., Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies for a series of discussions on the tension between data access and confidentiality.

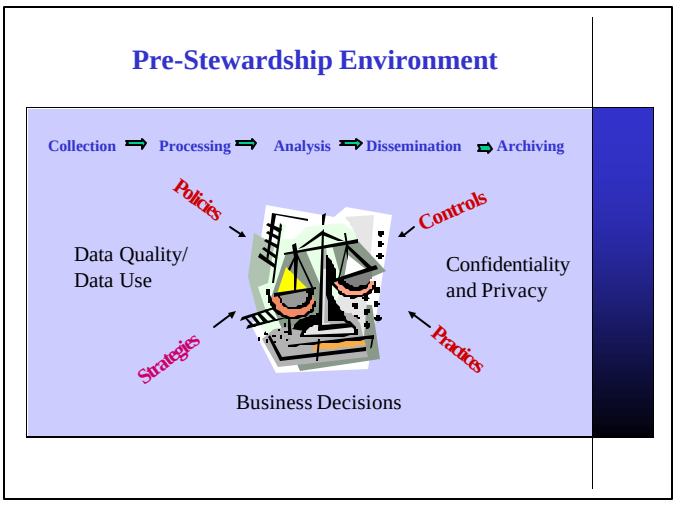
[15] Title 13 United States Code, Section 23(c) provides for the Census Bureau to "utilize temporary staff, including employees of Federal, State, or local agencies or instrumentalities, and employees of private organizations to assist the Bureau in performing the work authorized by this title," but only if such temporary staff is sworn to observe the limitations imposed by section 9 [which establishes confidentiality provisions].

A sound *data stewardship structure* within which such issues can be weighed provides a forum where the Census Bureau's can make balanced business decisions – data quality and access on one side of the scale and privacy and confidentiality on the other. The concept of "stewardship" is borrowed from environmentalists – the objective being to create a *sustainable balance* that supports one's needs over the long term.

Establishing a Basic Data Stewardship Structure

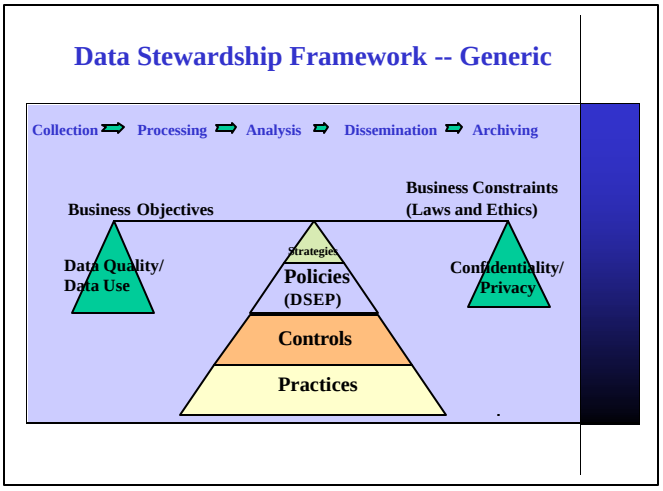
While data stewardship principles may exist, they are not always well coordinated or integrated, and/or they are applied in an ad hoc manner, depending on the particular circumstances involved. **Chart 1** demonstrates how business decisions that affect data-related operations -- collections, processing, analysis, dissemination, and archiving -- can become unbalanced and lose a corporate focus when there is no integration of strategies, policies, controls or practices, or they are not used systematically to make business decisions.

Chart 1 --



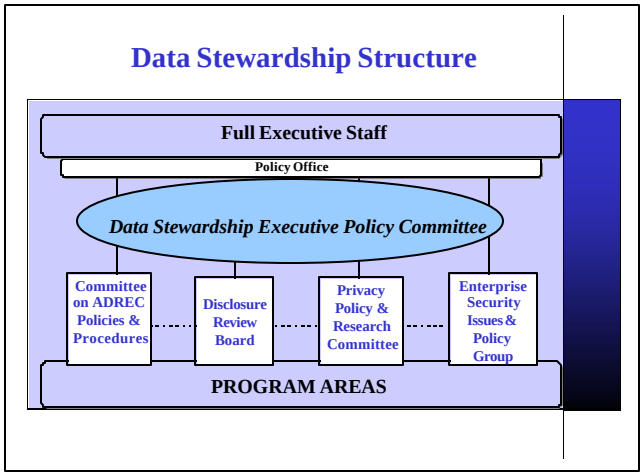
If strategies, policies, controls, and practices are fully integrated, the organization has a better chance of ensuring that business decisions will lead to the desired outcome. **Chart 2** illustrates how an otherwise *ad hoc* approach can be stabilized, achieving balance between business objectives and constraints. This better supports the data related operations.

Chart 2 --



The Census Bureau annually updates its 5-year strategic plan and communicates its strategic goals to employees and external stakeholders. In June 2001, the Census Bureau moved to address policy issues more consistently by establishing the *Data Stewardship Executive Policy (DSEP) Committee*. The DSEP Committee is composed of top bureau executives who are charged with identifying and developing policy issues related to data stewardship. This executive decision-making body is staffed by the Policy Office and supported by the analyses and recommendations of four DSEP staff committees: the Committee on Administrative Records Policy and Procedures (CARPP), the Disclosure Review Board (DRB), the Privacy Policy and Research Committee (PPRC), and the Enterprise Security Information and Policy (ESIP) Committee (see **Chart 3**).

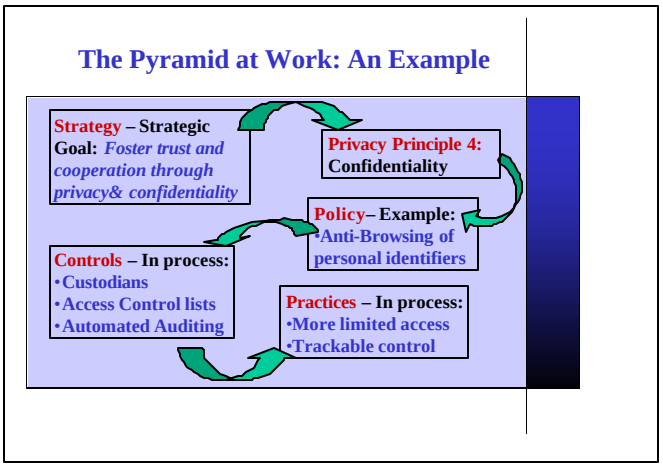
Chart 3 --



One goal of the DSEP Committee is to ensure that strategic goals, corporate ethics, policies, controls, and operational practices are integrated and consistent. This means that strategic goals are shaped by corporate ethics and drive policies. Policies in turn drive the creation of organizational controls, and these controls incorporate practices that ensure compliance. For

example, as shown in **Chart 4,** one of the Census Bureau's strategic goals is to *foster trust and cooperation through privacy and confidentiality.* In support of this goal, the Census Bureau developed a set of ethical standards called Privacy Principles, one of which is *Confidentiality.* This Privacy Principle resulted in the Census Bureau adopting a policy *prohibiting the browsing of records with personal identifiers by employees* and others who may have access to those records. The Census Bureau is currently working to establish *access control and auditing procedures,* such as identifying data custodians in each division responsible for monitoring access to personal identifiers. The result will be that fewer employees will have access to sensitive records, and those that do will have all their interactions with the data tracked and monitored by an automated audit system.

Chart 4 --



The DSEP structure has been successful in systematically establishing policies and procedures in several key areas. Accomplishments include the release of an *Administrative Records Handbook,* and documenting procedures for the negotiation, acquisition, access, and use of administrative record data. The DSEP Committee also has finalized a policy on appropriate data access and use for non-employees with Census Bureau Special Sworn Status. It is currently completing an analysis of how well existing policies support the Privacy Principles.

While the primary responsibility of the DSEP Committee is to serve as the policy-making body, it also gives considerable attention to controls and practices. However, translating policy decisions into day-to-day operational practices is a highly human resource-intensive activity. As a result, policy implementation is moving ahead more slowly than was originally anticipated. The Census Bureau has handled this challenge, in part, by establishing a new Policy Associates Program, which details competitively selected Census Bureau program staff for one year to the Policy Office to help implement new data stewardship policies.

Data Stewardship and the Use of Administrative Records

The benefits and stewardship of linked survey and administrative data, the subject of this panel, are of great interest to the Census Bureau's DSEP Committee, which uses its data stewardship framework to guide and support use of administrative records for statistical purposes. Using the

approach introduced in **Chart 4** above, the Census Bureau first looked to its strategic plan and whether administrative record data would support its goals. The Bureau's strategic goal of "*Fostering an environment that supports innovation, reduces respondent burden, and ensures individual privacy,"* supports use of data from administrative records. They minimize the cost of direct data collection, reduce the burden on respondents, improve and enhance census and survey collections, and enable the development of improved data products that inform public policy. This strategic goal drives the development of policies that balance the benefits of administrative record use against privacy and confidentiality concerns, particularly given that these benefits are primarily derived from linking administrative records to other datasets.

Policy issues surrounding use of administrative records are identified by the DSEP Committee, with subsequent policy analysis and recommendations developed by the CARPP (see **Chart 3** above). In addition to weighing the needs of the data user community and the public, the CARPP must give special consideration to the Census Bureau's data providers, including managing and safeguarding data in accordance with their legal authorities and policy requirements. The CARPP and the DSEP Committee have established a number of procedures for managing the use of administrative records at the Bureau.

Procedures for managing administrative records include consistent review criteria for all proposed projects; centralized custodial functions to control data access on a "need-to-know" basis; and centralized tracking of administrative record projects. In addition, personal identifiers on administrative records (e.g., Social Security Number and name) are maintained in a restricted environment by the custodian. Identifiers are stripped from the records before they are released to researchers. When necessary, the custodian replaces the personal identifiers with a "Protected Identification Key," or "PIK," to enable record linkage. Currently, the CARPP is developing a policy to guide the Bureau's record linkage activities, again seeking the balance between developing relevant, high-quality data products and providing appropriate privacy and confidentiality protections to respondents.

Enhancing the Basic Structure

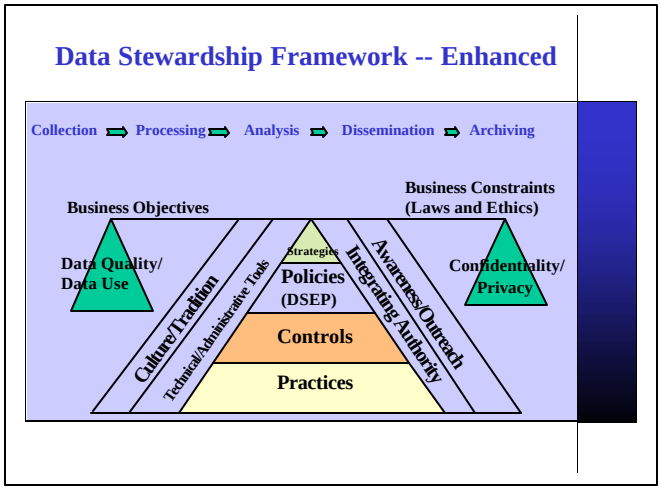
Although the basic data stewardship structure provides a mechanism for balancing data quality and access with privacy and confidentiality, that balance is still somewhat precarious. Looking back at the generic framework in Chart 2, it is useful, then, to consider ways to further stabilize this structure.

The Census Bureau has considered a number of sources for guidance in strengthening its data stewardship approach. First, it conducted a benchmarking exercise, making structured inquiry of six best practice-oriented private and government organizations about their policies, agency structures, and roles with regard to privacy. It also conducted a literature review consisting of recent privacy research both at the Census Bureau and elsewhere. The Census Bureau also drew on a General Accounting Office report issued in April 2001, *Record Linkage and Privacy: Issues in Creating New Federal Research and Statistical Information*, which provides a toolkit of

approaches to support data stewardship.[16] Lastly, the DSEP Committee commissioned an evaluation of the DSEP structure (executive body plus four staff committees). The evaluation targeted four areas for improvement -- the need to focus on employee awareness of the data stewardship structure; include stakeholders in policy discussions; be more systematic in assessing the operational impacts of policies; and restructure the role of the Security staff committee. The assessment activities also identified four key components that can help stabilize the data stewardship structure – culture and tradition, technical and administrative tools, awareness and outreach, and an integrating authority.

As shown in **Chart 5,** adding these steps to the data stewardship pyramid helps achieve a more stable balance between data access and use, on the one hand, and data protection, on the other.

Chart 5 --



- *Culture and tradition* form the basis for a statistical agency's approach to data stewardship. Al Zarate, Confidentiality Officer at the National Center for Health Statistics (NCHS) describes the Census Bureau as having a "culture of confidentiality."[17] Some organizations have cultures that focus predominantly on access to information. In an academic environment, for example, information sharing is the lifeblood of learning. The primary focus is on sharing research, not limiting access. Other organizations, like the National Security Agency, place a priority on keeping information highly controlled and access limitation is paramount. Survey organizations would not continue to do business without a focus on both confidentiality and access. The Census Bureau's culture and tradition fit this model well.

[16] U.S. General Accounting Office, *Record Linkage and Privacy: Issues in Creating New Federal Research and Statistical Information.* GAO-01-126SP, April 2001.

[17] Al Zarate, *Government Perspective on Data Stewardship for Statistical Data.* Paper presented for panel, "Statistical Data Stewardship in the 21st Century," Joint Statistical Meetings, New York, NY, August 11, 2002.

- *Technical and administrative tools* play an important role in a well-grounded data stewardship structure. Today, most organizations control disclosure by providing safe settings, where data can be used for legitimate statistical purposes, and by releasing safe data, where the data have been modified to hamper those who attempt to identify individual respondents. These tools allow organizations to more effectively accomplish the business objective of providing access to data while also ensuring confidentiality. They also play a role in restricting access and limiting uses within the organization. Need-to-know access and file-level auditing ensure that employees are not tempted to browse records or give others access, regardless of the motive. In deciding what tools to apply, the organization must be aware of external threats, assess the physical constraints on users, and take into consideration the impact on utility of the data for intended research.

- *Awareness and outreach* activities help ensure that business decisions are based on the valid concerns of external stakeholders, including respondents, privacy advocacy groups, and the data user community. Without adequate research and data on privacy attitudes and behaviors and data needs, it is easy to fall into an endless loop of supposition and speculation in the policy development process. The Census Bureau has conducted privacy attitude surveys for the past decade, to measure the public's awareness of confidentiality requirements and gauge concerns over the use of administrative records. Attitude surveys, focus groups, and cognitive interviews play an important role in understanding awareness of organization practices and identifying practices that may be misunderstood or not be acceptable. Messages that are conveyed to employees and to the public help reassure that data uses are important and that protections are appropriate. Message wording benefits from cognitive testing to ensure that what is intended is what is understood.

  An agency's marketing activities also support the agency's outreach efforts by emphasizing the organization's objectives and constraints and how its culture, tools and legal authority enforce its approach to data stewardship. It is critical, however, that messages accurately reflect practice (i.e., the "talk matches the walk") -- saying you do something when you don't can be worse than not saying anything at all.

- An *integrating authority* is critical to ensure integration of strategies, policies, controls and practices and to make most effective use of culture, tools and awareness. This typically entails a role for persons or groups to decide or advise on policies, controls and practices. The National Center for Health Statistics (NCHS) enlists its confidentiality officer for this purpose, who provides internal advice on data protection and access decisions. The Canadian government has established a Privacy Commissioner, who provides counsel and direction on matters affecting the privacy of Canadian citizens. Statistics Canada also has a privacy and confidentiality officer. In other instances, agencies are subject to Institutional Review Boards that review and approve survey research affecting human subjects. NCHS and the Census Bureau have also established Disclosure Review Boards to review and approve all publicly released data. Lastly, there is a trend among U.S. institutions to name a Chief Privacy Officer whose responsibility it is to implement privacy policies across the organization. Legislation recently enacted to establish a Department of Homeland Security requires affected federal agencies to establish a Chief Privacy Officer.

In short, there are several non-mutually exclusive options for establishing an integrating authority, all providing varying degrees of control. Some are purely internal, some external, and some provide a combination of the two orientations. The use of external decision makers is controversial and often resisted, but part of that resistance stems from a concern that such counsel generally lends itself to advocacy of privacy and confidentiality to the exclusion of balancing those concerns against the agency's need to provide quality data products. A redirection of the integrating authority's focus to a balanced data stewardship approach may alleviate this concern.

Conclusion

At this writing, the Census Bureau is deliberately working towards full implementation of the enhanced data stewardship framework illustrated in Chart 5. There are several data stewardship issues that will influence the way the Census Bureau – and the federal statistical community in general -- will function this decade. The impact of recent legislation like the USA Patriot Act and future implementation of new data sharing legislation (H.R. 2458), which passed through Congress in November 2002, need to be assessed and addressed. Additional challenges continue to arise.

As the Census Bureau explores the potential of using administrative records for statistical purposes, it needs a clear policy on record linkage methodology and standards for obtaining informed consent from respondents to conduct such matches. Also, administrative record procedures must include adequate controls on access and use of these data, which must be maintained in accordance with the requirements of the providing agencies. The Census Bureau is currently responding to new Office Management and Budget requirements for Privacy Impact Assessments, building on the Privacy Principles developed within the parameters of the data stewardship structure. A broad range of disclosure limitation approaches that permit safe release of data for public policy uses, must be developed, including contracting with experts to attempt unauthorized links of public data sets, and developing synthetic data sets to permit public users access to data while reducing the risk of identifying respondents.

Lastly, a key point bears repeating: developing and maintaining a viable data stewardship structure requires a significant commitment and investment of resources from an agency. Nevertheless, this more structured approach to data stewardship is integral to striking a balance between the tensions inherent in meeting data user needs and honoring the privacy and confidentiality of its respondents. In the end, privacy and confidentiality -- which are typically perceived as business constraints – can actually enable an agency's mission and business objectives by establishing the public's trust and cooperation as respondents.

Acknowledgments

The authors wish to thank Eloise Parker for her role in the preparation of this paper and Eloise Parker, Wendy Alvey, and Ta Shunna Marshall for their assistance with the presentation.

Recommended Resources on Data Access and Confidentiality

Doyle, Pat. Julia I. Lane, Jules J.M. Theeuwes and Laura V. Zayatz. *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies.* North-Holland, published in conjunction with the Census Bureau (2001).

Duncan, George T. Thomas B. Jabine, Virginia A. de Wolf, Eds. *Private Lives and Public Policies.* Panel on Confidentiality and Data Access. Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, and National Research Council. Washington, DC: National Academy Press (1993).

Federal Committee on Statistical Methodology (FCSM) (May 1994). *Report on Statistical Disclosure Limitation Methodology*. (Statistical Working Paper 22). Washington, DC: Office of Management and Budget, Office of Information and Regulatory Affairs, Statistical Policy Office.

Holz, V. Joseph, Robert Goerge, Julie Balzekas and Francis Margolin. *Administrative Data for Policy-Relevant Research: Assessment of Current Utility and Recommendations for Development.* A Report of the Advisory Panel on Research Uses of Administrative Data of the Northwestern University/University of Chicago Joint Center for Poverty Research (January 1998).

H.R. 2458. E-Government Act of 2002, containing the Confidential Information Protection and Statistical Efficiency Act of 2002. Legislation passed both Houses of Congress by November 15, 2002; pending presidential signature at this writing.

Privacy Protection Study Commission. *Personal Privacy in an Information Society* (July 1997). The principle of functional separation is addressed in Chapter 15.

U.S. Census Bureau. *Administrative Records Handbook.* May 2001. For inquiries about the handbook, please contact Eloise Parker, Administrative Records Coordinator, U.S. Census Bureau, FOB 3, Room 2430, Washington, DC 20233; (301) 763-2520; eloise.k.parker@census.gov.

U.S. General Accounting Office. *Record Linkage and Privacy: Issues in Creating New Federal Research and Statistical Information.* April, 2001. GAO-01-126SP.

Zarate, Alvan. *Government Perspective on Data Stewardship for Statistical Data.* Panel, "Statistical Data Stewardship in the 21st Century," Joint Statistical Meetings, New York, NY, August 11, 2002. (American Statistical Association CD-ROM Proceedings in process.)

Zarate, Alvan O. Jacob Bournazian and Virginia de Wolf. *Integrating Federal Statistical Information and Processes*. Federal Committee on Statistical Methodology (FCSM) Committee on Data Access and Confidentiality (November 8-9, 2000). <www.fcsm.gov/cdac/cdacpaper.pdf >.

SSA Policy Applications of Administrative Data Linked to SIPP[18]

Howard M. Iams
Social Security Administration

Abstract

The Social Security Administration (SSA) conducts policy analysis with the data from the Survey of Income and Program Participation matched to extracts from SSA's administrative records. SIPP represents the social characteristics of the U.S. population; SSA administrative records contain information necessary to administer the Old Age and Survivors and Disability Insurance Programs and the Supplemental Security Income program. SSA assesses the impact of policy changes to programs it administers on the distribution of income and poverty with these SIPP matched data. Using these matched SIPP records, SSA develops micro-simulation models to assist policy evaluation. These include models of eligibility and participation in the Supplemental Security Income and the Qualified Medicare Beneficiary programs as well as the retirement income and life histories of future retirees from the baby boom, World War II, and Depression birth cohorts. SSA also describes the beneficiaries served by its programs with these SIPP matched data. This paper discusses examples of these uses by SSA.

I. Introduction

The Social Security Administration's (SSA) Office of Policy relies extensively on the Census Bureau's Survey of Income and Program Participation (SIPP) matched to Social Security Administration records of benefits and lifetime earnings. A major focus is the impact of Social Security policy alternatives on the distribution of income to various sub-populations. A second is the development of statistical simulations of a projected population for policy evaluation. Linked data also describe who is being served by the programs administered by SSA. The programs include Title II benefits for Old Age Insurance, Survivor's Insurance, and Disability Insurance and Title XVI Supplemental Security Income benefits for disability and old age. The purpose of this paper is to briefly describe examples of these uses at SSA.

The SIPP matched data combine the SIPP survey information with SSA's administrative records. The content of the SIPP is well known (see the user's guide, U.S. Department of Commerce 2001), and the data are publicly available from the Census Bureau. Less well known are SSA's administrative records containing the material necessary to administer the Social Security Act (see Panis et. al. , 2000). [19] The matched SIPP permits analysts to use detailed SSA program

[18] The positions in the paper represent the author's professional judgement and do not represent the position of the Social Security Administration.

[19] The Numident includes basic information on birth and death dates. The Master Beneficiary Record contains monthly benefit status and payment amounts for Title II programs from 1951 to current month, while the Supplemental Security Record for the Title XVI program contains monthly benefit information from 1974 to the current month. The records include the SSA 831 Form for application for disability from 1974 to the current month. The Master Earnings File (MEF) contains detailed earnings information including Medicare taxable earnings and uncovered earnings. The Summary Earnings Record extract from the MEF contains Social Security taxable earnings and quarters of coverage for each year from 1951 to the current year minus 2 years. The Detailed Earnings Record extract from the MEF contains information from the W-2 tax form including total earnings, self-employment income, and nontaxable earnings for defined pension plan accounts.

information in combination with the socioeconomic and demographic information contained in SIPP. Through a joint agreement, SSA and the Census Bureau match individual respondent information provided in SIPP to the SSA records for administering the program for respondents providing Social Security numbers in the survey. They match about 90 percent of the adults in the 1990-1993 panels, about 85 percent of the adults in the 1996 panel, and about 74 percent of children in the 1996 panel.[20] SSA and the Bureau restrict access to these matchable administrative records to sworn census agents with approved research projects. The processing of the restricted data must take place at a secure Census Bureau or SSA site.

II. Policy Estimates

A primary use of SIPP matched data is the distributional impact from policy changes. This paper reviews three policy analyses conducted recently at SSA: cost neutral policies for increased widow benefits, childcare credits, and the removal of the retirement earnings test. The SIPP matched data were necessary for analysis of the distributional impact of policy change. The important function of the SSA administrative records is to provide SSA administrative details on benefits and lifetime earnings. Many survey respondents do not know these administrative details or would imperfectly recall a lifetime history. Examples would include the extent to which earned retired-worker benefits offset higher auxiliary benefits as a spouse or survivor and the lifetime history of annual earnings taxed for Social Security purposes (which changes across time). The important function of SIPP is to provide socioeconomic and demographic characteristics of a nationally representative sample including income, assets, marital history, fertility history, and pension coverage. In addition to these characteristics, the SIPP links husbands and wives in married couples. [21] Analysis of specific Social Security policy options requires both sets of information contained in the SIPP linked to SSA administrative data.

Widow Benefit Change

Older widows are much more likely to live in poverty than older married women. Because most aged widows receive Social Security benefits, one option for increasing widows' income would be to increase their Social Security benefits. The 1994-96 Advisory Council on Social Security (1996) proposed an increase in survivor benefits with some financing from reducing spouse benefits. This proposal would address both equity and adequacy issues connected with widow benefits.[22] Auxiliary benefits create inequities because wives/widows with their owned earned

[20] The incomplete matching of respondents to their own administrative records could influence results if the omission is selective. Several applications mentioned in this paper do not compensate for the 10-15 percent of adults without linkage other than reweighting population totals. SSA's microsimulation of the baby boomer's retirement (called Modeling in the Near Term or MINT) statistically generates an administrative linkage using a nearest neighbor or "hot-deck" linkage to a similar SIPP respondent. Analysts of beneficiary children use survey data when linkages are not available because of the lower match rate.

[21] SSA records only identify couples when a spouse/survivor is drawing benefits based on their current or former spouse's earnings record. In 2003, this includes about two-thirds of aged wives and most aged widows. No marital link is possible for those without benefits or those with only their own earned benefits.

[22] The increase in widow benefits would provide more adequate retirement income to qualifying widows, primarily survivors of couples with a working wife as well as a working husband. The spouse benefit reductions would affect couples with a non-working wife or a wife with much lower earnings than her

benefits often do not receive higher benefits than if they had not worked (Iams and Sandell 1998).

An analysis of the impact of such a change needs SIPP matched data to make the estimates. The analysis requires separate measures of each spouse's earned and auxiliary benefits that must be derived from SSA records. The estimate also requires the offset of the auxiliary benefit by the earned benefit for dually entitled beneficiaries (approximately one-third of beneficiary wives and two-thirds of widows). Most dually entitled beneficiaries would not know this information and could not report it in a survey. SIPP provides information on family income, poverty, and links husbands and wives, which is absent from SSA records.

Based on analysis of SIPP matched data for married couples, this policy shift would moderately decrease poverty rates among older women by reducing the poverty rate of widows slightly more than increases in the rate for couples (Iams and Sandell 1998, Table 2; Sandell and Iams 1997).

Childcare

Advocates have argued that periods of full-time child care reduce women's Social Security benefits, but perhaps more importantly, they argue that this has had a greater impact on minority and lower income women because they have more children. The legislative proposals in the 1980s would delete a few years with no earnings (called dropout years) because of full-time child care from the Social Security worker benefit computation, thereby increasing the lifetime average earnings and earned benefits of mothers.

The analysis to test the effect of providing additional dropout years for childcare could not be made without SIPP matched data. SIPP's fertility topical module identifies the birth years of children. But the policy test requires identifying the years with no earnings, which is not available from the SIPP. The SSA administrative data provide each year's earnings taxed for Social Security purposes. In addition, the policy analysis requires identification of women expected to receive only their own retired worker benefits, because changes in a woman's earned benefits have no impact on income if she receives higher benefits as a wife or widow. This requires SSA matched earnings records to estimate expected retired-worker and auxiliary benefits.

---

husband. Wives can receive half of their husband's benefits and widow's can receive their husband's full benefits, without paying Social Security taxes on any earnings of their own. (Divorced persons can receive these benefits if married for at least ten years.) Wives and widows can receive Social Security benefits based totally on their husband's earnings, based totally on their own earnings, or based on a combination (termed dual entitlement where earned benefits offset higher auxiliary benefits). About two-thirds of wives and the majority of widows receive their benefits based on their husband's benefit either as auxiliary or dual benefits. This creates an inequity between couples and survivors of couples with a working spouse and those without a working spouse. Those with a working spouse receive lower benefits than those without a working spouse given a similar level of total couple earnings over a lifetime. Changes that increase equity either reduce benefits of couples or survivors of couples with nonworking spouses or increase the benefits of couples or survivors of couples with working spouses. Lowering Social Security benefits may reduce the adequacy of retirement income. Thus, options to increase equity often reduce the adequacy of benefits.

Using SIPP matched data, Iams and Sandell (1994) estimated the impact of childcare dropout years on benefits expected for women born in the 1930s and 1940s. They found that childcare dropout years would increase the retirement benefits of some women, but the estimated benefit increases were small, were more likely for more privileged socioeconomic groups, and were lower among women born in the early baby boom than those born in the depression (Iams and Sandell 1994, Table 3 and Table 6). Iams and Sandell conclude that subsidizing child-care dropout years is not a well targeted policy, and the impact will decline over time as fewer women drop out of the labor force to care for young children.

Retirement Earnings Test

What is the impact of eliminating the retirement earnings test (RET) which reduces Social Security benefits of working beneficiaries with earnings above specified levels? SSA wanted to estimate the impact of legislation passed in 2000 that eliminated the earnings test for working beneficiaries aged 65-69. Although most agree the financial incentives of the RET affect earnings behavior, the size of the impact has been ambiguous for high and low earners.

The SIPP matched data provided the information needed for a study of the effects of the legislative change. SSA benefit records identify the benefits in each month of each year, and SSA earnings records contain annual Social Security taxable earnings. The SIPP data provide personal characteristics such as gender, educational attainment, health limits, per capita family income, and self-employment that would indicate differential effects on various groups of beneficiaries.

The study looked at changes in earning or not earning income, earnings levels, and applications for benefits. Removal of the earnings test in 2000 was not related significantly to changes in the presence of earnings (Song 2002). This suggests that it didn't change the decision to work or not work among beneficiaries aged 65-69. The earnings test removal significantly increased the earnings of high earners but not middle and low earners (Song 2002, Table 9). The removal also was associated with slightly increased applications for benefits among persons aged 65-69.

III. Micro-Simulations

SSA conducts policy evaluations with micro-simulation models created from SIPP matched to SSA administrative records. This paper discusses two models—Modeling Income in the Near Term (MINT) which projects life histories of the aged population 20 years from now in 2022 and a Financial Eligibility model for Supplemental Security Income and other means-tested programs.

Financial Eligibility Model

Policy analysis related to the Supplemental Security Income (SSI) program requires SIPP matched data. SSI pays benefits to the aged and nonaged disabled with limited income and limited assets which SIPP identifies.[23] The SSA administrative data are used to clarifying benefit

[23] SSI also requires the nonaged to have disabling health limitations which can be inferred from SIPP information.

eligibility status and actual benefits received from Social Security and the SSI program (Huynh et. al. 2001, Table 1 and Table 2).

SSA has developed a Financial Eligibility model that can be used to address a wide range of policy issues related to SSI, Social Security, Medicare, Medicaid and other programs. These include the following:

- What is the rate of participation in SSI and other means-tested programs? Is there a substantial pool of eligibles that do not participate in the program? Why? Davies et. al. (2002) find that about three-fifths of eligibles participate in SSI. Rupp and Sears (2000) and Sears (2002, Table 1) also find about three-fifths of eligibles participate in Qualified Medicare Beneficiary, Special Low-Income Medicare Beneficiary, and Qualified Individual buy-in programs which pays Part-B Medicare premiums with Medicaid funds.

- What are the costs and benefits of potential modifications of SSI program rules? The model can provide estimates on changes in program cost, number of eligibles, number of participants, average benefits, and distributional outcomes such as effects on the poverty rates and the poverty gap. The model is capable of estimating the potential effects of changes to the SSI program, such as the asset test, earned and unearned income disregards. For example, if SSI expenditures increased by 3 percent through changes in the Social Security benefit exclusion, then the poverty gap of aged women would decrease 1.1% (Rupp et. al. 2001, Table 3).

- If policy makers consider a range of alternative interventions, which one is the most effective? SSA has developed a methodology of cost-equivalent comparisons that can be used to assess which one of several policy alternatives are most effective in improving desired outcomes at given levels of funding availability. For example, Rupp et. al. (2001, Table 3) find that modifying the SSI asset limits is a relatively effective change in reducing poverty among elderly women.

- What are the interactions between SSI program changes and other programs? What is the effect of proposed changes in other programs, such as Social Security on SSI participation and cost? For example, how do proposals to introduce a minimum Social Security benefit affect SSI? What changes in SSI are necessary to facilitate desired distributional outcomes under a Social Security minimum benefit? What is the effect of changes in SSI eligibility rules on Medicaid participation and cost?

- What is the likely size of the SSI program in terms of costs and participation in the medium term? How do different demographic and socioeconomic factors, as well as potential policy changes affect this? For example, what is the likely effect of the increased proportion of successive cohorts with Social Security insured status and the aging of the baby boom generation on SSI participation and program cost?

SSA continues to develop and improve the Financial Eligibility Model to accomplish these objectives with the most recent SIPP data on income and assets matched to SSA records.

MINT

The Modeling Income in the Near Term (MINT) microsimulation model is designed to study the retirement of the baby boom birth cohort as well as the World War II and Depression birth cohorts. Policy makers have a strong interest in the differential effects of policy changes on the benefits, total income, and poverty level of the retiree population, as well as its subgroups. Of particular concern to policy makers is the economic well-being of future retirees in the baby boom cohort – those born between 1946 and 1964. Not only is this the largest birth cohort in U.S. history, but the earliest baby boomers will be eligible for retirement in 2008, and without program changes the Social Security (OASDI) Trust Fund is projected to be exhausted in 2041 (The Board of Trustees Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds, 2002). Aside from its sheer size, the baby boom cohort has distinguished itself from earlier cohorts in a number of ways that reflect the culture of the post-world War II period. The baby boom cohort experienced "unprecedented prosperity" and increased educational opportunities and attainment, as well as major changes in marital patterns and in the lifetime employment and earnings of women (Farley 1996; Levy 1998; O'Rand and Henretta 1999). Because of structural changes in mortality, marriage, lifetime earnings, and work patterns, we would expect the impact of policy changes to differ between current retirees and future retirees in the baby boom cohort.

When changes occur across time, policy analysis of the current beneficiary population may be misleading. Analysis of the future population targeted by legislation is preferable. This approach takes into account birth cohort differences and diversity and, consequently, is sensitive to shifts across cohorts in socio-economic relationship such as in women's lifetime earnings and work patterns. Accordingly, Modeling Income in the Near Term (MINT) projects the life histories of the baby boom cohort and the aged population to 2022.[24] SSA can estimate the impact of alternative Social Security policies on total income and poverty for subgroups defined by race, educational level, and marital status of the baby boom cohort in retirement.

The MINT projection of life histories relies heavily on the SIPP matched data. To enhance the data for analysis, MINT combines the SIPP panels of 1990, 1991, 1992, and 1993. The policy universe for most analyses is the surviving population born from 1931 through 1960 that is expected to reach retirement age and to receive Social Security retirement and survivor benefits in 2022. [25] The matched data provide important information that supplements the SIPP reported data. Statistical projections make use of these longitudinal SSA data to estimate life histories until death. SSA administrative records measure the annual earnings history, the monthly benefit history, and date of death through 1999. The MINT model makes independent statistical

[24] The U.S. Social Security Administration (SSA) created MINT with substantial input from the Brooking Institution, the RAND Corporation (Panis and Lillard 1999), and the Urban Institute (Toder et. al. 1999; Toder et. al. 2002). For a summary of the work completed by the Brookings Institution, RAND, and the Urban Institute for the initial MINT model see Butrica, et. al. 2001. Toder et. al. (2002) document the revision of MINT completed in 2002.

[25] Those born 1961-64 were dropped from the analysis because with fewer years of real data we are less confident in their projections of retirement income. The SIPP reported data for a person born in 1960 would be at age 30 in the 1990 panel, 31 in the 1991 panel, 32 in the 1992 panel, and 33 in the 1993 panel.

projections until death for each SIPP respondent's lifetime earnings, retirement income (Social Security benefits, pensions, assets, and earnings of working beneficiaries), and marital changes. The 1990-1993 panels of SIPP for middle aged persons born in 1931-1960 directly measures such choices as educational attainment, marriage and divorce history, current employment, pension plan participation, and savings.

MINT projects substantial changes in the characteristics of the baby boom retirees compared to earlier birth cohorts from World War II and the depression. Butrica and Iams (1999, Table 2) document with MINT the importance of both marital histories and earnings records to the projected Social Security benefits of married couples. MINT projects that spouse and widow benefits will be less important to the baby boom cohort than to earlier cohorts born in the depression and World War II (Butrica, Iams and Sandell 1999, Chart 2). MINT also projects that the proportion of women who divorce will be higher among the baby boom cohorts than earlier cohorts, but the proportion of these women eligible for benefits as a divorced spouse will decline (Butrica and Iams 2000, Table 3 and Chart 2). This occurs because MINT projects divorced women in the baby boom to be more likely to have their own earned retired-worker benefits, but they are less likely to have at least ten years of marriage necessary to be eligible for spouse/widow Social Security benefits.

Using the MINT data system, Toder et. al. (2002, Chapter 9) describe the characteristics of the aged population in 2020 and the retirement population at age 62 and age 67. These tables describe the projected change in socio-economic and demographic characteristics among the baby boom compared to earlier cohorts born in the 1930s and during World War II. MINT projects the baby boom cohort of beneficiaries at age 62 and age 67 to be more educated, to contain more minorities, and to contain fewer married couples than earlier cohorts. MINT projects retirement wealth among the baby boom to increase with shifts toward more income from pensions as well as non-financial wealth. MINT projects average levels of retirement income at age 67 to be higher in the early baby boom cohort than the depression cohort, but similar to the late baby boom cohort.

IV. Descriptions of Beneficiaries

SSA also produces several reports of the socioeconomic and demographic background characteristics of its current beneficiaries using SIPP matched records. These reports describe the characteristics of beneficiaries served by SSA and the importance of SSA administered benefits as an income source. The SSA record match identifies the SSA program beneficiaries and benefit amounts actually paid to beneficiaries (Huynh et. al., 2001). The SIPP based characteristics are unavailable from SSA records used in administering its programs. Tabulations include SIPP based demographic characteristics, sources of income, family income, poverty level, family and household size, household type, home ownership and receipt of assistance for energy, for housing, for Food Stamps, for health insurance. [26] The Performance and Accountability Report

[26] For example, the SSI Annual Statistical Report (2001d) reports characteristics of Supplemental Security Income Title XVI recipients, and the Annual Statistical Report on the Social Security Disability Insurance Program (2001a) reports characteristics of Disabled Insurance Title II

(Social Security Administration 2001c) contains measures of adequacy of income of beneficiaries including the reduction in the poverty gap due to SSI benefits, SSI as a percent of total income, and the percent participating in an employer sponsored pension plan.

Conclusion

SIPP data linked with SSA administrative data benefit from the strengths of surveys and administrative data. The linked data have become a critical source of information for policy analysis, evaluation of legislation, and statistics to inform policymakers.

Bibliography

The Board of Trustees, Federal Old-Age and Survivors Insurance and Disability Trust Funds. 2002. THE 2002 ANNUAL REPORT OF THE BOARD OF TRUSTEES OF THE FEDERAL OLD-AGE AND SURVIVORS INSURANCE AND DISABILITY TRUST FUNDS. Washington, D.C.: U.S. Government Printing Office.

Butrica, Barbara and Howard M. Iams. 1999. "Projecting Retirement Income of Future Retirees with Panel Data: Results from the Modeling Income in the Near Term (MINT) Project. Social Security Bulletin. Vol. 62, No. 4, pp. 3-8.

Butrica, Barbara and Howard M. Iams. 2000. "Divorced Women at Retirement: Projections of Economic Well-Being in the Near Future". Social Security Bulletin. Vol. 63, No. 3, pp. 10-24.

1994-1996 Advisory Council on Social Security. 1996. Reports of the Technical Panel on Assumptions and Methods Technical Panel on Trends and Issues in Retirement Savings and Presentations to the Council. Vol. II. Washington, D.C.: Social Security Administration.

Butrica, Barbara, Howard M. Iams, and Steven H. Sandell. 1999. "Using Data for Couples to Project the Distributional Effects of Changes in Social Security Policy". Social Security Bulletin. Vol. 62, No. 3, pp. 20-27.

Butrica, Barbara, Howard M. Iams, James H. Moore, and Mikki D. Waid. 2001. Methods in Modeling Income in the Near Term (MINT I). ORES Working Paper Series No. 91. Washington, D.C.: Office of Research, Evaluation and Statitistics.

Davies, Paul, Minh Huynh, Chad Newcomb, Paul O'Leary, Kalman Rupp, and Jim Sears. 2002. "Modeling SSI financial eligibility and simulating the effect of policy options". Social Security Bulletin. Forthcoming.

Farley, Reynolds.1996 THE NEW AMERICAN REALITY: WHO WE ARE, HOW WE GOT THERE, WHERE WE ARE GOING. New York: Russell Sage Foundation.

Huynh, Minh, Kalman Rupp, and James Sears. 2001. The assessment of Survey of Income and Program Participation (SIPP) benefit data using longitudinal administrative records. Paper presented at the Federal Statistical Committee on Methodology meeting. SIPP Working Paper Series No. 238. Washington, D.C.: Census Bureau.

---

beneficiaries. SSA also includes SIPP information in the Annual Report to the President and Congress on the Supplemental Security Income Program (Social Security Administration 2001b)

Iams, Howard M. and Steven H. Sandell. 1994. "Changing Social Security Benefits to Reflect Child-Care Years: A Policy Proposal Whose Time Has Passed". Social Security Bulletin. Vol. 57, No. 4 (Winter), pp. 10-24.

Iams, Howard M. and Steven H. Sandell. 1998. "Cost-Neutral Policies to Increase Social Security Benefits for Widows: A Simulation for 1992". Social Security Bulletin. Vol. 61. No. 1, pp. 34-43.

Levy, Frank. 1996. The New Dollars and Dreams: American Incomes and Economic Change. New York: Russell Sage Foundation.

National Research Council. 1997. Assessing Policies for Retirement Income: Needs for Data, Research, and Models. Edited by Constance F. Citro and Eric A. Hanushek. Washington, D.C: National Academy Press.

O'Rand, Angela M. and John C. Henretta. 1999. *Age and Inequality: Diverse Pathways through Later Life.* Boulder, Co: Westview Press.

Panis, Constantijn, Roald Euller, Cynthia Grant, Melissa Bradley, Christine Peterson, Randall Hischer, and Paul Steinberg. 2000. SSA Program Data User's Manual. June. Santa Monica, California: RAND.

Panis, Constantijn and Lee Lillard. 1999. Near Term Model Development Part II. Final Report. Santa Monica, Ca: RAND.

President's Commission to Strengthen Social Security. 2002. Report of the President's Commission to Strengthen Social Security. Washington, D.C.

Rupp, Kalman and James Sears. 2000. "Eligibility for the Medicare Buy-in Programs, Based on a Survey of Income and Program Participation Simulation". Social Security Bulletin. Vol. 63, No.3. pp. 13-25.

Rupp, Kalman, Alexi Strand, and Paul S. Davies. 2001. "The Potential of the Supplemental Security Income Program to Fight Poverty Among Elderly Women." Proceedings of the Institute for Women's Policy Research 2001 Conference, The Status of Women: Facing the Facts, Forging the Future.

Sandell, Steven H. and Howard M. Iams. 1997. "Reducing Women's Poverty by Shifting Social Security Benefits from Retired Couples to Widows". Journal of Policy Analysis and Management. Vol 16, No. 2 (Spring), pp. 279-297.

Sears, James. 2002. " 1996 QMB/SLMB". Unpublished paper. Washington, D.C.: Social Security, Office of Policy.

Social Security Administration. 2001a. Annual Statistical Report on the Social Security Disability Insurance Program. Washington, D.C.: Social Security Administration, Office of Policy, Office of Research Evaluation and Statistics.

Social Security Administration. 2001b. Annual Report to the President and Congress on the Supplemental Security Income Program. Baltimore, Md: Social Security Administration, Office of the Chief Actuary.

Social Security Administration. 2001c. Performance and Accountability Report Program. Baltimore, Md: Social Security Administration, Office of Finance, Assessment, and Management.

Social Security Administration. 2001d. SSI Annual Statistical Report. Washington, D.C.: Social Security Administration, Office of Policy, Office of Research Evaluation and Statistics.

Toder, Eric, Cori Ucello, John O'Hare, Melissa Favreault, Caroline Ratcliffe, Karen Smith, Gary Burtless, and Barry Bosworth. 1999. Modeling Income in the Near Term- Projections of

Retirement Income Through 2020 for the 1931-1960 Birth Cohorts. Final Report. Washington, D.C.: The Urban Institute.

Toder, Eric, Lawrence Thompson, Melissa Favrealt, Richard Johnson, Kevin Perese, Caroline Ratcliffe, Karen Smith, Cori Uccello, Timothy Waidmann, Jillian Berk, Romina Woldmariam, Gary Burtless, Claudia Sahm, and Douglas Wolf. 2002. Modeling Income in the Near Term: Revised Projections of Retirement Income Through 2020 for the 1931-1960 Birth Cohorts. Final Report. Washington, D.C.: The Urban Institute.

U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau. 2001. SIPP Survey of Income and Program Participation User's Guide. Washington, D.C.

Discussing Potok and White's Papers Presented in Session 7: Stewardship of Linked Survey and Administrative Data

Olivia Blum
Israel Central Bureau of Statistics

Potok and White scan policies and restrictions that statistical agencies put upon them. The deriving question is why. Why do we go through Acts of - privacy, confidentiality, asking for consent, access to information etc.? Why a secondary use of administrative data accelerates the need to refer to these deeds and acts?

First, these are all implementation symbols of social norms and values. This is the social glue that we use and create day by day. Theses acts come to corroborate the social contract within the national group. The second function is ensuring the survival of the statistics organization through legal agreements with the direct users and with the public.

The statistics bureau is responsible for providing the users with quality data under changing circumstances. No bureau can allow itself stagnation with regard to attitude/policy, ways of action and tools. People, social structure and processes, technology, they all change and therefore social values and behavior. Adaptation to theses changes serves the quality data objective in the long run, meaning that it is required to maintain the functionality and therefore the mere existence of the organization.

As for the public, the agreement with the public has a give and take pattern. The data collector asks for private information and gives in return analyzed information that allows the policy makers to act efficiently and effectively and allows the individual to choose, based on empirical findings, what to eat, where to live etc. Private data and the consent to link individual administrative files are given in return to meta information and aggregated results and in return to explicit way of handling these data: Storage in secured sites, no accessibility to people and uses not specified, etc. Breaking this contract means no data, no quality data, no consent to link records, no support in a changing reality and changing environment, and therefore, no justification to the existence of the organization.

The acts and the careful processes of handling linked data, as described in both papers, bring about additional costs since the pure professional considerations are not the only guiding lights when coming to link administrative records. The resultant questions aim toward the quality data market in a broader perspective: Who are the participants in this market? What are the mechanisms to make it stable? What does the statistics agencies have an influence on?

There are three core participants: The direct data supplier, whom the data describe, the data user and the statistics agency who demands the data. The challenge derives from the statistics agency's role to serve and protect both, the supplier and the user.
As for the operating forces and mechanisms, I would like to make an analogy to the economic commodities market, which has two basic parameters, quantity and price. In the quality data market the detailed data represents the quantity while the quality represents the price. In this market the data supplier prefers to be less exposed while the data demander is seeking more

detailed data (see supply and demand curves in diagram1). The statistics agency and the user consider rich data, obtained by linking records, as quality data. However, the first is obliged to protect the privacy of the data suppliers whilst the last does not. The equilibrium point in this market is not stable; everyone wants to get out of it.

There are several mechanisms to be engaged in stabilizing the equilibrium point:

1. Acts, controls, policies, practices, as described in Potok and White's papers. These restrictions move up the whole supply curve, i.e., the public is willing to allow the statistics agency to link records and to have more detailed information for the same price in quality terms (see diagram2).

2. Pushing the suppliers up along the supply curve, which means reducing antagonism by overt presentation of the benefits drawn from rich linked data and by encouraging and enabling the public to use statistics on a daily basis.

3. Partnerships or business relations with the suppliers of the administrative records, in the public and government sectors, in the private and business sectors. This is a mechanism that comes to ensure the obtaining of the administrative data. It is a prerequisite to the existence and stability of this market.

4. Pulling the demand curve of the statistics agencies, vertically, toward a less invasion of privacy with no quality loss. Meaning, reducing the correlation between quality and quantity, which can be done by developing methodologies that enable the statistical estimates to rely on less detailed information (see diagram3).

5. Reducing antagonism of the public by visible fairness of the redistribution center. The state administration serves as a redistribution center of the national resources. Although it seems to be irrelevant to the statistics world and statistics agencies have no control over it, the conduct of public administration with regard to benefits, subsidiaries, infrastructure investments etc. has a direct influence on the cost/benefit analysis of the individual when asked to supply data or to give consent to use linked data.

Potok and White focus on the first mechanism as activated in their statistics agencies (US Census Bureau and Statistics Canada, respectively). It is the one that statistics agencies have a more influence on. This mechanism stipulates the approval of record linkage not only vis a vis the public, but also within the statistics bureau and vis a vis the relevant government oversee functions. However it is not clear how far the supply curve can be pushed and when additional restriction costs more than its contribution.

The second mechanism, in which the empirical findings are either published or made accessible to the public, is presented as an integral part of the record linkage program in White's paper. Data suppliers, whether they are individuals or administrative data holders (third mechanism), should have an ongoing interest to supply the data and to allow its use.

The forth Mechanism is an ongoing challenge for today's statisticians while the feasibility of the implementation of the fifth one is unclear.

Diagram 1

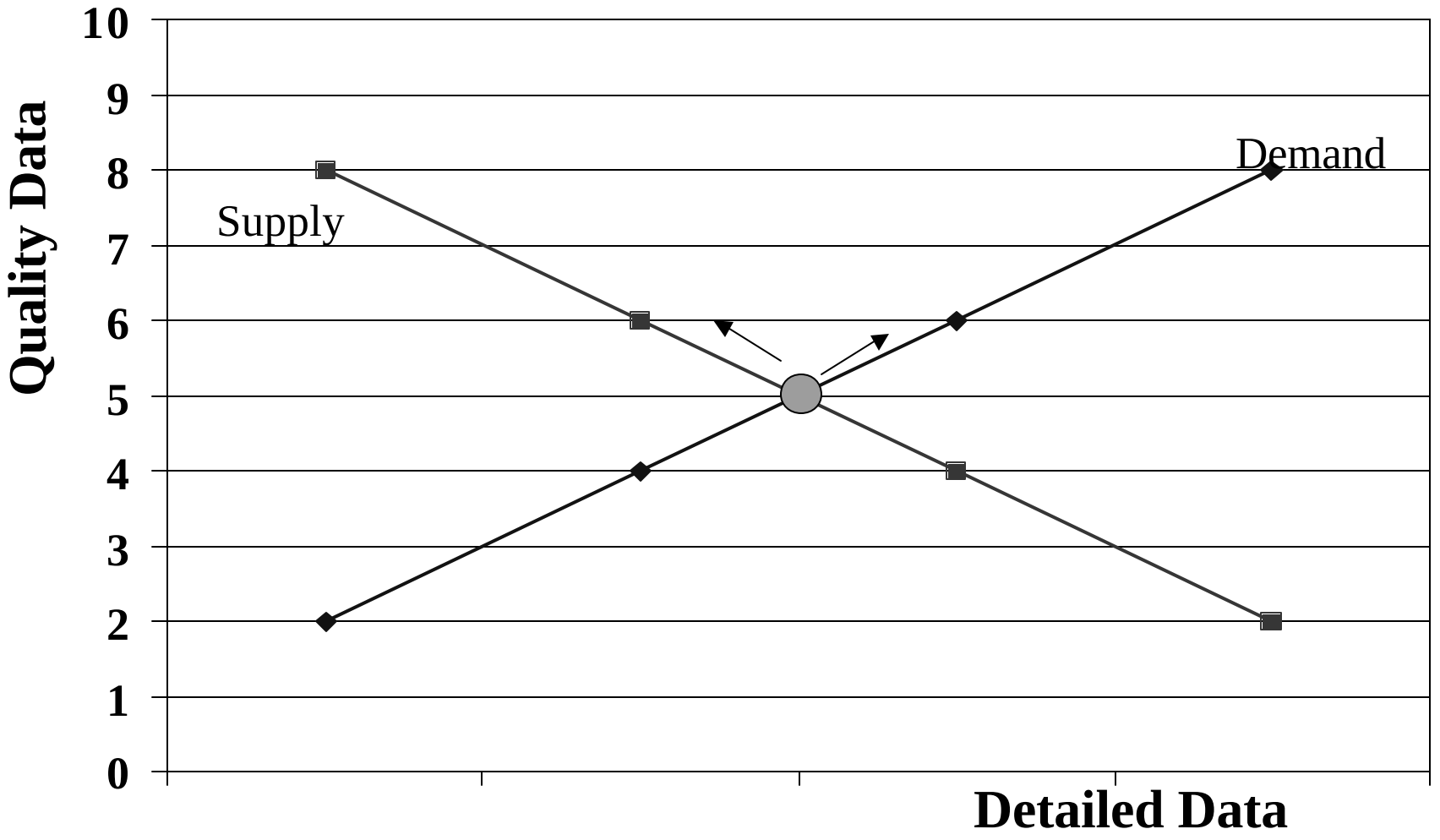


Diagram 2

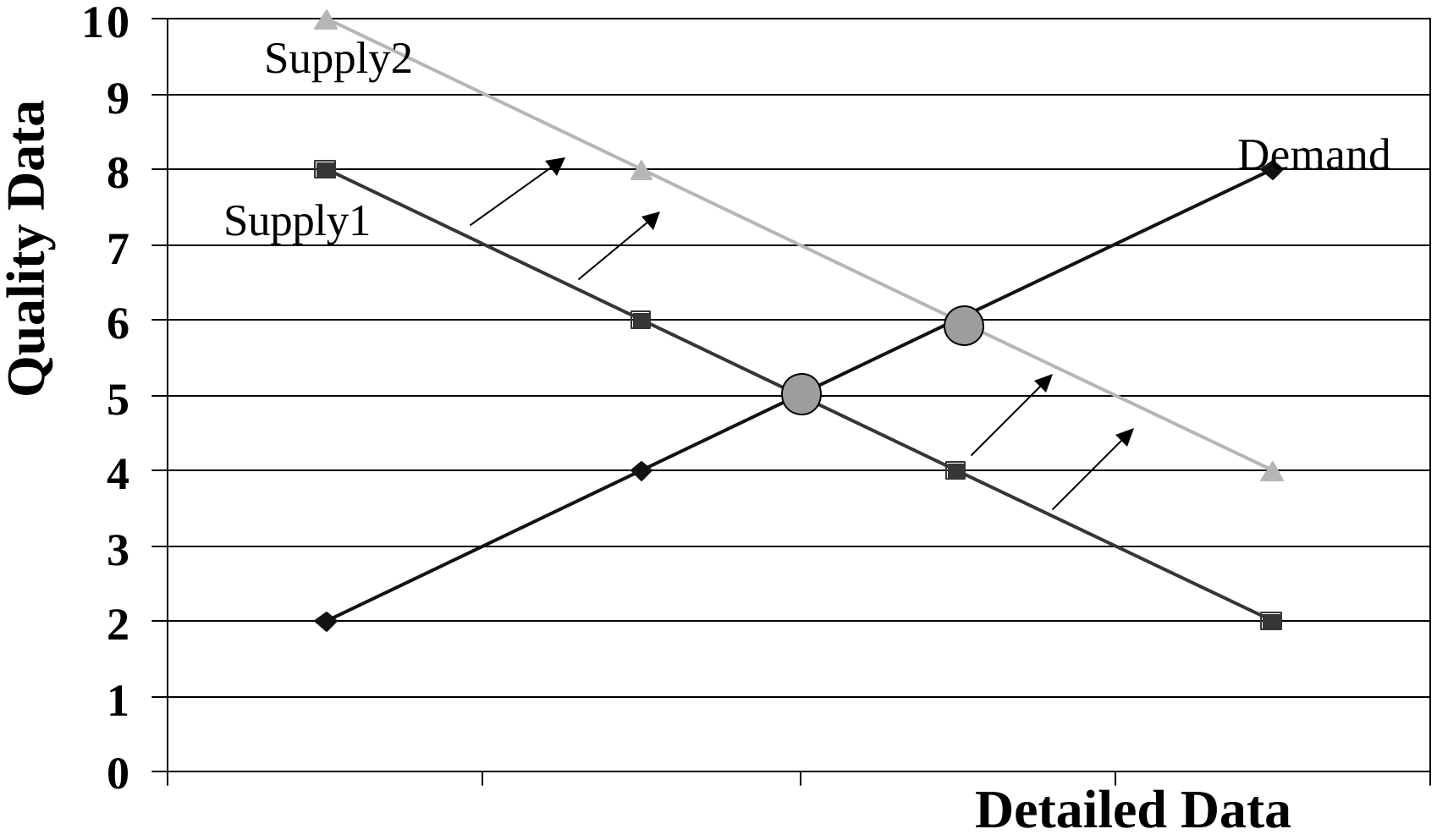
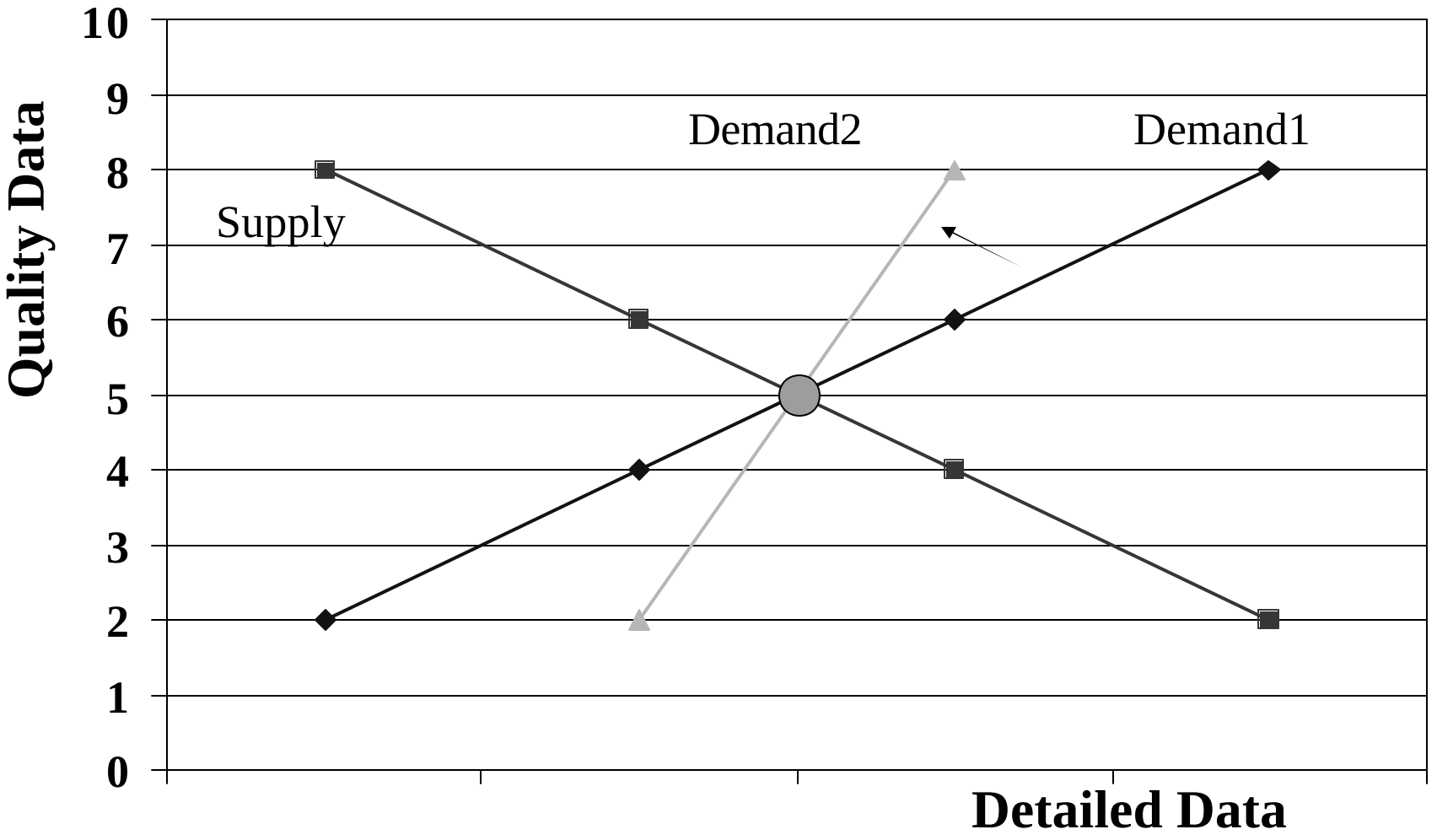

Quality Data
10
9
8
7
6
5
4
3
2
1
0
Supply2
Supply1
Demand
Detailed Data

Diagram 3


Quality Data
10
9
8
7
6
5
4
3
2
1
0
Demand2
Demand1
Supply
Detailed Data

Session 8

Capitalizing on Technology to Enhance Survey Reporting

A Comparison of the Random Digit Dialing Telephone Survey Methodology with Internet Survey Methodology as Implemented by Knowledge Networks and Harris Interactive

Jon A. Krosnick and LinChiat Chang
Ohio State University

Introduction

With their response rates declining and costs rising, telephone surveys are increasingly difficult to conduct. At the same time, Internet data collection is emerging as a viable alternative, in two forms. Some firms are distributing computer equipment to national samples recruited through RDD calling, and other firms are attracting volunteer respondents and then building panels of those individuals with some demographic characteristics distributed as they are in the nation. Most firms assemble panels of respondents who provide data on a regular basis.

Just as the survey industry was initially reluctant to embrace the telephone when it emerged decades ago as an alternative to face-to-face interviewing in respondents' homes, the field is currently uncertain about the costs and benefits of a shift to Internet-based data collection. The practical advantages of this approach are obvious: quick turn-around time, easy presentation of complex visual and audio materials to respondents, consistent delivery of questions to and collection of responses from respondents, the flexibility to allow respondents to complete questionnaires whenever they like, lack of the pressure to move quickly that is typical of telephone interviews, and the ability to track a respondent's answers across repeated waves of questioning. But potential drawbacks are obvious as well: literacy ability to read questions and navigate web pages is required, as is proficiency with a computer keyboard (and mouse when one is used); the lack of interviewers' modeling of professionalism and commitment to the task may compromise respondent attentiveness and motivation; lack of ability for an interactive conversation between a respondent and an interviewer may preclude clarifying the meanings of ambiguous questions; samples may be of uncertain representativeness, and more. Some of these potential drawbacks are overcome by internet data collection via devices other than computers (e.g., WebTV), but most remain.

Given the obvious practical advantages of Internet-based data collection, it seems worthwhile to conduct object tests of this relatively new method in direct comparison with the dominant alternative methodology: telephone interviewing. To do so, we commissioned a set of side-by-side surveys using a single questionnaire to gauge public opinion and voting intentions regarding the 2000 U.S. Presidential Election from national samples of American adults.

Data were collected by three houses: The Ohio State University Center for Survey Research (CSR), Knowledge Networks (KN), and Harris Interactive (HI). The CSR did RDD telephone interviewing. KN recruited respondents via RDD telephone interviews and equipped them with WebTV, which then permitted Internet data collection. HI respondents joined a panel after seeing and responding to invitations to participate in regular surveys; the invitation appeared on the Excite search engine web page and in various other places as well. These respondents also completed Internet surveys.

This report describes just a few of the preliminary results from our investigation. We have conducted extensive analyses of the obtained data and have much more to do analytically. The findings reported here capture a few of the general patterns we see in the data, and we look forward to providing much more extensive and detailed reports of our findings in the near future.

We compared the data from these various surveys in a number of ways:

1. We compared the demographic characteristics of the three samples to the demographic characteristics of the nation as a whole (assessed by the U.S. Census Bureau's March 2000 CPS Supplement).

2. We compared the distributions of responses to opinion and behavior questions across the three houses, expecting one of two possible patterns to be observed. If respondents answer less carefully on the Internet because of the lack of an interviewer to motivate and assist them, we thought respondents might select midpoints on rating scales more often than did telephone respondents (posited to be a form of survey satisficing; Krosnick, 1991). But if Internet respondents answer more carefully because they feel less rushed than telephone respondents do, Internet respondents might select midpoints of rating scales less often than telephone respondents. We also thought that because HI respondents were purely volunteers, their motivation to provide accurate data and therefore their response quality might exceed that of the other houses.

3. We evaluated the reliability of individual questions. If Internet respondents answer less precisely, we would expect to see higher reliability from the telephone respondents. The reverse pattern of reliabilities would indicate greater care in responding by the Internet respondents. And again, the HI respondents might have provided more reliable responses because they were volunteers.

4. We investigated the extent to which respondents manifested another form of survey satisficing: non-differentiation (i.e., identically answering a series of questions using a single rating scale). We thought this response pattern could be greater or could be less among the telephone respondents as compared to the Internet respondents, depending upon whether the Internet mode inspires more or less satisficing. If HI respondents' motivation was highest, they might have manifested the least non-differentiation.

5. Finally, we gauged the quality of responses by assessing predictive validity; stronger statistical relations between variables that theory says should be related to one another is generally taken to indicate greater respondent precision in providing the self-reports. Again, we expected that predictive validity could be either greater among the telephone respondents or less among those respondents as compared to the Internet respondents. And if HI respondents were most motivated, their predictive validity might have exceeded that of the other houses.

Data Collection

Data were collected by all three houses in two waves. The first wave of data collection was conducted before the election campaign began, in June and July. Then shortly after election day, respondents again answered questions. During the pre-election wave, respondents predicted their presidential vote and reported a wide range of attitudes and beliefs that are thought to drive vote choices. During the post-election wave, respondents reported whether they had voted and for whom they had voted.

Approximately 1,500 respondents were interviewed pre-election by telephone by the CSR. Approximately 5,000 respondents provided data to KN pre-election, and approximately 2,300 respondents provided data to HI pre-election. The CSR and HI data collections involved administering each questionnaire entirely, which lasted about 30 minutes on the telephone pre-election. KN broke the questionnaire up into three parts and administered one part per week for two consecutive weeks, took one week off, and administered the final part the next week.

Details on response rates and field periods are provided in Table 1. The pre-election response rate is highest for CSR and lower for KN. The rate at which people invited by HI to complete the pre-election survey did so is lower than the response rates for either CSR or KN. Similarly, about four-fifths of CSR and KN respondents who provided data pre-election also did so post-election, whereas this figure was 45% for HI.

Our comparisons across houses were done after weighting the samples. The weights applied to the KN and HI data were provided to us by those houses, and we generated the weights applied to the telephone data using CSR's standard procedure.

Demographic Representativeness

Table 2 shows the demographic characteristics of respondents in the CSR, KN, and HI surveys, when samples were not weighted, as well as CPS data for comparison. Under each column of percentages for a demographic variable is the average deviation of the results from the CPS figures.

In general, the average deviations are generally not huge, and sample representativeness is never dramatically poor in terms of the percentage point deviation of any survey estimate from the population. The two largest percentage point discrepancies appear between the HI and CPS percentages for people who graduated from high school and got no more education (deviation = 21 percentage points) and individuals with incomes less than $25,000 (deviation = 17.9 percentage points). Most discrepancies are much smaller than these in terms of percentage points.

The telephone survey sample manifests the smallest average deviation for three variables (education, income, and age). For two other variables (race and gender), the KN sample is more similar to the population than is either the telephone survey sample or the HI survey sample. The HI sample consistently manifests the largest average deviations from the population. As shown in the bottom row of the table, the average deviation for the telephone sample is 4.0%, 4.3% for KN, and 8.7% for HI.

Consistent with other previous studies, the telephone sample under-represents the least educated individuals and over-represents the most educated individuals. The same bias is apparent in the KN sample and even more apparent in the HI sample. Likewise, the telephone sample under-represents the lowest income individuals and over-represents higher income individuals; this bias is again more strongly apparent in the KN sample and even more apparent in the HI sample.

Again consistent with prior work, the telephone sample under-represents the youngest and oldest individuals, and these same biases are even more apparent in the KN and HI samples.

Telephone samples typically under-represent African-American respondents, and this was true here for the CSR sample, and the KN and HI samples evidenced this same bias even more strongly.

Finally, the telephone sample over-represented women, whereas the HI sample over-represented men; the KN sample's gender balance closely matched the population.

One way to summarize the discrepancies between houses is to correlate the figures in each of the first three columns of numbers in Table 2 with the numbers for the CPS in the last column. These correlations are .96 and .94 for CSR and KN, respectively, and .87 for HI. This approach again indicates nearly comparable representativeness for the CSR and KN data and less representativeness for the HI data.

Table 3 shows the distributions of the demographics after the weights have been applied to the data. As shown in the last row of the table, weighting considerably shrank the demographic deviations from the population (as should occur, of course), making the houses equivalently accurate.

Distributions of Responses

Next, we turn to examining some substantive responses to the survey questions.

Turnout. Table 4 presents post-election reports of turnout. With more than 70% of CSR and KN respondents and more than 90% of HI respondents reporting that they voted in 2000, these surveys manifest the same bias that all post-election surveys do. This may be due to self-selection: people especially interested in politics may have been especially likely to choose to participate in surveys about politics. The HI respondents also manifested the most frequent reports of having usually voted in past elections, suggesting that this sample was the most politically involved, whereas the rates for CSR and KN were quite similar.

Candidate Preference. Voters' reported choices of Presidential candidates differed between houses (see Table 4). Majorities of CSR and KN voters said they voted for Al Gore, whereas a majority of the HI voters said they voted for George W. Bush. Among non-voters, a clear plurality preferred Al Gore. Again, the CSR and KN results were quite comparable, whereas the HI non-voters manifested a more pronounced preference for candidates other than Gore and Bush.

Party Identification. The distribution of party identification confirmed two of the trends we have seen thus far (see Table 5). First, the CSR and KN data are quite similar, and the HI data are more different. Second, the HI respondents were less likely than the CSR and KN respondents to be Independents who do not lean toward either party, and the HI respondents were most likely to report strong party identification, which is again consistent with the idea that the HI respondents were the most politically involved.

Knowledge About Politics. Our pre-election questionnaire included a 5-item quiz of respondents' factual knowledge about politics, and Table 6 shows that the Internet respondents were more knowledgeable than were the telephone respondents. The average percent of questions answered correctly was 53% for CSR, 62% for KN, and 77% for HI, again suggesting the highest political involvement for the latter sample.

Other Attitudes and Beliefs. On most other measures of attitudes and beliefs, HI respondents chose the extreme ends of rating scales more often than the other respondents, while CSR respondents tended to choose the mid-points of rating scales most often. One example is displayed in Table 7, which shows the distributions of thermometer ratings of attitudes toward President Bill Clinton, Al Gore, and George W. Bush (0= least positive, 50=neutral, and 100=most positive).

Measurement Reliability

We were able to estimate the reliabilities of the measures by building a structural equation model involving two indicators of candidate preferences gathered at both waves: reported vote choice (predicted at pre-election and actual post-election) and the difference between thermometer ratings of Gore and Bush. The model posited that both measures were indicators of a latent variable (i.e., true candidate preference) at both waves, and this latent variable was allowed to manifest instability across waves. From this model, we could estimate the reliabilities of the measures (which appear in Table 8).

The CSR and KN samples yielded very comparable reliabilities, whereas the HI sample yielded notably higher reliabilities. The latter group's higher reliabilities may be attributable to more effortful reporting by those respondents and/or may be due to the HI sample containing more people who naturally answer survey questions with less random error (i.e., highly educated respondents). The structural equation modeling approach does not offer an easy way to control for demographic differences between the samples, so we cannot test these two explanations directly.

Non-Differentiation

The questionnaire included various batteries of questions using the same rating scale, and we calculated a non-differentiation score for each battery. We then standardized these scores and averaged them together to yield a single non-differentiation score for each respondent.

As shown in Table 9, the average standardized non-differentiation score was comparably high for the CSR and KN respondents and notably lower for the HI respondents. And as the regression coefficients in the first row of Table 10 show, the HI non-differentiation rate was significantly lower than those for CSR and KN, which were not significantly different from one

another. This pattern remained when we controlled for differences between houses in levels of education (see row 2 of Table 10).

As the final row of Table 10 shows, though, controlling for differences between houses in terms of political knowledge revealed significantly more non-differentiation in the KN sample than in the CSR sample ($b=.06$, $p<.05$) and the HI sample ($b=-.07$, $p<.01$). Thus, the KN respondents appeared to have satisficed most according to this measure, and the HI respondents did so the least.

Predictive Validity

Finally, we examined data quality via predictive validity. These tests are all predicated on the assumption that respondents' candidate preferences should be correlated to at least some degree with the array of variables that are thought to be determinants of vote choices. We therefore conducted binary logistic regressions predicting vote choice (coded dichotomously: Bush vs. Gore) with each of its posited predictors.

These simple logistic regressions tell a consistent story: the Internet data manifest higher predictive validity than do the telephone data across the board, often substantially so. One set of illustrations of this pattern appears in Table 11. Here, the predictors are respondents' perceptions of how national conditions would change if Bush or Gore were elected President, and the dependent variable is candidate preference. The coefficients shown in columns 2 and 3 are larger than the comparable coefficients in column 1, attesting to higher predictive validity for the Internet respondents. As the first two columns of numbers in Table 12 attest, the CSR's predictive validities are consistently significantly smaller than those of KN and HI.

Note also that the predictive validity coefficients for HI (in column 3) are consistently larger than those for KN (in column 2), suggesting that HI's volunteer respondents were more precise in their reporting. As the third column of Table 12 shows, two of these five differences are statistically significant.

These differences might be attributable to differences in sample composition. That is, the KN and HI samples were higher in education and political knowledge than the CSR sample, and the HI sample was higher in education and political knowledge than the KN sample. If education and political knowledge enhance predictive validity (which they very well might), this could be responsible for the appearance of differences between the houses.

As columns 3, 4, and 5 of Table 12 show, almost all of the differences between houses are smaller when controlling for demographics and political knowledge and for interactions of the demographics and political knowledge with attitudinal predictors than when not controlling for these variables. However, all but two of the significant differences between houses remain significant after controlling for demographics and political knowledge and interactions of them with the attitudinal predictors. Therefore, the differences between houses are only slightly attributable to sample composition differences.

Specific Conclusions

These results and many others we have obtained but not reported in this memo support a set of specific conclusions:

1) Differences between the telephone and Internet samples in terms of distributions of variables or data quality were rarely huge.

2) The CSR sample was most representative of the population; the KN sample was nearly as representative; and the HI sample was least representative.

3) The Internet samples over-represented high social status individuals more than the telephone sample did, and, relative to the CSR and KN samples, the HI sample over-represented individuals highly knowledgeable about politics, individuals highly involved in politics, and individuals who voted for George W. Bush.

4) Answers given by HI respondents contained the least random error and the least systematic error attributable to survey satisficing. Rates of random error were comparable for CSR and KN, and the CSR respondents manifested the highest rates of satisficing. The differences in systematic measurement error appeared even when controlling for differences in sample composition in terms of demographics and political knowledge.

5) Reports of attitudes collected over the Internet manifested higher predictive validity than reports of attitudes collected over the telephone, and HI respondents occasionally manifested higher predictive validity than did KN respondents. The differences in predictive validity appeared even when controlling for differences in sample composition in terms of demographics and political knowledge.

General Conclusion

This study suggests that Internet-based data collection represents a viable approach to conducting representative sample surveys. Internet-based data collection compromises sample representativeness, more so when respondents volunteer rather than being recruited by RDD methods. But Internet data collection improves the accuracy of the reports respondents provide over that rendered by telephone interviews.

Reference

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. Applied Cognitive Psychology, 5, 213-236.

Table 1: Sample Sizes, Response Rates, and Field Periods

	OSU Center for Survey Research	Knowledge Networks	Harris Interactive
Pre-election Survey			
Eligible Households	3,500	7,054	12,523
Participating Respondents	1,506	4,933	2,306
Response Rate	43%	28%[a]	NA[b]
Cooperation Rate[c]	51%	31%	
Panel Completion Rate[d]		70%	18%
Start Date	June 1, 2000	June 1, 2000	July 21, 2000
Stop Date	July 19, 2000	July 28, 2000	July 31, 2000
Post-election Survey			
Eligible Households	1,506	4,143[e]	2,306
Participating Respondents	1,206	3,416	1,028
Response Rate	80%	82%	45%
Start Date	Nov 9, 2000	Nov 8, 2000	Nov 9, 2000
Stop Date	Dec 12, 2000	Nov 21, 2000	Nov 26, 2000

[a]This figure is the product of 89% (the rate at which eligible RDD-sampled telephone numbers were contacted for initial telephone interviews) and 56% (the rate at which contacted households agreed to participate in the initial telephone interview and agreed to join the KN panel) and 80% (the rate at which households that agreed to join the KN panel had the WebTV device installed in their homes) and 70% (the rate at which invited KN panel respondents participated in the survey).

[b]A response rate cannot be calculated for the HI survey, because respondents volunteered to join their panels, rather than being recruited through "cold call" contacts.

[c]This is the rate at which people who were contacted through "cold calling" and invited to participate in the CSR survey or join the KN panel ended up completing the pre-election questionnaire for this study.

[d]This is the rate at which people who had agreed to join the KN or HI panel completed the pre-election questionnaire for this study.

[e]Of the 4,933 who completed all of the first three instruments, 790 members were excluded from assignment to the follow-up survey for the following reasons: (a) temporarily inactive status (being on vacation, health problems etc.), (b) some individuals had been withdrawn from the panel, and (c) some individuals had already been assigned to other surveys for the week of the election.

Table 2: Demographic Composition of Unweighted Pre-election Samples

		OSU Center for Survey Research	Knowledge Networks	Harris Interactive	2000 CPS March Supplement
Education	Some high school	7.0%	6.7%	2.0%	16.9%
	High school grad	31.3%	24.4%	11.8%	32.8%
	Some college	19.6%	32.3%	36.6%	19.8%
	College grad	30.1%	26.0%	25.8%	23.0%
	Postgrad work	12.0%	10.6%	23.7%	7.5%
	TOTAL	100.0%	100.0%	100.0%	100.0%
	N	1504	4925	2306	
	Average Error	4.6%	7.4%	13.9%	
Income	<$25,000	19.0%	14.3%	12.6%	30.5%
	$25-50,000	36.9%	32.5%	32.3%	28.3%
	$50-75,000	22.0%	27.5%	25.9%	18.2%
	$75-100,000	12.9%	13.8%	14.8%	10.1%
	$100,000	9.2%	11.9%	14.5%	12.5%
	TOTAL	100.0%	100.0%	100.0%	100.0%
	N	1138	4335	1976	
	Average Error	6.0%	6.8%	8.6%	
Age	18-24	10.0%	7.8%	8.0%	13.2%
	25-34	17.9%	19.1%	21.2%	18.7%
	35-44	24.5%	25.8%	21.5%	22.1%
	45-54	20.7%	23.0%	27.9%	18.3%
	55-64	12.1%	12.4%	15.5%	11.6%
	65-74	9.4%	7.7%	4.8%	8.7%
	75+	5.5%	4.2%	1.0%	7.4%
	TOTAL	100.0%	100.0%	100.0%	100.0%
	N	1496	4923	2306	
	Average Error	1.7%	2.7%	4.6%	
Race	White	78.5%	86.4%	89.6%	83.3%
	African American	9.7%	6.9%	3.6%	11.9%
	Other	11.8%	6.7%	6.8%	4.8%
	TOTAL	100.0%	100.0%	100.0%	100.0%
	N	1490	4721	2183	
	Average Error	4.7%	3.3%	5.5%	
Gender	Male	45.1%	49.2%	60.1%	48.0%
	Female	54.9%	50.8%	39.9%	52.0%
	TOTAL	100.0%	100.0%	100.0%	100.0%
	N	1506	4910	2306	
	Average Error	2.9%	1.2%	12.1%	
TOTAL AVERAGE ERROR		4.0%	4.3%	8.7%	

Table 3: Demographic Composition of Weighted Pre-election Samples

		OSU Center for Survey Research	Knowledge Networks	Harris Interactive	2000 CPS March Supplement
Education	Some high school	17.1%	12.3%	7.9%	16.9%
	High school grad	32.7%	33.5%	36.5%	32.8%
	Some college	19.8%	28.5%	26.9%	19.8%
	College grad	21.7%	18.2%	19.8%	23.0%
	Postgrad work	8.6%	7.4%	9.0%	7.5%
	TOTAL	100.0%	100.0%	100.0%	100.0%
	N	1504	4925	2250	
	Average Error	0.5%	3.8%	4.9%	
Income	<$25,000	19.0%	18.0%	24.8%	30.5%
	$25-50,000	37.1%	35.3%	29.8%	28.3%
	$50-75,000	22.4%	25.8%	20.6%	18.2%
	$75-100,000	13.4%	11.9%	11.6%	10.1%
	$100,000	8.1%	9.0%	13.0%	12.5%
	TOTAL	100.0%	100.0%	100.0%	100.0%
	N	1138	4335	1917	
	Average Error	6.4%	6.5%	2.3%	
Age	18-24	13.5%	9.8%	14.0%	13.2%
	25-34	15.3%	19.1%	18.9%	18.7%
	35-44	22.7%	22.8%	21.8%	22.1%
	45-54	17.8%	19.8%	20.4%	18.3%
	55-64	12.4%	13.4%	10.4%	11.6%
	65-74	12.5%	9.7%	12.3%	8.7%
	75+	5.8%	5.5%	2.2%	7.4%
	TOTAL	100.0%	100.0%	100.0%	100.0%
	N	1496	4923	2250	
	Average Error	1.6%	1.5%	1.9%	
Race	White	83.3%	82.8%	81.1%	83.3%
	African American	11.9%	10.0%	12.3%	11.9%
	Other	4.8%	7.2%	6.6%	4.8%
	TOTAL	100.0%	100.0%	100.0%	100.0%
	N	1490	4721	2132	
	Average Error	0.0%	1.6%	1.5%	
Gender	Male	46.9%	49.2%	48.2%	48.0%
	Female	53.1%	50.8%	51.8%	52.0%
	TOTAL	100.0%	100.0%	100.0%	100.0%
	N	1506	4910	2250	
	Average Error	1.1%	1.2%	0.2%	
TOTAL AVERAGE ERROR		1.9%	2.9%	2.2%	

Table 4: Post-election Vote-Related Questions (Weighted Samples)

		OSU Center for Survey Research	Knowledge Networks	Harris Interactive
Usually Voted in Past Elections?	Yes	74.4%	70.2%	83.7%
	No	21.0%	22.4%	13.3%
	Ineligible	4.6%	7.4%	3.0%
	TOTAL	100.0%	100.0%	100.0%
	N	1204	3408	1028
Voted in 2000 Presidential Election?	Yes	76.5%	72.2%	90.9%
	No	23.5%	27.8%	9.1%
	TOTAL	100.0%	100.0%	100.0%
	N	1205	3406	1028
Candidate Choice of Voters	Gore	49.9%	52.5%	43.5%
	Bush	46.6%	42.9%	50.1%
	Other	3.5%	4.6%	6.3%
	TOTAL	100.0%	100.0%	100.0%
	N	881	2406	920
Candidate Preference of Non-voters	Gore	47.2%	50.2%	48.6%
	Bush	36.4%	34.1%	27.1%
	Other	16.4%	15.6%	24.3%
	TOTAL	100.0%	100.0%	100.0%
	N	253	732	91

Table 5: Party Identification (Weighted Samples)

	OSU Center for Survey Research	Knowledge Networks	Harris Interactive
Strong Republican	12.1%	12.4%	18.1%
Weak Republican	15.3%	13.5%	11.9%
Independent-Leans toward Republicans	8.6%	8.4%	8.8%
Independent-Does not Lean	23.3%	23.6%	13.6%
Independent-Leans toward Democrats	9.8%	8.7%	9.9%
Weak Democrat	17.6%	17.0%	19.0%
Strong Democrat	13.3%	16.4%	18.5%
TOTAL	100.0%	100.0%	100.0%
N	1458	4803	2250

Table 6: Percent of Correct Answers to Political Knowledge Quiz Questions (Weighted Samples)

	OSU Center for Survey Research	Knowledge Networks	Harris Interactive
Do you happen to know what job or political office is now held by Trent Lott?	21%	23%	40%
Whose responsibility is it to determine if a law is constitutional or not?	64%	78%	83%
How much of a majority is required for the U.S. Senate and House to override a presidential veto?	42%	60%	73%
Which political party currently has the most members in the House of Representatives in Washington?	64%	77%	80%
Which party would you say is more conservative?	61%	70%	73%
Average Percentage of Correct Responses per Respondent	53%	62%	77%
N	1506	4935	2250

- Average percentage of correct responses per respondent was significantly different between all pairs of houses

Table 7: Pre-election Thermometer Ratings (Weighted Samples)

Target	Rating	OSU Center for Survey Research	Knowledge Networks	Harris Interactive
President Bill Clinton	0-10	24.9%	26.9%	36.3%
	11-20	5.0%	3.6%	3.4%
	21-30	7.7%	7.7%	5.5%
	31-40	5.3%	4.3%	4.5%
	41-49	1.8%	2.3%	2.0%
	50	14.7%	11.3%	8.0%
	51-60	8.3%	6.7%	4.7%
	61-70	6.6%	5.8%	5.4%
	71-80	12.2%	14.9%	10.1%
	81-90	6.4%	8.0%	9.0%
	91-100	7.3%	8.5%	11.2%
	TOTAL	100.0%	100.0%	100.0%
	MEAN	45.4	46.5	42.6
	STD DEV	32.0	33.8	36.6
	N	1491	4698	2249
Al Gore	0-10	12.3%	18.9%	25.4%
	11-20	5.1%	4.1%	4.1%
	21-30	6.8%	8.7%	7.4%
	31-40	8.1%	7.3%	5.2%
	41-49	2.3%	3.2%	2.2%
	50	23.4%	17.1%	12.8%
	51-60	11.8%	9.2%	8.0%
	61-70	8.5%	7.0%	5.5%
	71-80	14.1%	13.9%	14.2%
	81-90	4.3%	5.7%	7.7%
	91-100	3.2%	4.9%	7.4%
	TOTAL	100.0%	100.0%	100.0%
	MEAN	49.6	47.1	46.4
	STD DEV	25.4	29.0	32.8
	N	1481	4716	2248
George W. Bush	0-10	9.6%	14.9%	18.4%
	11-20	2.3%	3.6%	4.6%
	21-30	5.9%	8.0%	8.9%
	31-40	6.5%	8.0%	5.6%
	41-49	3.3%	3.6%	3.9%
	50	20.8%	17.6%	13.5%
	51-60	13.5%	9.0%	7.1%
	61-70	10.0%	6.2%	5.6%
	71-80	19.3%	16.5%	13.7%
	81-90	5.6%	7.0%	7.1%
	91-100	3.3%	5.6%	11.6%
	TOTAL	100.0%	100.0%	100.0%
	MEAN	54.7	50.6	50.9
	STD DEV	24.4	28.4	31.7
	N	1483	4726	2249

Table 8: Reliabilities of Thermometer Ratings and Vote Choice Measures (Weighted Samples)

	OSU Center for Survey Research	Knowledge Networks	Harris Interactive
Pre-election Thermometer Rating Difference	.69	.68	.86
Pre-election Vote Choice	.94	.91	.96
Post-election Thermometer Rating Difference	.64	.65	.81
Post-election Vote Choice	.88	.88	.91
N	869	2459	910

Table 9: Average Extent of Non-Differentiation in Each House (Weighed Samples)

	OSU Center for Survey Research	Knowledge Networks	Harris Interactive
Average non-differentiation	.07 N=1478	.08 N=4847	-.05 N=2250

- CSR and KN are not significantly different from one another.
- HI is significantly different from the other two houses.
- Non-differentiation scores are standardized.

Table 10: Unstandardized Regression Coefficients Testing Differences Between Houses in the Extent of Non-Differentiation (Weighted Samples)

	Tests of Differences Between Houses			
	CSR vs. KN	CSR vs. HI	KN vs. HI	N
House Only	.01 (.03)	-.12** (.03)	-.13** (.03)	8574
Controlling for Education	.01 (.03)	-.11** (.03)	-.13** (.03)	8565
Controlling for Education and Political Knowledge	.06* (.03)	-.01 (.03)	-.07** (.03)	8565

*p<.05, **p<.01

- Standard errors are in parentheses.
- For each pair of houses (e.g., CSR vs. KN), a negative coefficient means more non-differentiation in the first listed house than the second, and a positive coefficient means more non-differentiation in the second listed house than the first.

Table 11: Effects of Expected National Conditions if Candidate is Elected (Bush - Gore) on Pre-election Vote Choice (Bush=0, Gore=1) (Weighted Samples)

	OSU Center for Survey Research	Knowledge Networks	Harris Interactive
Economy	7.19	9.38	9.48
	(.48)	(.35)	(.48)
	N=1052	N=3544	N=1994
Foreign Relations	6.23	8.35	10.23
	(.43)	(.31)	(.54)
	N=1056	N=3545	N=1994
Crime	5.51	8.45	8.78
	(.40)	(.32)	(.45)
	N=1073	N=3548	N=1994
Race Relations	6.07	8.41	9.79
	(.46)	(.34)	(.53)
	N=1069	N=3548	N=1994
Pollution	3.40	5.76	5.88
	(.29)	(.22)	(.28)
	N=1064	N=3548	N=1994

- Probit coefficients appear above standard errors in parentheses.
- Expected national conditions if each candidate was elected were reported on 5-point scales ranging from “much better” to “much worse,” coded to range from 0 to 1.

Table 12: Tests of Difference Between Houses in Predictive Validity Using Pre-election Vote choice as the Dependent Variable (Weighted Samples)

Performance Domain	MODEL 1			MODEL 2		
	CSR vs. KN	CSR vs. HI	KN vs. HI	CSR vs. KN	CSR vs. HI	KN vs. HI
Economy	1.45*	1.88*	.43	1.11	1.06	-.05
	(.72)	(.83)	(.67)	(.74)	(.86)	(.68)
Foreign Relations	1.90**	3.86*	1.95**	1.61**	3.39**	1.79*
	(.60)	(.78)	(.68)	(.62)	(.81)	(.70)
Crime	3.12**	3.25**	.13	2.59**	2.66**	.07
	(.55)	(.64)	(.57)	(.57)	(.64)	(.56)
Race Relations	2.72**	4.39**	1.67*	2.47**	3.86**	1.40*
	(.62)	(.78)	(.70)	(.64)	(.81)	(.71)
Pollution	2.42**	2.81**	.38	2.03**	2.48**	.45
	(.40)	(.46)	(.40)	(.44)	(.51)	(.44)

$^{+}$p<.10; * p<.05; ** p<.01

- Probit coefficients appear above standard errors in parentheses.
- MODEL 1 tests simple differences between houses.
- MODEL 2 tests differences between houses controlling for demographics and political knowledge.

Use of Responsive Virtual Human Technology to Enhance Interviewer Skills Training[27]

Michael W. Link, Ph.D., Polly P. Armsby, BA, Robert Hubal, Ph.D, and Curry I. Guinn, PhD.

Abstract

Research on survey non-response suggests that advanced communication and listening skills are among the best strategies telephone interviewers can employ for obtaining survey participation, allowing them to identify and address respondents' concerns immediately with appropriate, tailored language. Yet, training on interaction skills is typically insufficient, relying on role-playing or passive learning through lecture and videos. What is required is repetitive, structured practice in a realistic work environment.

This research examines acceptance by trainees of an application based on responsive virtual human technology (RVHT) as a tool for teaching refusal avoidance skills to telephone interviewers. The application tested here allows interviewers to practice confronting common objections offered by reluctant sample members. Trainee acceptance of the training tool as a realistic simulation of "real life" interviewing situations is the first phase in evaluating the overall effectiveness of the RVHT approach. Data were gathered from two sources -- structured debrief questionnaires administered to users of the application, and observations of users by researchers and instructors. The application was tested with a group of approximately fifty telephone interviewers of varying skill and experience levels. The research presents findings from these acceptance evaluations and discusses users' experiences with and perceived effectiveness of the virtual training tool.

Responsive Virtual Human Technology (RVHT) involves the use of natural language processing and an emotive behavioral engine to produce natural, interactive dialogues with intelligent, emotive virtual-reality (VR) agents. RVHT has great potential for use in training interaction skills, such as those required for effective survey interviewing. However, our understanding of how people interact with responsive virtual humans (a.k.a. intelligent agents) is quite limited. Better understanding requires employing RVHT in training applications and conducting systematic use, usability, perception, and training-effectiveness assessments. Important questions yet to be answered include:

- Do intelligent agents make learning more accessible?
- How willing are students to accept intelligent agents as interactive partners in learning?
- What skills can be acquired, practiced, and validated using RVHT?
- What is involved in providing a convincing simulation of human interaction, realistic enough for the student to suspend disbelief and acquire skills that will transfer to a "live" environment?

Users' interactions with RVHT applications are little studied and poorly understood. The research presented here (and the larger research program from which it is drawn) provides an initial assessment of some of the issues associated with user interface design, user acceptance of computer-based training, and perceptions of the effectiveness of the training tool. As part of this assessment, usability assessments were conducted using instructor observations and a structured questionnaire. The assessment involved the use of an RVHT-based training tool for refusal avoidance at the outset of a telephone interview. Approximately fifty telephone interviewers of

[27] This work was supported by a research grant from the National Science Foundation (Grant No. EIA-0121211) and by a Strategic Capability Development Award from RTI (R9898-002).

varying experience levels, ages, genders, races, and educational backgrounds took part in the assessment.

Background

Intelligent agents are being used in fields as diverse as computer generated (military) forces (Hill, et. al., 1998), manufacturing (Regian, Shebilske, and Monk, 1992), medicine (Miksch, Chang, and Hayes-Roth, 1996), and theater (Loyall and Bates, 1997; Lundeberg and Beskow, 1999). Intelligent agents have not been employed in training on interaction skills, although such skills are critical in a number of fields. Therefore, advanced technologies for training these "soft skills" can be a considerable asset in training. There remain, however, questions that must be answered if intelligent agents are to reach the level of sophistication required for robust interaction skills training.

Interaction skills training is certainly a new educational area in which to apply advances in information technology, such as virtual reality (VR) and agent technology. To date, VR has been shown to be effective for equipment training (Adams, 1996), maintenance training (Barnett, Helbing, Hancock, Heininger, and Perrin, 2000), simulation of military field exercises (Shlechter, Bessemer, and Kolosh, 1992), and maneuvers (Magee, 1995), and acquisition of spatial knowledge (Ragian, Shebilske, and Monk, 1992). It can be used for interaction with unobservable processes or abstract concepts (Dede, Salzman, and Loftin, 1996), tasks that are costly or dangerous to perform (Loftin and Kenney, 1994), and for gaining situation awareness (Maggart and Hubal, 1998). VR systems have become steadily smaller, faster, cheaper, and easier to use (Psotka, 1995). RTI International has integrated a spoken natural language assistant with a VR-based maintenance training environment to enhance ease of use and facilitate instruction (Guinn and Montoya, 1998). Other relevant research efforts in enabling spoken interaction with virtual humans include work done at the University of Pennsylvania (Badler, Phillips, and Webber, 1993), MIT Media Lab (Cassell and Vilhjalmsson, 1999), University of Southern California (Lindheim and Swartout, 2001), and Oregon Graduate Institute (Cole et al, 1999; Massaro et. Al, 1998).

RVHT is a relatively recent advance in training technology. Few researchers have begun integrating emotion models with agents (Becheiraz and Thalmann, 1998; Elliott, 1993; Gratch, 2000; and Klein, 1998), and none for interaction training. Portraying emotions in a virtual human, it is argued, requires clearly defined emotional states, action that shows thought processes, and accentuation to reveal feelings (Bates, 1994). In general, lifelike "pedagogical agents" can lead to improvements in problem-solving ability and can engage and motivate trainees (Johnson, Rickel, and Lester, 2000; Lester et. al, 1997). Most importantly, RVHT can open entirely new capabilities for computer-based training of interpersonal skills, and can provide the benefits of reduced training costs, individualized tutoring, and greater student convenience that are associated with computer-based training (Field, et. al., 1999).

Today, interaction skills training usually relies on peer-to-peer role playing or passive learning through videos. These approaches lead to a critical training gap, because the students are limited in the practice time and the variety of scenarios that they encounter. Nevertheless, it is exactly this practice that leads to significant on-the-job benefits.

Table 1 (adapted from Hubal, et al. 2000) presents a comparison of approaches to interaction skills training. Constraints imposed by the current approach include insufficient time in the classroom to conduct effective practice sessions, forced and unrealistic role-playing exercises, and little time or ability for individual feedback and coaching to trainees from the instructor. By using virtual humans to simulate realistic interactions, RVHT increases the amount of time trainees spend acquiring and practicing critical skills, reduces passive learning (information and skills are retained better through active learning), improves the realism of practice sessions, and enables intelligent tutoring (Graesser et al, 2000).

Table 1. Comparison of Training Approaches

Role	Traditional Approach	Role-player	RVHT Approach	Role-player
Trainee (e.g., medical practitioner, police recruit, survey interviewer)	Student's ability to learn dependent on: ❑ relevance of role-play scripts, ❑ time available during training to conduct role-plays or mock interviews, ❑ acting ability of role-play Partner, ❑ observations made by role-play Partner and/or by Instructor.	Student	Student's ability to learn enhanced by: ❑ using numerous age-appropriate role-play or mock interview scripts, for more practice of critical skills, ❑ interacting with different virtual role-play partners, ❑ knowing that actions are observed and tracked, ❑ ability to replay interaction.	Student
Conversation Partner (e.g., patient, mentally disturbed consumer, household respondent)	❑ Partner must be present, available. ❑ Partner must act out a role that s/he will not always understand (non-essential learning activity). ❑ Partner is of a specific gender/age/ethnicity, limiting realism of practice.	Other person (e.g., actor, other student, Instructor)	❑ Ability to simulate conditions impossible with a human. ❑ Standardization of responses. ❑ Different virtual partners of gender/age/ethnicity and having different personalities.	Virtual human
Observer/ Evaluator	❑ Role-play Partner must take on second role, again a role not taken in live environment. ❑ Role-play Partner, if other student, is in passive learning mode.	Other person	❑ Ability to track all interactions with virtual role-play partner for use in feedback, guidance, assessment. ❑ Knowledge of all characteristics of virtual partners.	Second virtual human
Coach/Tutor	❑ Instructor must rely on role-play Partner for assessment of Student when not actually witnessing interaction. ❑ Only means of replaying interaction is through video, requiring an additional person and equipment.	Instructor or Supervisor	❑ Virtual tutor has ability to guide learning as it occurs. ❑ Instructor can use automatically collected interaction information for assessment & replay, as well as actually witness interaction. ❑ Instructor can convey "what-if" scenarios.	Second virtual human Instructor

We stress that using virtual humans as interaction partners has disadvantages as well as advantages. Most importantly, the current state-of-the-art does not produce fully realistic conversational partners. Advances in utilizing natural language dialog features and behavior models will add tremendously to the realism. From a larger perspective, though, one must understand that virtual training is simply one component of training. Just as a trainee must "skin his/her knuckles" on actual machines in validating maintenance and diagnostic skills, so a trainee

must interact with people in validating interaction skills (Helms, Hubal, Triplett, 1997). Virtual environments, though, offer advantages in reliability, repetitiveness, flexibility, throughput, and distribution that lead directly to overall cost-effectiveness of training (Field, et al, 1999).

Mechanics of the Training Application

One of the most difficult skills for a telephone interviewer to learn – and for an instructor to teach – is gaining cooperation from sample members and avoiding refusals. In telephone interviewing in particular, the first 30 seconds on the telephone with a sample member is crucial. Sample members almost automatically turn to phrases such as, "I don't do surveys," "I don't have time," "I'm just not interested" to avoid taking part in surveys. Non-response research suggests that the best approach to obtaining participation is for the interviewer to immediately reply with an appropriate, informative, tailored response (Camburn, Gunther-Mohr, & Lessler, 1999; Groves & Couper, 1998; Groves, 2002). How can the interviewer learn and then practice those responses before the survey begins, without creating more refusals during their first few weeks at work by being placed on the telephone unprepared?

The approach tested here involves the use of an RVHT-based application to simulate the environment an interview faces during the first 30 to 60 seconds of a telephone interview. The application allows interviewers to practice their skills in gaining cooperation in a self-paced, realistic environment. The software is designed such that interviewers begin with an introduction and then need to respond to a series of these objections or questions raised by the "virtual respondent." The interviewer's responses are captured electronically and processed by a natural language speech processor. Based on the content of the interviewer's speech, the software launches another objection/question or ends the conversation by either granting the interview or hanging-up the telephone (see Figure 1).

The application uses speech recognition and a behavior engine (for determining the intelligent agent's emotional state) to produce natural dialogues with the trainees. The speech recognizer uses a basic dictionary of common words as well as a specific dictionary for each turn of a conversation. The specific dictionary consists of up to 200 words based on behavioral observations of real world events. These specific dictionaries are dynamic, therefore, changing with each turn of the conversation. During the development of the application tested here, the researchers monitored live interviews and behavior coded the responses of interviewers and sample members. These behavioral observations were then modeled, using the dictionaries and the emotional state behavior engine. Thus the specific dictionaries created for capturing responses from an interviewer to a respondent who said, "I'm too busy" in a harsh tone varied somewhat from the dictionaries created for when the respondent gave the same objection but in a softer, more reasoned tone. As trainees used the application, the emotional state of the virtual respondent varied from scenario to scenario, thus giving trainees exposure to an array of objections and emotional states. The scripts launched by the RVHT program were recorded in both a male and a female voice to add variety to the program. In all a total of six basic objections were recorded in four different tones of voice for both a male and female virtual respondent. Thus a total of 48 different practice scenarios could be offered to the trainees.

Assessment of the RVHT-based Interviewer Training Application

A primary goal of the overall research program of which this study is a part is to determine if RVHT can be an effective technology for interaction training across a broad spectrum of ethnic and socioeconomic backgrounds, jobs, and job levels. In particular, we investigate whether users find RVHT interactions accessible and acceptable. The effectiveness of this technology depends upon its ability to provide appropriate learning experiences, its ability to engage the trainee, and its acceptability to disparate users.

An "accessible" user interface is one that is easy to learn and easy to use, and can result in measurable goals such as decreased learning time and greater user satisfaction (i.e., acceptance) (Weiss, 1993). Characteristics of easy to learn and easy to use interfaces have been described as having navigational and visual consistency, clear communication between the user and application, appropriate representations, few and non-catastrophic errors, task support and feedback, and user control (Nielsen, 1993; Norman, 1993; Sneiderman, 1992; Weiss, 1993).

The assessment provided here of the interviewer training module is based on researcher / instructor observations, and user debriefings in the form of a questionnaire. Empirical data were collected on users' observed ability to interact with the application as well as their perception of the interaction. The training application was tested with a group of approximately 50 telephone interviewers of varying ages, races, experience and education levels. Trainees who participated in the assessment used the application to practice communication and thinking skills required with real conversation partners. These skills involve the use of adaptive strategies, listening and responding to the other's concerns.

To evaluate the *accessibility* of the application we focused on the following:

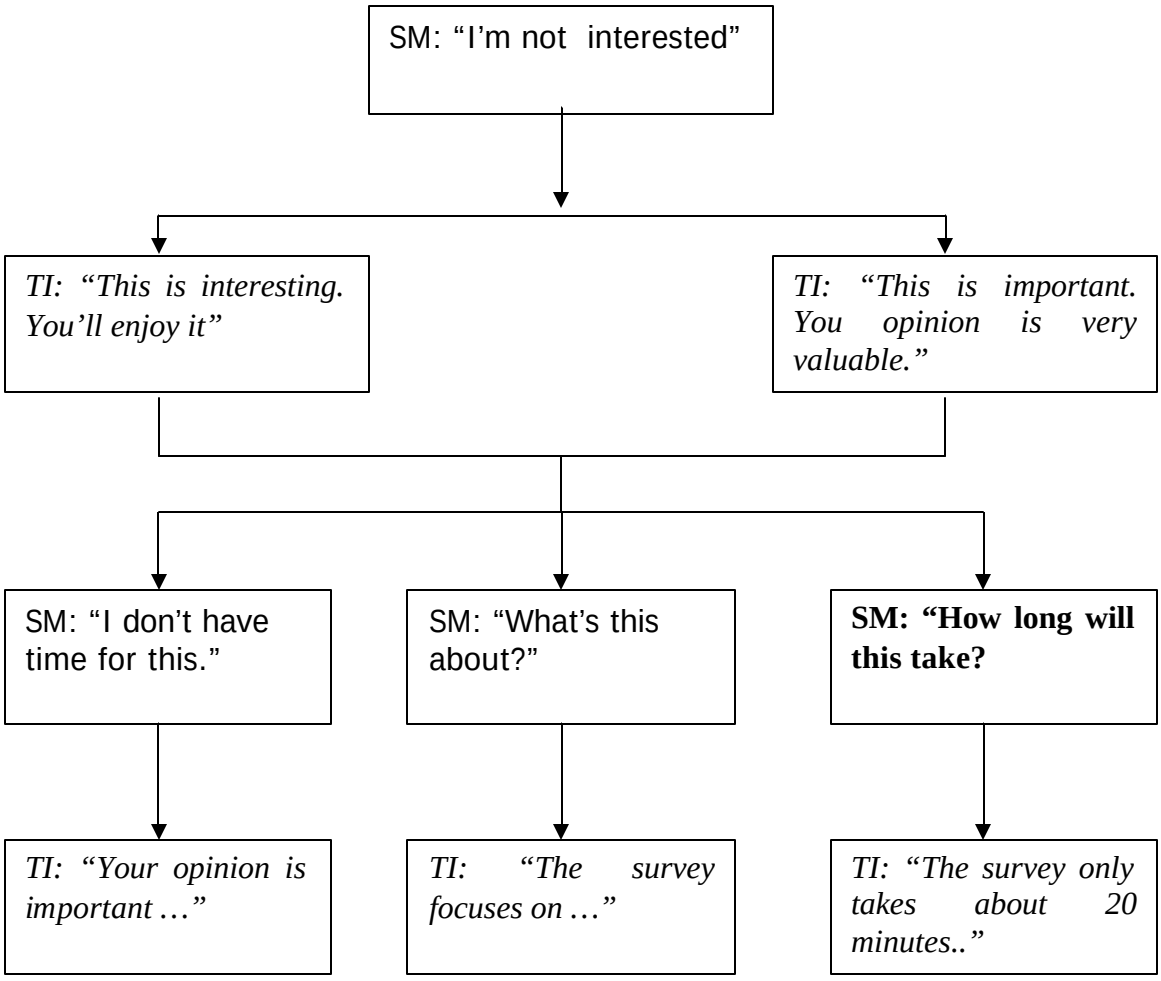
- Do users understand the basic features of the application?
- Are users able to complete each task and exit the application?
- Do users understand where they are in the application?
- Are different users (e.g., based on age, time on the job, and education level) equally able to use the application?

Instructor/researcher observation was used to assess more directly the interaction between the user and the training application, addressing questions such as:

- When there are problems (e.g., the virtual human seems to respond inappropriately), what are user reactions?
- Are inappropriate responses due to a programming error, misunderstanding in the interaction, or incorrect user behavior?
- What knowledge engineering improvements will lead to better recovery by the application when inappropriate responses occur?

Analysis of these questions will provide clues as to how smoothly the application runs, or when and why difficulties arise in its use.

Figure 1
Example of Dialogue Flow



The question of whether and why participants "accept" or "reject" the virtual training environment is also central to this research. To evaluate *acceptance* of the application by the trainees, we debriefed participants using a structured questionnaire to gauge reactions and engagement in the application. In particular we are interested in the following:

- Are the virtual humans realistic enough for the users? Why or why not?
- How fast and accurate is the speech recognition?
- When recognition is inaccurate, does the application respond reasonably?
- Overall, do the users "buy into" the virtual environment?
- Could trainees detect changes in the emotive states of the virtual human using only audio cues?
- Did the trainee perceive any gains in skills from using the application?
- Would they use the application again and/or recommend it's use by others?

While some of these acceptance measures may be particular to the specific application tested, most help in gaining a general understanding of user satisfaction and affect with RVHT.

As part of the evaluation process, data were collected using a questionnaire filled out by the interviewers and notes made by instructors and researchers who observed the training sessions. The questionnaire asked questions related to users' perceptions of the realism of the interactions with the "virtual human," ease of use of the software, the perceived effectiveness of the training sessions, and some basic background characteristics of the users. In all, a diverse group of 48 interviewers filled-out the questionnaires (96% of the software users). A breakdown of some of the demographic characteristics of this set of users is provided on Table 2. Finally, each training session was observed by either the researchers or training instructors, who made notes of their observations. These observations are included as part of the analysis.

Findings

The questions posed to the interviewers were designed to assess their perceptions and experiences in using the RVHT training tool in four basic areas: ease of use of the software, realism of the training environment, impact on skill development, and desire to recommend or use the software again. Although this is the first detailed look at how users interact emotive intelligent agents for soft-skills development, we can formulate some hypotheses regarding how different types of users might respond based on how users generally differ in their use and acceptance of other computer-based tools. For example, we might expect to find that trainees who are younger, have more education, and are more comfortable using computers in general to have fewer difficulties in using the system. Likewise, we might expect that more experienced interviewers might not find the training tool as useful as inexperienced interviewers because the more experienced interviewers will have already developed and honed their refusal avoidance skills (a supposition that mirrors the finding of Groves, 2002). To examine possible differences in accessibility and acceptance of the program, we cross-tabulated all of the closed-ended questions in the questionnaire with the demographic variables listed on Table 2. Significant differences are noted in the text.[28]

[28] Because of the small number of observations (N=48) we also created dichotomous variables for both the dependent variables (collapsing scales where possible) and independent variables (collapsing or combining variables with 3 or more values). These variables were also examined to determine if

Table 2
Demographics of RVHT Trainees

Characteristic	N	%
Sex		
Male	12	25%
Female	36	75%
Education		
High School/GED	2	4%
Some College	12	25%
Four Year Degree	25	52%
Advanced Degree	9	19%
Age		
18-21	7	15%
22-29	17	35%
30-39	8	17%
40-49	7	15%
50+	9	18%
Race		
African-American	34	70%
White	7	15%
Hispanic	7	15%
Experience		
< 1,000 hours	19	40%
1,000 – 1,999 hours	17	35%
2,000+ hours	12	25%
Comfort with Keyboard		
Slow-touch typing	15	31%
Fast-touch typing	33	69%

significant differences among subgroups could be identified. Significance was evaluated at the ($p < .10$) level.

Table 3
Interviewer's Evaluation of the RVHT Training Software

	Extremely	Very	Somewhat	Not Too	Not At All
In general, how easy was the application to use?	52.1% (25)	31.3% (15)	12.5% (6)	4.2% (2)	0 % (0)
In general, how realistic did you find the overall conversation with the "virtual respondent"?	2.1% (1)	14.6% (7)	43.8% (21)	16.7% (8)	22.9% (11)
In general, how realistic did you find the objections, concerns, questions posed by the "virtual respondent"?	12.5% (6)	35.4% (17)	39.6% (19)	8.3% (4)	4.2% (2)
How easily could you determine the "virtual respondent's" emotional state or attitude based on the tone of his/her voice?	22.9% (11)	43.8% (21)	29.2% (14)	4.2% (2)	0% (0)
How easily could you determine the "virtual respondent's" emotional state or attitude based on the words used or objectives raised by him/her?	8.3% (4)	54.2% (26)	27.1% (13)	10.4 % (5)	0% (0)

Ease of Use of the Application

Training software should be accessible to users; that is, it should be relatively easy to use. As shown on Table 3, users of the RVHT software seemed to find it very accessible to use, with 84% indicating the software was either extremely easy or very easy to use (52% extremely, 31% very, 13% somewhat, 4% not too, 0% not at all). Nearly everyone found the written instructions (96%) and the verbal instructions (98%) that accompanied the training to be clear and accurate. Only eight (17%) of the 48 trainees indicated that they required additional assistance to use the training software (after the initial training received by all trainees).

The only significant difficulty encountered by the users were "insufficient memory" errors received on some of the training stations. The version of the application tested did, at times, use up considerable CPU memory. Once the machines were adjusted to handle the software memory requirements, the error messages were no longer an issue.

Realism of the Training Environment

The promise of RVHT-based training tools is that they can simulate a "real" environment, thereby allowing trainees repetitive practice in conditions that are as close as possible to what they will encounter on the job. For this particular application, the "virtual respondent" needed to mirror the behaviors and emotions of real respondents encountered when doing live interviewing. This means delivering an array of objections to the trainees in different tones of speech and emotional levels in a fast-paced manner. Interviewers were asked a series of

questions to try to assess how well they accepted the virtual environment as a substitute for real work conditions. In other words, do they "buy-into" the virtual environment?

The answer is somewhat mixed. In general, trainees did not find the virtual environment to be realistic and they cited two primary reasons: the slowness of the response of the "virtual respondent" and the limited number of different objections/questions offered by the "virtual respondent." They did, however, find the responses that were offered to be realistic and stated that they could detect and respond to changes in tone and emotional cues offered by the "virtual respondents." A majority of the trainees also indicated that they felt the sessions helped them to improve their skills needed at the outset of an interview either somewhat or a lot.

When asked, *In general, how realistic did you find the overall conversation with the 'virtual respondent*,' 17% said they thought it was extremely or very realistic, 44% said it was somewhat realistic, 17% not too realistic and 23% not at all realistic (see Table 3). Slowness of the "virtual respondents" in replying (due to the lag caused by the speech recognizer as it interpreted the interviewer's responses and determined the next script to launch) was the primary problem cited by interviewers. Over three-quarters (77%) of the users felt the response time was too slow (4% felt it was too fast and 19% indicated the speed was just right). Perhaps not surprisingly, trainees who describe themselves as "fast-touch typists" were more likely than those who indicated they were "slow-touch typists" to say the response time was too slow (82% fast-touch vs 67% slow-touch; $p < .08$ chi-sq). Interviewers who are more comfortable at a keyboard and who, it can be surmised, tend to get through an interview faster were the ones most put-off by the perceived slowness of the response time.

The trainees were, however, more positive when evaluating the realism of the objections and questions offered by the "virtual respondent." A plurality (48%) indicated that the content of what was said was either extremely or very realistic, with 40% saying it was somewhat realistic, 8% not too realistic, and 4% not at all realistic. They also felt it was relatively easy to determine the emotional state of the virtual respondent based on the tone of voice they heard (23% extremely easy, 44% very easy, 29% somewhat easy, and 4% not too easy; no one indicated that they could not determine the avatar's emotional state from the tone of the "virtual human's" voice). Likewise, the content of the speech used by the avatar was also a good cue to trainees as to the "virtual human's" emotional state: 8% extremely easy to tell, 54% very easy, 27% somewhat easy, 10% not too easy, 0% not at all easy.

Being able to recognize changes in the emotional state of the virtual respondent changed – at least in the minds of many trainees – how the interviewer approached the situation. Nearly 60% indicated that they behaved differently in the practice scenario based on the tone of the virtual respondent's voice. Interestingly, a higher percentage of women than men reported reacting differently to the changing tone of the avatar's voice (women 67% v. men 33%, $p < .04$ chi-sq.). Similarly, 54% said they treated the situation differently based on the actual words used by the avatar in expressing a concern or voicing an objection. There were, however, no differences between men and women on this question. When asked how they behaved differently, interviews said they tended to soften and take a more conciliatory tone when the virtual respondent seem to grow more hostile or angered, and they mirrored the tone when the virtual respondent seemed more pleasant. Likewise, they reported tailoring the content of their responses to try to meet the

objections or questions of the virtual sample member rather than simply moving forward with their script. It seems, therefore, that the both the content of the objections raised by the virtual respondent and the emotional behavior of the "virtual human" were generally accepted by the trainees and caused them to react differently within the various training scenarios.

When asked in an open-ended format to list some of the problems with the realism of the software, many cited the slowness and others indicated that the limited number of objections raised by the virtual respondent made the sessions less realistic than what they encounter on the telephone. Because this was the first iteration of the software, a conscious decision was made at the design phase to maintain a limited set of six main objections and questions ("I'm not interested," "I'm too busy," "What is the survey about?", "I don't have time right now," "How was I selected?", and "How long will this take?"). These six responses, however, were recorded in four different tones of voice (ranging from calm to upset) and recorded in both a male and a female voice. A total of 48 possible practice scenarios were, therefore, actually possible (6 responses * 4 tones of voice * 2 sexes). It appears, however, that while the interviewers do recognize and react to the different emotional cues they obtain from the different scenarios, they don't necessarily process these as being very distinct. They focus more on the actual content of the argument (regardless of the tone of voice or whether the voice is a male or female) when considering how diverse the scenarios offered are. In designing future versions of the software this will need to be considered to increase interviewer acceptance of the training tool as a realistic simulation of the environment within which they must work.

Impact on Skill Development

The purpose for allowing trainees to operate within a virtual environment is to allow them to develop and hone essential skills before entering the "real" environment, thereby reducing the amount of "on the job" skill development required. For telephone interviewers, this means an opportunity to practice their skills at gaining cooperation at the outset of an interview. Practice in a virtual environment, it is hoped, will allow interviewers – particularly new interviewers – to develop, practice, and hone these skills before getting on the telephone. New interviewers can do considerable damage at the outset of a telephone study, generating a large number of refusals as they gain comfort and confidence on the telephone. If practice within a virtual environment at the beginning of a project can reduce the numbers of initial refusals even modestly, then the training program will have value. While longer-term assessments of the effectiveness of the RVHT software will need to include examination of more objective measures of improved performance, this preliminary assessment focused on the user's assessment of the impact of the training on their own skill development.

Trainees were asked to evaluate if they thought the RVHT software increased their abilities in six different areas (see Table 4). Nearly three-quarters of the trainees felt that the practice sessions increased a lot or somewhat their ability to respond to questions and concerns by sample members. Approximately 56% felt it helped them a lot or somewhat in better gaining respondent cooperation at the outset of an interview. Likewise, over half felt it helped in their ability to adapt to differences in respondents' tone or voice or perceived moods and to adapt to differences in the speed and pace of different sample members' speech. About half of the trainees also thought that the sessions helped them a lot or somewhat in avoiding refusals at the outset of an interview.

Table 4
Interviewer's Perceptions of Effectiveness of RVHT Training Software

	A Lot	Somewhat	A Little	Not at All
Respond to questions / concerns raised by sample members	25.0% (12)	47.9% (23)	16.7% (8)	10.4% (5)
Better gain respondent cooperation during the first seconds of a call	25.0% (12)	31.3% (15)	29.2% (14)	14.6% (7)
Enhance your ability to adapt to differences in respondents' tone/mood	25.0% (12)	29.2% (14)	29.2% (14)	16.7% (8)
Think on your feet	20.8% (10)	39.6% (19)	27.1% (13)	12.5% (6)
Enhance your ability to adapt to differences in respondents pace of speaking	18.8% (9)	33.3% (16)	27.1% (13)	20.8% (10)
Avoid refusals at the outset of an interview	16.7% (8)	35.4% (17)	31.3% (15)	16.7% (8)

Once again, while more objective measures of increased ability to gain cooperation from sample members are needed in the longer-term evaluation of this training tool, it does appear that trainees perceive an increase in their ability to deal with various facets of the opening of an interview as a result of their training sessions.

Would They Use The RVHT Training Tool Again?

An effective training tool is also one that trainees should enjoy using, would use again, and recommend to others (see Table 5). Approximately two-thirds (65%) of the users said that they found using the RVHT software to be fun and enjoyable. Interestingly men were significantly more likely than women to say that they found the sessions to be enjoyable (92% men vs. 56% women, $p < .05$ chi-sq). Nearly three-quarters (73%) said they would like to use the software again. In addition, 83% said they would recommend the program as a training tool for other interviewers. In open-ended responses, a number of interviewers indicated that it would be a very good practice vehicle for new or less experienced interviewers.

Conclusions

This initial assessment of an RVHT-based training tool for telephone interviewers provides some valuable insights into how trainees access and accept virtual environments as practice labs and "virtual humans" as training partners. There were aspects of the training program that interviewers clearly liked, such as the ability to do repeated practice of frequently asked questions, being able to distinguish different emotional states from the tone of voice and speech content of the virtual respondent, and the opportunity to learn to think on their feet in a simulated environment before being placed into a live interviewing situation.

Table 5
Recommendation for Future Use of RVHT Training Tool

Assessment Questions:	**Yes**	**No**
Would you recommend the RVHT program as a training tool for other interviewers?	83% (40)	17% (8)
Would you like to use the RVHT program again as a training tool?	73% (35)	27% (13)
Was using RVHT fun and enjoyable?	65% (31)	35% (17)

There were also aspects that the interviewers did not like, such as the slowness of the response of the virtual respondent and the perceived lack of variety in the scenarios that were presented. This provides constructive feedback for the engineering and improvement of the software. Adding additional scenarios is a relatively easy process, involving research into the "normal" flow of such scenarios and simple scripting and programming. The responsiveness issue is a more fundamental matter, reflecting the current state-of-the-art in speech recognition. For virtual training partners to be more readily accepted, the underlying speech recognition technology needs to be improved, providing quicker, more efficient processing of the input from interviewers and more rapid launching of responses by the virtual respondent. While our research focused on a specific training application, the results have implications for a broader range of training and educational RVHT-based tools. The lessons learned here can be used to inform the development of tools in these other areas.

We do not anticipate RVHT-based training will replace instructor-led training, but we expect that combinations of RVHT-based training and instructor-led training will significantly reduce training development costs (with new development tools) and training delivery costs, while increasing trainee throughput and maintaining training effectiveness and consistency. As an additional return-on-investment, RVHT-based training can provide inexpensive, focused sustainment (i.e., refresher) training. We feel it is important to continue to investigate more robust and effective RVHT models and more efficient means of creating the models, to better understand user preferences and acceptance of RVHT, and to determine how best to use RVHT in combination with other training methods to provide cost-effective training on critical interaction skills.

References

Adams, N. (1996). A Study of the Effectiveness of Using Virtual Reality to Orient Line Workers in a Manufacturing Environment. Motorola University, unpublished dissertation.

Badler, N.I., Phillips, C.B., & Webber, B.L. (1993). Simulating Humans: Computer Graphics, Animation, and Control. Oxford Univ. Press.

Barnett, B., Helbing, K., Hancock, G., Heininger, R., & Perrin, B. (2000). An Evaluation of the Training Effectiveness of Virtual Environments. Presented at the Interservice/Industry Training, Simulation and Education Conference. November 30, 2000, Orlando, FL.

Bates, J. (1994). The Role of Emotion in Believable Agents. Communications of the ACM, Special Issue on Agents, July, 1994.

Becheiraz, P., & Thalmann, D. (1998). A Behavioral Animation System for Autonomous Actors personified by Emotions, Proceedings of the First Workshop on Embodied Conversational Characters (WECC '98), Lake Tahoe, California.

Camburn, D.P., Gunther-Mohr, C., & Lessler, J.T., (1999). Developing New Models of Interviewer Training. International Conference on Survey Nonresponse, Portland, OR, October 28-31, 1999.

Cassell, J., & Vilhjálmsson, H.H. (1999). Fully Embodied Conversational Avatars: Making Communicative Behaviors Autonomous. Autonomous Agents and Multi-Agent Systems: 2, 45-64.

Cole, R., et. al. (1999). New tools for interactive speech and language training: Using animated conversational agents in the classrooms of profoundly deaf children. In Proceedings of ESCA-MATISSE ESCA/SOCRATES Workshop on Method and Tool Innovations for Speech Science Education, London, UK.

Dede, C., Salzman, M., & Loftin, B. (1996) ScienceSpace: Research on Using Virtual Reality to Enhance Science Education. In P. Carlson & F. Makedon (Eds.), Proceedings of the 1996 ED MEDIA Conference (pp. 172-177). Charlottesville, VA: Association for the Advancement of Computers in Education.

Elliott, C. (1993). Using the Affective Reasoner to Support Social Simulations. In Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence, pp. 194-200, Chambery, France, August 1993. Morgan Kaufmann.

Field, S.S., Frank, G.A., Helms, R.F., & Hubal, R.C. (1999). Army Learning & Training Effectiveness Symposium. Final Report, March 30, 1999. Submitted to Battelle RTP Office, Subcontract # DAAH04-96-C-0086, Agreement #99020, Delivery Order #0366, Dated February 12, 1999.

Graesser, A., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R., & the Tutoring Research Group (2000). AutoTutor: A simulation of a human tutor. Journal of Cognitive Systems Research, 1,35-51.

Gratch, J. (2000). Modeling the Interplay Between Emotion and Decision-Making. Proceedings of the Ninth Conference on Computer Generated Forces & Behavioral Representation, May 16-18, 2000, Orlando, FL.

Groves, R. (2002). Principles and Practices in Non-response Reduction. Presentation at the 2002 Respondent Cooperation Workshop sponsored by the Council for Marketing and Opinion Research, New York, NY.

Groves, R., & Couper, M. (1998). Nonresponse in Household Interview Surveys. New York: John Wiley & Sons, Inc.

Guinn, C.I., & Montoya, R.J. (1998). Natural Language Processing in Virtual Reality, Modern Simulation and Training, pp. 44-55.

Helms, R.F., Hubal, R.C., & Triplett, S.E. (1997). Evaluation of the Conduct of Individual Maintenance Training in Live, Virtual, and Constructive (LVC) Training Environments and their Effectiveness in a Single Program of Instruction. Final Report, September 30, 1997. Submitted to Battelle RTP Office, Subcontract # TCN 97031, Delivery Order #0027, Dated April 16, 1997.

Hill, R., Chen, J., Gratch, J., Rosenbloom, P., & Tambe, M. (1998). Soar-RWA: Planning, Teamwork, and Intelligent Behavior for Synthetic Rotary Wing Aircraft, Proceedings of

the Seventh Conference on Computer Generated Forces & Behavioral Representation, May 12-14, 2000, Orlando, FL.

Hubal, R.C., Kizakevich, P.N., Guinn, C.I., Merino, K.D., & West, S.L. (2000). The Virtual Standardized Patient-Simulated Patient-Practitioner Dialogue for Patient Interview Training. In J.D. Westwood, H.M. Hoffman, G.T. Mogel, R.A. Robb, & D. Stredney (Eds.), Envisioning Healing: Interactive Technology and the Patient-Practitioner Dialogue, 133-138. IOS Press: Amsterdam.

Johnson, W.L., Rickel, J.W., & Lester, J.C. (2000). Animated pedagogical agents: face-to-face interaction in interactive learning environments. International Journal of Artificial Intelligence in Education, 11.

Klein, G. (1998). Sources of Power. MIT Press.

Lester, J., Converse, S., Kahler, S., Barlow, S., Stone, B., & Bhogal, R. (1997). The Persona Effect: Affective Impact of Animated Pedagogical Agents. In S. Pemberton (Ed.), Human Factors in Computing Systems: CHI'97 Conference Proceedings. 359-366. New York: ACM Press.

Lindheim, R., & Swartout, W. (2001). Forging a New Simulation Technology at the ICT. Computer, 34(1), 72-79.

Loftin, R.B., & Kenney, P.J. (1994). Virtual Environments in Training: NASA's Hubble Space Telescope Mission. 16th Interservice/Industry Training Systems & Education Conference, Orlando, FL.

Loyall, A.B., & Bates, J. (1997). Personality Rich Believable Agents That Use Language. Proceedings of the First International Conference on Autonomous Agents, February 1997, Marina del Rey, CA.

Lundeberg, M., & Beskow, J. (1999). Developing a 3D-Agent for the August Dialogue System. Proceedings from AVSP '99, Santa Cruz, CA.

Magee, L.E. (1995). Virtual Reality and Distributed Interactive Simulation for Training Ship Formation Manoeuvres. Proceedings of the 36th NATO Defense Research Group (DRG) Seminar.

Maggart, L.E., & Hubal, R.C. (1998). A Situation Awareness Model. In S.E. Graham & M.D. Matthews (Eds.), Infantry Situation Awareness: Papers from the 1998 Infantry Situation Awareness Workshop. U.S. Army Research Institute: Alexandria, VA.

Massaro, D.W., Cohen, M.M., Beskow, J., Daniel, S., & Cole, R.A. (1998). Developing and Evaluating Conversational Agents. In Proceedings of Workshop on Embodied Conversation Characters (WECC), Lake Tahoe.

Miksch, S., Chang, K., & Hayes-Roth, B. (1996). An Intelligent Assistant for Patient Health Care. Knowledge Systems Laboratory Report No. 96-19, Stanford University.

Mills, K.C., Parkman, K.M., Smith, G.A., & Rosendahl, F. (1999). Prediction of driving performance through computerized testing: High-risk driver assessment and training. Transportation Research Record, 1689, 18-24.

Nielsen, J. (1993). Usability Engineering. Boston: Academic Press.

Norman, D.A. (1993). Things That Make Us Smart . Reading, MA: Addison-Wesley.

Psotka, J. (1995). Immersive Training Systems: Virtual Reality and Education and Training. Instructional Science 23, 405-31.

Regian, J.W., Shebilske, W.L., & Monk, J.M. (1992). Virtual reality: An instructional medium for visual-spatial tasks. Journal of Communication 42, 136-149.

Rousseau, D., & Hayes-Roth, B. (1997). Improvisational Synthetic Actors with Flexible Personalities. KSL Report #97-10, Stanford University.

Shlechter, T.M., Bessemer, D.W., & Kolosh, K.P. (1992). Computer-Based Simulation Systems and Role-playing: An Effective Combination for Fostering Conditional Knowledge. Journal of Computer Based Instruction 19, 110-114.

Schneiderman, B. (1992). Designing the User Interface: Strategies for Effective Human-Computer Interaction. Reading, MA: Addison-Wesley.

Weiss, E. (1993). Making Computers People-Literate. San Francisco: Jossey-Bass.

Discussion Paper: Capitalizing on Technology to Enhance Survey Response

Carol C. House
National Agricultural Statistics Service

This session of our workshop focuses on "Capitalizing on Technology to Enhance Survey Response". Before commenting on the two papers presented in the session, I will speak more broadly about this topic. First I ask: what do we mean by "technology" in this context of survey response. There exists, in fact, a fairly broad smorgasbord of related technologies, each providing tools useful in the survey process. It may be helpful for discussion purposes to group these technologies into the following categories:

- Tools for Use by Interviewers
- Tools for Use by Respondents
- Tools for Use by the Survey Designer or Survey Administrator

These are not mutually exclusive categories, and a given technology may be an excellent tool in more than one area.

Moving to the concept of "enhancing survey response", we can fashion a similar taxonomy. Clearly, as survey practitioners we want both more response (i.e. higher response rates) and improved quality for the responses that we do receive. In order to increase responses rates, we generally need either to make it easier for potential respondents to respond, i.e. reduce the burden of responding, or we need to provide more compelling reasons why they should spend their time and give up information. Thus, technology may be used to enhance survey response in any of the following ways:

- Make it easier to respond
- Communicate compelling reasons to respond
    - Facilitate quality responses

It is informative to look at which tools enhance response in which ways. Figure 1 provides a matrix of technologies, with columns defined by the entity who uses the technology and rows defined by how the technology enhances response. This is not intended to be a complete listing of technologies, and I may have left out some important examples. However, the clustering of technologies provides some insight.

Figure 1: Technology tools to enhance survey response, categorized by who uses the tool and the purpose of the tool.

PURPOSE OF TOOLS	TOOLS FOR USE BY		
	INTERVIEWER	**RESPONDENT**	**DESIGNER OR SURVEY ADMIN.**
EASE BURDEN OF RESPONSE	CATI / CAPI Wearable technology Geo-positioning devices	CASI Touch-tone / Voice Recognition Fax Email Web Collection Other EDR Methods	**CATI / CAPI**
COMMUNICATE REASONS TO RESPOND			RVNT – Training
IMPROVE RESPONSE QUALITY	CATI / CAPI Wearable technology Geo-positioning devices	Web Collection CASI CAPI	CATI / CAPI RDD Methods OCR / Scanning

Early technologies in the survey response area were CATI and CAPI (computer assisted telephone interviewing and personal interviewing respectively). In reference to this matrix, these technologies are tools utilized by interviewers. They automate the flow through the questionnaire and provide consistency checks on responses. Thus CATI and CAPII clearly belong in the bottom left cell, reserved for technologies that are tools for use by interviewers to enhance the quality of survey responses. By assisting with the flow through complex questionnaires, these tools may also make it easier for respondents to respond. Therefore, CATI and CAPI also appear in the top left cell. These important technologies simultaneously serve as tools for use by the survey designer and the survey administrator. For example, most CATI systems have a call scheduler who can be utilized by the survey administrator or field director to schedule or reschedule an interview at a convenient time for the respondent, thus easing the burden of responding.

As we look at this matrix, there are tools for interviewer use, respondent use, and for use by the survey designer or administrator. There are tools that help ease response and tools that enhance the quality of responses. However, there are gaps in the matrix cells showing technologies designed to help communicate the reason why a respondent should cooperate. Perhaps the single greatest factor in getting individuals to respond on surveys is to give them sufficient reason (from the respondent's perspective) to do so. It is apparent that our development of technology has generally ignored this important area. Thus this discussant highly recommends future efforts be directed along these lines.

I wish specifically to point out that one of the two papers in this session directly addresses a portion of this area. The paper authored by Link, Armsby, Hubal and Guinn discuss technology that enhances the survey administers' ability to train interviewers on how to avoid refusals during the first 30 seconds of a telephone interview. Thus, it is a tool for survey administrators that helps communicate to potential respondents why they should complete the survey. The other paper in this session looks at Web collection (a tool for respondents) and how that tool enhances the quality of response.

With that overview, we next look more closely at both papers. The Chang and Krosnick paper compares random digit dial (RDD) telephone surveys with two different approaches to Internet surveys. The paper provide results from both a field study and a laboratory study, utilizing surveys of individuals generally focused on political opinions and activities. The paper addresses two areas of concern: the representativeness of the responding sample and the quality of the responses supplied. This was an excellent paper. It is very useful for the survey community to have some work with actual measures of these qualities so that somewhat objective comparisons can be made, particularly concerning Internet surveys.

The paper begins with an excellent discussion of the potential differences between the three modes of collection, and is worth reading for this alone. To the long list of items discussed under response quality, one might add an item concerning the time intervals available for response. There may be a difference in how suitable the different modes are when responses are needed within a very tight time interval. Web collection provides the respondent with more flexibility in terms of when to respond, while a telephone calls pressures the respondent to "do it now". Krosnick indicates that this does not appear to be an issue for the types of surveys addressed in his paper. However, it may affect some Federal agencies' data collections. In trying to publish time sensitive reports, we in NASS often have a window of only a few days for survey response.

The authors provide a wealth of detailed results in comparing different components of quality. For example, they conclude that the Knowledge Network (KN) samples are comparable to the RDD samples in terms of demographic representation, and when weighted all three samples were fairly close. One interesting result concerns measure of non-differentiation. (*Respondent's lack of differentiation among questions with similar response scales is indicative of mindless responding patterns.*) KN respondents receive WebTV equipment in exchange for their participation in surveys, and they showed the greatest amount of non-differentiation. This raises the issue of the use of response incentives in a more general context. Do they lead to quality issues such as non-differentiation? Clearly more research in this area is needed.

Finally, I agree with the authors concluding comments that it is important to compare these different modes of collection in the framework of how they can be complements in future mixed mode designs.

The Link, Armsby, Hubal, and Guinn paper discusses the use of responsive virtual human technology (RVHT) as a tool for teaching refusal avoidance skills to telephone interviewers. The paper addresses some early work in this area. As mentioned earlier, this discussant compliments the authors for working on issues which help fill the "gap" in technology aimed at helping communicate the reasons to respond.

The technology simulates reactions of respondents during the first 30 seconds of a telephone call, and allows interviewers to practice their own reactions and responses. The RVHT tools allow more repetitions for lower cost of this important part of interviewer training. The authors report on early attempts which did not provide completely realistic simulations but which provide optimism for future progress. The simulations ran more slowly than reality, which caused some complaints by experienced interviewers. The real benefit of this type of training is likely to fall to inexperienced interviewers. Similarly the concern with the "slower than life" simulation may give new interviewers false expectations.

This discussant looks forward to following future work on this technology. It will be useful to see results using inexperienced interviewers in a split experiment which compares their subsequent response rates.

I thank the authors, organizer and audience for a very interesting session on technology.

Capitalizing on Technology to Enhance Survey Reporting Discussion

Alan R. Tupek
US Census Bureau

I would like to commend the authors for their thoughtful papers and their innovative work in advancing the use of technology in survey methods. We have come a long way in applying technology to survey methods. The innovative work presented in these papers illustrates just how far we have come.

I will organize my remarks into three broad themes --

I. Innovations, especially innovations that involve process automation, often lead to discoveries that weren't envisioned at the outset.

II. Since this is a conference sponsored by the Federal Committee on Statistical Methodology, what should the role of the Federal Government be as it relates to these projects?

III. Looking ahead. What might be on the horizon that's fueled by these endeavors?

I'll begin with surprises --

I. Innovations, especially innovations that involve process automation, often lead to discoveries that weren't envisioned at the outset.

CATI and CAPI did not reduce response error in surveys, as least as we traditionally measure it through reinterview surveys. In the CPS, CATI and CAPI had little effect on the month-in-sample differences. Computer-assisted methods did not make robots out of interviewers. What it did do was to change the complexity of the survey instruments so that concepts could be measured more precisely. CATI and CAPI also did not reduce costs, but provided the potential to improve collection methods and validation of results.

So what surprises might there be for the paper presented by Polly Armsby and Michael Link? The innovation that they describe is the use of a wide-range of technologies, they call "responsive virtual human technology," to train interviewers to handle the first few seconds of the interview process. While this is truly revolutionary in many respects, a significant side benefit is a learning process about refusal conversions that would not otherwise have occurred. The 48 scenarios discussed in the paper are just the start. Is there an optimum response for each of the 48 scenarios? Or, maybe there are a few preferred responses that work across many of the scenarios. It's telling that interviewers are reacting differently to different verbal cues, yet interviewers ignore the mood and sex of the virtual respondent. Should the interviewers pay more attention to these factors?

In the elections the other day, many voters had to deal with new voting equipment. Here in Montgomery County, we went from punch cards with hanging chads and having to make sure the punch card was properly seated in its place, to touch screen machines. The "improvement"

resulted in long lines, probably due to unfamiliarity with the new machines, plus fewer machines compared to the punch card system. Some machines in Montgomery County were misprogrammed. This happened with the new machines in Florida too. I'm sure no one is surprised to hear that. I can't help but think about the potential for programming errors with the touch screen machines. At least with the punch cards, you still have the cards to fall back on as we saw in Florida two years ago. Did the voting officials insist on level one programming? Was there independent programming of all systems?

Let's move on to Internet data collection. The Census Bureau has experimented with Internet data collection. It was available for the short-form in the decennial census. In the American Community Survey, we conducted a split panel test of Internet collection. The households selected for the Internet panel were given the option of responding by Internet but were also sent a questionnaire that they could mail back. The surprise in this test is that the combined mailback and Internet response rate was lower in the Internet panel than the mailback rate in the control panel. Not only did we only get a handful of responses by Internet – around 5% – but it actually reduced the mailback response rate significantly. There are a number of theories as to why this happened. And, we plan to investigate them in future tests of Internet data collection.

When we began developing one-stop shopping for federal statistics, we hadn't envisioned the research program that would evolve. We thought of FedStats as a portal, though we didn't call it that – we called it a window to statistics available on the federal agency websites. Now, through the leadership of Valerie Gregg and Marshall DeBerry, FedStats is a conduit to improving the statistical literacy of the nation.

Let me move on to my second topic.

II. Since this is a conference sponsored by the Federal Committee on Statistical Methodology, what should the role of the Federal Government be as it relates to these projects?

Does it make sense that the government is not at the table?
What should the role of the government be on the use of new technologies?
Should it fund the types of projects we've heard about today?

Regarding the paper presented by Polly Armsby and Michael Link, maybe the question should be "is the survey methods community rich enough to take this highly technical research that is being applied to refusal conversion from its infancy to maturity?" Or, should we let the Defense Department move the technical aspects of this research along to the point that the incremental research cost for the survey community is minimal? The Census Bureau was given the opportunity to participate in this endeavor and we decided to pass. I was one of the ones at the table. We understood the potential, but decided it was too risky. We didn't think the technology was there yet. Let some other research community move this along and when it's mature, we'll take advantage of it. I don't know the right answer. The Census Bureau needs to take risks. It needs to find the right balance between risky long-term projects, short-term research, and production activities. As we've heard from the authors, a lot of progress has been made in the

year or so since the Census Bureau was given the opportunity to participate. More progress than I would have expected. If I knew then what I know now, I might have voted differently.

The vital national surveys conducted by the Federal Government, like the Current Population Survey, the National Health Interview Survey, the Consumer Expenditure Survey, the National Crime Victimization Survey, and the American Housing Survey are not likely to embrace the Internet as the sole mode of data collection. Just as the Federal Government uses CATI as one mode in a multi-mode approach for several of these surveys, the Internet will also be used in this way. For similar reasons, the Census Bureau has not embraced RDD surveys except in special situations. The Census Bureau puts a great deal of resources in developing sampling frames that cover the entire target population. In addition, many of its household surveys achieve response rates of over 90 percent. If the Census Bureau did not set and achieve these goals for the vital national surveys, then where will the gold standard come from? How would the other survey organizations know how to weight their survey results?

It's heartening to know that representative samples do make a difference. In the paper by Jon Krosnick, the self-selected samples from the Harris Interactive Surveys provided skewed results even when weighted to detailed characteristics from the Current Population Survey.

III. Looking ahead. What might be on the horizon that's fueled by these endeavors?

I think there's something to be learned from the mandated Internet training instruments that seem to be springing up. For example, I am required to learn about the rules and procedures for using my government-issued credit card for travel. If I don't complete an online training course by such and such a date I must forfeit my credit card. You can skim the pages on the online course as quickly as you can move your mouse and click to the next screen. However, in the end, there's a quiz that requires you to answer most of the questions correctly. Otherwise, you don't pass and you must repeat the exercise.

The Knowledge Network practice of providing a WebTV box and service is something to build on. I'd call this a creative use of incentives. Unfortunately, it's costly, but the good thing is that the cost is spread over a lot of surveys.

There's no end to what one might imagine as uses for the Responsive Virtual Human Technology (RVHT). The training possibilities are endless. But, how can it be used in the interviewing process itself, especially over the web? How might it help with language difficulties, illiteracy, or persons with disabilities? Maybe RVHT can be a virtual boss? "What's the status of that project," delivered in either a sad, had, glad, or mad tone of voice. The human boss needs only to follow-up on the sheepish responses.

In conclusion, I want to thank the presenters and their co-authors for their groundbreaking work and for a stimulating discussion on improving survey methods

Session 9

Providing Adequate Technical Support

Confessions Of A Survey Guy

Stephen Dienstfrey
Schulman, Ronca, & Bucuvalas, Inc.

I'm a survey guy. I learned how do to survey's during my graduate studies in Sociology. It seemed I had a knack for statistics and methodology. A professor asked me to help him with a study he was doing of residents of public housing. This led to a job with a political polling firm, to a short stint at the Bureau of the Census working on a survey of Inmates of Local Jails, to AMTRAK doing consumer research, to the Energy Information Administration doing a survey of Industrial Energy Consumption, to the Department of Veterans Affairs where I was responsible for surveys of the veteran populations, and for the last 5 ½ years I have been back in the private sector.

I just covered 35 years of my professional life – and until I worked for VA documentation was never a problem: either no one ever asked me a question about a data set I put together, or I wasn't there long enough for people to ask me a question.

Let me step back a minute. I know meta data is more than documentation about data files – it's archival material, it's bibliographic material, it is a whole host of other types of data about data in a format that can be searched using common terms and I know that there are a lot of people that are working on developing standards to standardize meta data formats. But I'm a survey guy, and to me it is documentation.

(And when did 'documentation' become 'meta data'? I can't decide if this is like doctors calling 'anti-histamines' 'histamine antagonists' or religious leaders calling 'sin' 'transgression'.)

So while I know we are talking about meta data, my experience is with documentation and I am going to confine my remarks to that specific area.

My first contact with the vagaries of documentation came while I was at VA doing some work with the 1980 Census Public Use Microdata files. VA had ordered some special tabulations of veterans from the Bureau. These tabulations were used as the basis for the re-calibration of the Veterans Population Model. The results of this recalibration had been distributed throughout VA and to states. However, the counts we ran from the PUMS tapes did not match the counts from the special tabs. After much trial and error we realized that the definition of "veteran" in the PUMS included individuals who were still on active duty. These individuals could easily be removed by using the current employment questions where those on active-duty were identified. This information may have been available in the PUMS documentation, but it seemed a little intimidating and no one bothered to read it.

One would have thought that this experience would have instructed me on the need for adequate documentation. But, alas, this was not the case.

The 1980 Census was to first time that women were asked if they had served in the military. This important group of veterans was further studied in the 1987 Survey of Female Veterans.

Prior to releasing copies of the data we prepared SAS and SPSS data files where all the variables and values were labeled. So much for documentation.

As the file was used outside VA people would call with questions. While everything was relatively fresh in my mind, it was not a problem. But as time went on and the calls became less frequent, it became more and more time consuming to run down the answers. At no time during this experience did I bother to keep notes about what was asked and what the answers were.

One would have thought that I learned from this experience, but a similar pattern took place with other VA studies.

More recently I have been more of a user of public data sets than a provider. We have downloaded files from various health related surveys as well as from the Bureau of the Census. Generally, the documentation is both voluminous and written for those who are initiated in the theology of the survey not those of us who are novices. Specifically, the variables have names that sometimes defy understanding, and computed variables are identified but the computations were not there. One data set had 6 different weights with the instructions to use the one that was most appropriate without giving the user who may not understand the subtle difference between "household" and "individual' weights.

So if we can agree that there is a need for complete and understandable documentation, how do we go about making sure that it happens? We could give the task to OMB, and while that might work for data produced by federal agencies and the key word here is "might", I think it would be unwise to create yet another onerous bureaucratic requirement. Further, this would do nothing for data that was created outside of the federal umbrella.

Another possibility is work on people's guilt. But this hasn't worked in the past and I can see of no reason for it to work in the future.

It seems to me that if there is to be adequate documentation it needs to be a transparent by-product of the entire survey process from start to finish.

Using my current job as a model, let's walk through the steps that take place when we conduct a study. From my experience in a variety of jobs, I think these are fairly standard.

- Develop the questionnaire.
- Program the questionnaires for Computer Assisted Telephone Interviewing (CATI) with range checks, automatic skips, internal codes, etc.
- Edit the data – code open ends, collapse data, etc. – to final data set.
- Create analytical variables by combining and/or recoding data from the previous steps during the reporting phase.

What I suggest is a software suite that takes the final questionnaire (along with any notes about specific questions or series of questions that can be carried along with the document) and feed that into a CATI program that is very close to final. As the CATI program is modified, notes are automatically kept on what was changed. As the data are edited, more notes are kept. The codes

used for open-ends are also attached. All of this is carried along to the data set similar to the information SPSS displays with the variable tab. The syntax used to create an analytical variable is also attached to the data set.

The result may not be the complete documentation that archivists might like, but it is a very good first step. More importantly, it might help the casual data user, such as myself, to get a good idea of what is going on. And when I have to call and ask for further help I sound less like a complete idiot.

But what do I know; I'm a survey guy.

Training and Staff at the U.S. Census Bureau to Create Metadata and to Provide Consultation with Users

Pat Doyle
U.S. Census Bureau[1]

There is widespread corporate and project level support for documentation and user assistance across the Economic, Decennial, and Demographic Directorates of the Census Bureau. This paper focuses on the programs in the Demographic Directorate. It covers the formal and informal procedures for training staff to provide documentation and technical assistance, guidelines for preparing documentation for demographic programs, corporate level support for documentation preparation and tools to support documentation efforts, organizational support for staff to provide technical assistance, and methods of obtaining technical assistance. The paper concludes with a discussion of issues related to survey documentation and technical assistance and some thoughts for the future.

Staff Training

The Census Bureau strongly supports staff development and encourages staff to provide technical assistance. The formal training program consists of a 55-hour classroom training program spread over six weeks, which is targeted to new staff in the Demographic Surveys Division but available to all interested staff. This program covers all aspects of the design, development, administration, processing, and documentation of demographic surveys. Its intent is to give survey staff specializing in one area of the survey process an overview of the other aspects of the survey process and to introduce them to the various Census Bureau organizational units that play a role in the survey. This gives survey staff a broad perspective on the Census Bureau and on the range of expertise employed in conducting surveys. This will enable staff to effectively route specific questions from users and customers that require expertise in a particular part of the survey or its design and execution.

As part of this demographic survey training program, a three-hour session is devoted to documentation and to the importance of documentation to the wise use of Census Bureau data products and to the continued support for Census Bureau surveys from other agencies. This training program offers examples of how secondary data users and funding agencies rely on data documentation to carry out their work and illustrates both good and bad examples of documentation. It is an interactive, hands-on class intended to motivate the Census Bureau staff not only to develop good documentation but to provide users with methods of access to that information that is helpful and efficient.

The demographic directorate actively participates in the Joint Program in Survey Methodology (JPSM program) offered by the Universities of Maryland and Michigan and Westat. The Census Bureau cosponsor this program along with other federal agencies, and the demographic program encourages its staff to take advantage of the various seminars and classes offered through that

[1]This paper reports the results of research undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope to that given to official Census Bureau publications, and is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

program. A number of staff have received masters degrees in survey methodology through that program with sponsorship from the Demographic Directorate, and many more have taken (and continue to take) courses and attend seminars.

Guidelines

The demographic directorate developed guidelines for documenting all its surveys, to govern the type and extent of documentation needed for the data products. These guidelines encourage the staff to prepare documentation at three-levels: summary, user, and technical. The summary level material is intended to provide a quick and easily accessible overview of the survey and its data products. The user-level information is a comprehensive set of documentation for individuals using the data products, written in a style that is informative but not overly technical and thus does not require the user to be an expert statistician or methodologist to comprehend its content. The technical information is the most comprehensive description of the survey and its processes but can be fairly dense and difficult for a novice to follow. It is devoted to individuals with a need to know on a particular point and to experts in various components of the survey process.

The general guidelines cover all aspects of the survey design, development, and execution of the survey. They focus on producing documentation for all users, both internal and external to the Census Bureau. They cover content of the data collection instruments as well as data products, procedures for data collection and data processing, and survey and data quality.

One specific product included in the overall documentation guidelines is the quality profile. The demographic area has separate guidelines devoted explicitly to the organization and routine production of quality profiles for demographic surveys. The emphasis in these guidelines is to focus the reader's attention on the relative quality of Census Bureau surveys within the survey industry, emphasizing the high response rates and other high quality features. The guidelines also stretch the scope of quality profiles to touch on areas not routinely included, such as the outcome of instrument pretesting and the field procedures used to minimize nonresponse. These guidelines (along with similar ones developed for the Economic Directorate) are in the pipeline to become formal standards.

Corporate Support and Tools

Traditionally, survey documentation has been prepared by technical staff who are not always skilled in the fine art of writing for a broad audience (like users of public use files). With the new documentation guidelines, the Demographic Surveys Division provides a technical writer to assist the projects in refining the presentation of technical materials, so that they meet the needs of secondary data users. Thus, we encourage staff to use their technical staff to write the technical-level documentation and to use the technical writer to prepare the summary-level material.

The Census Bureau recognizes that production of good documentation can be handled most efficiently if there are good support tools. In fact the agency has sponsored and continues to sponsor a number of documentation tools.

One of those is the instrument documentation system (IDOC), the first attempt to address the serious gap in the automation of complex survey questionnaires, which was (and is) the lack of interpretable instrument documentation. IDOC grew out of an effort by the Association of Public Data Users to focus attention on the impact this lack of documentation would have on intelligent use of the data. Its concept was broad, in that it used the notion of generating documentation directly from the instrument: storing it in a neutral format that could then be converted to a format like HTML, so that all users with a web browser could access it and use all the features of a web browser to search and excerpt from the documentation. The system includes hypertext links to facilitate following all the possible paths an instrument might take.

Another tool is the survey development and documentation system (SuDDS), for the development and documentation of instruments. This system relies on a Microsoft Access database system to house question- or item-level documentation and specification; and an interface to facilitate development of question text, fills, flows, universes and response categories. This will enable developers to more easily specify and develop automated instruments, track their evolution over time through the development, testing, and refining process, and generate instrument documentation in the form of items booklets. It will also enable the Census Bureau to integrate results from the survey (such as the frequency of responses to a particular item by type of response) with the documentation. Finally, its long-term goal is to support the specification of the post-collection processing system within the same system that documents the instrument.

Another tool supported by the Demographic Directorate is Data FERRETT, a metadata driven data access system and a component of the American Fact Finder (AFF). It is a data access tool that encompasses and relies on item-level documentation to make it easier for analysts to gain access to survey microdata either through quick small tables or big extracts. Data FERRETT is distinguished from the decennial census component of AFF by its emphasis on survey rather than Census data and by its facility to support preparation of extracts with accompanying data file documentation.

Finally, the Demographic Directorate actively supports the Data Documentation Initiative (DDI), an international effort to establish standards for formatting data documentation to enable data archives and the like to establish generic documentation access systems applicable to all of their holdings. The DDI standard exists for both microdata files and for tabular data stored in machine-readable form. As such, it promises to significantly reduce the cost and resources required to share increasing complex information derived from surveys.

Technical Assistance

The Census Bureau encourages staff to give user questions high priority. Contact information for appropriate staff is widely-available. The staff are organized in different ways, depending on what phase of the survey process or demographic program they support: survey-specific experts, subject area experts, and methodologists. Each group has a different set of skills and knowledge, a different organizational structure, and slightly different approach to training. However, all groups are required to provide assistance to those who need it.

The survey-specific experts make up the operational staff that manages surveys, supports their administration, and guides the production of files and documentation. This group includes survey statisticians and programmers who are organized by survey and, within survey, by function (instrument development and administration versus data processing). Technical questions on file development, access, format and documentation are often fielded by staff from this group.

Training of this staff is a mixture of classroom training (the 55-hour training program developed and administered by the operational division for demographic surveys), courses like those at JPSM, and on-the-job training. They take phone calls, emails, and other inquiries from users; answer questions they can; and refer other questions to those who can provide answers.

For Census-sponsored surveys (and some reimbursable surveys) and for other demographic programs, the Census Bureau has subject area experts who develop the survey or program content and scope, recommend questions to be administered, review the data, develop the approach to post collection processing, and produce Census Bureau reports and technical papers in their area of expertise. Individuals in these groups are content specialists, typically with degrees in one of the social sciences, and their responsibilities span multiple surveys. Questions about specific variables and how to analyze specific concepts are referred to this group. They are organized first by subject area into two divisions (one focused on housing and household economic information and the other on demographic information) and by content (rather than by survey) within those divisions. Their work extends beyond demographic surveys to include the decennial census and its related studies. As with the operational staff, the subject area staff take phone calls, emails, and other inquiries from users; answer questions they can; and refer other questions to those who can provide answers.

Training opportunities for this staff include the 55-hour training program developed for the operational staff, as well as on-the-job training, Census Bureau seminars (by Census Bureau and non-Census Bureau researchers), and participation in external conferences and seminars—both as presenters and attendees.

To support state-of-the-art survey and sample design and analytic techniques, the Census Bureau has a number of methodologists specializing in a wide array of technical topics—including sample design, questionnaire design, and research methods. This group tends to be organized by the type of methodological work and by survey, but this is not exclusively the case.

Staff in this group provide assistance within the Census Bureau on the proper interpretation of statistical information in analysis and assist users in understanding the technical design features of the sample. As with the operational staff, they take phone calls, emails, and other inquiries from users; answer questions they can; and refer other questions to those who can provide answers. Some (particularly the sample design group) are actively involved in the preparation of documentation on surveys, routinely preparing the source and accuracy statements and periodically preparing quality profiles.

This staff participates in the 55-hour on the job training program, actively participates in the JPSM program, attends seminars and conferences, and presents technical papers on the procedures used by the Census Bureau.

Access to Technical Assistance

Users can get their questions answered by phone, through list serves, directly by E-mail or regular mail, through the internet, or through bulletin boards. There are also a variety of user groups, working groups, and advisory committees through which information can be obtained. Examples of contact information follow:

By Email:

Pop@census.gov (for domestic demographic questions).
Hhes-info@census.gov (for household and household economic questions).
Ipc@census.gov (for international demographic questions).
Cpshelp@info.census.gov (for questions on the Current Population Surveys).

Web Sites:

www.census.gov (general Census Bureau home page).
www.sipp.census.gov/sipp (information on the Survey of Income and Program Participation (SIPP)).
www.sipp.census.gov/spd/ (information on the Survey of Program Dynamics (SPD)).
www.census.gov/hhes/www/ahs.html (information on the American Housing Surveys).

List serves:

ACS Alert (for information on the American Community Survey (ACS)).

There are information groups within the subject area divisions whose primary job is to field questions on the demographic surveys (and other programs). They answer routine questions directly when they can and refer more technical questions to the appropriate expert. There are also some survey-specific groups within the Census Bureau to provide direct support to users of the survey. For example, the outreach and analysis branch within the Housing and Household Economic Statistics Division provides support for SIPP and SPD, and an area within the Demographic Surveys Division that advocates for the ACS, reaching out to users to determine their needs

The Demographic Directorate promotes and participates in the Census Bureau's user conferences which are heavily focused on the Decennial Census and related programs that produce a large amount of tabular information. General information about demographic surveys is included in these conferences and provided by the staff from the demographic directorate.

The directorate is also exploring the possibility of establishing user conferences for users of micodata files, which would be heavily focused on demographic surveys. The project is in the concept stage at the moment, and approval and implementation will depend on costs, resource requirements, and expected levels of participation.

Issues

Federal agencies in general are facing a number of difficult issues with documentation of demographic surveys, and the Census Bureau plans to continue its lead role in attacking these. The first issue, alluded to earlier, is the loss of good instrument documentation with automated surveys. Generally, a small or linear character-based automated instrument can be documented manually by compiling the image of the screens and describing the navigation possibilities. However, this is too resource-intensive for complex instruments and nearly impossible for instruments designed to take full advantage of graphical user interfaces (such as the Windows operating system). Two attempts have been made so far to address this problem (one of which is the IDOC system noted above), but neither of these has reached a level of maturity that yields an adequate substitute for the paper instrument that existed for surveys which were not automated.

A related issue is that the industry is not actively tackling the problem of the loss of good instrument documentation, leaving each survey project to fend for itself to compensate. This yields varying quality instrument documentation that is often incomplete in its explanation of all the possible wording variations and sequences of questions. The National Academy of Sciences has recently called attention to the problem (through a workshop sponsored by the Demographic Directorate and subsequent seminars) and encourages computer specialists to join the survey industry in tackling it but, as yet, no substantial source of funding has arisen to support these efforts.

The ever present issue is that there is never enough documentation for everyone, and there are never enough resources to do good comprehensive documentation—particularly for large complex surveys. Any pressure to save resources will affect documentation first, because the theory is that you can't document what you don't collect and we never have enough data to answer all of society's important questions. This is somewhat inconsistent with our goals, however, in that it is difficult to insist on quality uses of quality data if the documentation is inadequate. But that is a weak argument when there isn't enough money to collect all the data needed in the first place. It's also misguided in the sense that it is really an inefficient shift in the costs of using the data that are collected. If the data collector does not adequately document the data, the data user will have to do it him or herself. That is an inefficient use of resources, unless there is only one user and, hence, only one occurrence of the documentation effort.

To proceed to produce good documentation on a routine basis at minimal cost within the current budget environment requires a lifestyle change, accompanied by strong management to enforce that change. The lifestyle change is to view documentation as an integral part of the survey, that should be generated alongside the other survey processes rather than saved until the end. It is also to recognize that preparation of the documentation is most efficiently accomplished by a mix of technical staff and technical writers, rather than exclusively by one or the other. Finally, it is a recognition that developing and conforming to standards for documentation can support a more efficient and cost-effective documentation production system.

Another issue which may be unique to the Census Bureau is that we tend to have a different set of data accessible to staff than to external users. This is largely due to the disclosure avoidance measures applied to data files before they can be released. However, in some instances it is due

to a preference for using a more efficient file format for the data stored internally than for data distributed to a broad set of users. The different format and content make it difficult for the staff to answer technical questions that the users generate directly from observed anomalies in the data. The Demographic Directorate moved away from the use of the differently-formatted files, but it is unlikely to restrict its use to the disclosure-proved data because of the analytic restrictions imposed by the limited geography that is required for the disclosure avoidance.

One issue implicit in this paper is that the training provided to our staff does not include on-the-job training in the type of people skills that would make the technical assistance process go more smoothly. We hire good people and expect they can convey answers to technical questions in a manner that users can appreciate and understand. For the most part, we are lucky and this works.

The final issue to note is that expertise tends to reside with individuals rather than groups, so that the absence of one individual can create a void in the information available for users. More cross-training of topics and cross-fertilization of staff would be very helpful. The Census Bureau has begun to recognize the importance of this cross-fertilization in terms of staff development and succession planning, and any efforts that come to pass in that vein should benefit efforts to provide technical assistance.

The Future

The demand for documentation and technical assistance will continue to grow with the increasing complexity and sophistication of surveys and data collection programs. As a result, the federal statistical system (in general) and the Census Bureau (in particular) need to address the issues just noted. For its part, the Demographic Directorate intends to continue to support the development of documentation tools, continue to participate in the development of documentation standards, and hopefully will be able to implement a series of user conferences to assist users in the wise use of our products and to share information between us and the user community on uses of the data.

Session 10

Providing Small Area Estimates

Small Domain Estimation for the U.S. Current Employment Statistics Program: Management Implications of Multiple Stakeholders and Multiple Constraints

John L. Eltinge[29]
Bureau of Labor Statistics

1. Introduction

The Current Employment Statistics (CES) survey is a large-scale monthly establishment survey conducted for the Bureau of Labor Statistics through a joint federal-state program. This survey currently is completing a transition to a probability sample design that is intended to produce relatively precise estimators of monthly total employment for each state in the U.S. For some general background on the CES and related technical and policy issues, see, e.g., American Statistical Association (1994), Butani, Harter, and Wolter (1997), Butani, Stamas and Brick (1997), West et al. (1997) and Werking (1997).

Although the CES sample design and estimation work focused primarily on production of estimates at the state and national levels, many stakeholders have strong interest in estimation for considerably smaller domains, e.g., for a specified major industrial division within a given metropolitan statistical area. Consequently, the Bureau of Labor Statistics is developing model-based small domain estimation methods for the CES. In general, small domain estimation involves very rich set of technical, management and policy issues. For some general background, see, e.g., Fay and Herriot (1979), Platek et al. (1987), Ghosh and Rao (1994), Schaible (1996), Shen and Louis (1998), Singh et al. (1998), Schirm et al. (1999), Marker (1999, 2001), National Research Council (2000 a, b, 2001), Rao (2002) and references cited therein.

In keeping with the theme of this year's FCSM Statistical Policy Conference, "Challenges to the Federal Statistical System in Fostering Access to Federal Statistics," I will focus this presentation on some policy and management issues that appear to have a substantial effect on the degree and nature of many stakeholders' use of small area estimates. Specifically, I will suggest that for some small domain estimation programs, the presence of multiple stakeholders and multiple constraints can have a substantial effect on the development and implementation of our small area estimation methods, and on the best strategies for communication of small domain information to various groups of stakeholders. In addition, I will suggest that it can be useful to view small domain estimation methodology as a form of technology; and that the development and implementation of small domain programs may benefit from previous literature on the adoption and diffusion of technology.

2. Multiple Stakeholders and Multiple Utility Functions Small Domain Programs

Some small domain estimation programs have been developed primarily for a single relatively well-identified purpose. A prominent example of this is the he Small Area Income and Poverty Estimates, or SAIPE, program of the U.S. Census Bureau and the U.S. Department of Education) which is focused primarily on funding allocation formulas, and for which other uses

[29] Office of Survey Methods Research, PSB 1950, Bureau of Labor Statistics, 2 Massachusetts Avenue NE, Washington, DC 20212. Eltinge_J@bls.gov

are somewhat secondary. On the other hand, other small domain estimation programs, including the one currently under development for the Current Employment Statistics program, have been motivated by a wide range of potential uses, without a single dominant user, and without a corresponding dominant consensus regarding priorities among estimands and the objective functions appropriate for evaluation of estimator performance.

Instead, extensive discussions with states and other stakeholders indicated a wide range of views regarding, e.g., the relative importance of estimates of total employment, one-month change, three-month change and twelve-month change; and an equally wide range on the relative importance assigned to relative bias, absolute bias, relative variance and absolute variance. In addition, some stakeholders use published small domain estimates as one of several sources of information regarding local economic conditions, and do not use these estimates in ways that would lend themselves to a solid characterization of, e.g., the relative benefits of reducing bias or reducing variance.

Finally, we have also encountered some data users who focus heavily on published CES small domain estimates, but do so in a form that may be described as informal simultaneous inference. For example, a metropolitan area analyst may examine time plots of employment estimates for the past three to eighteen months for a given metropolitan area-industry combination, and then attempt to make statements like, "employment is approximately constant," "employment has an upward trend" or "employment growth is very similar to what we are seeing in the rest of the state." This suggests that it may be worthwhile for the BLS and other statistical agencies to consider two areas for additional work. First, the small domain literature has tended to focus a substantial amount of attention on estimation of variances or mean squared errors. Given the abovementioned attempts at informal inference, it may be appropriate to develop related procedures for simultaneous confidence sets for linear functions of a vector of monthly estimands; and to develop related user-friendly graphical tools. Second, some of the abovementioned informal analyses appear to involve a large number of implicit hypothesis tests. Consequently, it would be worthwhile to explore the extent to which we can help the analysts incorporate appropriate measures of uncertainty into such tests; and to explore the extent to which one may use "false discovery rates" (e.g., Benjamini and Hochberg, 1995; DuMouchel, 1999; Genovese and Wasserman, 2002; and Storey, 2002) and related tools to provide appropriate quantitative measures for the risks incurred in this type of multiple testing.

3. Multiple Constraints in Small Domain Programs

The literature on statistical policy often notes – either explicitly or implicitly – that practical constraints can have a dominant role in the development, implementation and perceived value of a given statistical program. See, e.g., Bonnen (1988), Eckler (1972), Felligi (1996), Kaysen et al. (1969), Lehnen (1988), Moser (1976), O'Hare and Pollard (1998), Parke et al. (1976), Reynolds (1988), Rosenberg and Myers (1977), Shiskin (1970) and Weiner (1974). Similarly, the current development of the CES small domain estimation program has been heavily influenced by practical constraints on, e.g., the relatively brief time between data collection and publication deadlines, the timely availability of specific forms of microdata, a preference for compatibility with previously implemented methodology for national- and state-level estimation, and compatibility with legacy production systems. In many cases, a full-scale quantitative

characterization of these constraints is not readily available, and work toward such a quantitative characterization can itself involve substantial costs and cognitive burden.

In contrast with this, the mathematical statistics literature in survey sampling tends to focus on optimization of an objective function (e.g., variance, mean squared error, or a pseudolikelihood expression) in the presence of a relatively complex stochastic structure induced by a complex sample design or a hierarchical model. In such work, constraints often are viewed as being relatively mild or otherwise of somewhat secondary interest.

Thus, in a qualitative sense, our optimization work follows the pattern displayed in Figure 1. (For some related discussion and partial exceptions to this, see, e.g., Ahsan and Khan, 1982; Cochran, 1977, Section 5A.3; Grzesiak and Johnson, 1989; Hansen et al., 1983; Harris, 1972; Kish, 1976; Neumann, 1999; and Renner, 1976.) On the other hand, the presence of substantial operational constraints, and the components of uncertainty associated with some of these constraints, could be characterized in a schematic form by Figure 2. Similarly, given the presence of many stakeholders with distinct utility functions, one could extend Figure 2 to include multiple objective-function curves.

4. Implications for Managers and Mathematical Statisticians

Comparison and contrast of the ideas in Sections 2 and 3 lead to some suggestions regarding efficient development of small domain estimation programs. First, for small domain work that is not dominated by a single objective (e.g., funding allocation formulas), some of the traditional optimization insights offered by mathematical statistics (as in Section 3) may be dominated by the presence of multiple utility functions, multiple constraints, and limitations of the quantification of these utility functions and constraints.

Second, this domination result is fundamentally an opportunity, rather than a barrier. From a pure research point of view, this provides a very rich set of mathematical statistics problems that, as we have seen, can be of very serious practical interest. In addition, this expanded set of problems appears to have substantial connections with some unresolved traditional problems in the analysis of survey data, e.g., issues raised by Hansen et al. (1983). This in turn potentially enhances the contribution that mathematical statistics can make to operational aspects of small domain programs. For example, comparison of Figures 1 and 2 suggests that one carry out a form of triage to identify specific constraints that may have the largest relative effect on the performance of the estimation program.

Third, this also leads to suggestions regarding efficient management of mathematical statistics research projects related to small domain estimation. Within agencies, mathematical statisticians constitute a relatively rare resource. Consequently, it is of interest to focus that resource on a moderate number of high-priority research areas that are most likely to lead to substantial improvements in agency production work. The ideas of Sections 2 and 3 suggest that for problems involving multiple utility functions and multiple constraints, one focus on cases involving sufficient common methodological structure and quantitative structure to lead to substantial improvements in perceived utility functions.

5. Access to Small Domain Estimates, and the Adoption and Diffusion of Related Innovations

Now let's return to the theme of this FCSM Policy Seminar, "Challenges to the Federal Statistical System in Fostering Access to Federal Statistics." For cases in which small domain estimates are used primarily for funding allocation formulas, it is possible that an agency may reasonably focus of development of appropriate methods, review of these methods by the National Academy of Sciences and other responsible outside scientific groups, and provision of public access to the estimates and associated measures of uncertainty. On the other hand, for a small domain program that does not have a single dominant purpose (e.g., the CES small domain program), it may not suffice to focus on relatively passive forms of access as such. In such a case, the practical value for multiple stakeholders may depend heavily on the extent (possibly limited) to which we can convey to these stakeholders a relatively refined sense of the information that is, and is not, conveyed by a given set of small domain estimates.

In thinking about programmatic ways in which we can address this need, it is useful to think about small domain estimation work as a technology, and to examine the extent to which we may obtain some management and statistical policy insights form previous studies of the ways in which multiple stakeholders explore and make decisions about a relatively new technology. These studies generally fall under the rubric of "adoption and diffusion of technology," and have arisen in several disciplines, including rural sociology, software engineering, military science, communications and marketing. A prominent early study by Ryan and Gross (1943) considered the processes by which groups of farmers in Iowa adopted hybrid seed corn. For detailed exposition and critique of the general literature on technology adoption and diffusion, see Rogers (1995) and references cited therein.

My initial reading of some of this literature led to several points of potential value for small domain work. Today, I'll emphasize five of these points, taken primarily from Rogers (1995). First, Rogers (1995, p. 11) defines an innovation broadly as, "an idea, practice or object that is perceived as new by an individual or other unit…." Thus, tables of small domain estimates, or new methods for the production and interpretation of these estimates would fit readily under this definition.

Second, Rogers (1995, p. 10) defines diffusion as a "process by which an innovation is communicated through certain channels over time among members of a social system." Note especially that this definition requires us to have a fairly concrete vision of the "members of the social system" who may use our estimates, and the channels through which we will communicate with them. This starts to push us beyond relatively passive notions of "access" and toward more active engagement with specific subsets of our multiple stakeholders.

Third, Rogers (1995, pp. 15-16) emphasizes five characteristics of innovations that he considers important in analysis of adoption and diffusion processes: relative advantage, compatibility, complexity, trialability and observability. Each of these clearly is applicable to small domain work, and may help us explain some of the responses of our stakeholders to new small domain estimation programs. For example, in keeping with the comments in Section 3, perceptions of relative advantage may be dominated initially by the ability to produce estimates for

subpopulations that were not previously available. Subsequently, these perceptions may be influenced by observation of both Type I and Type II error rates encountered in either formal or informal inferences drawn from reported small domain estimates and associated measures of uncertainty.

Fourth, Rogers (1995, pp. 28-30) distinguishes among three types of innovation decisions: optional, collective and authority-based. For example, competition in a free market among several comparable technologies may result in optional innovation decisisons. On the other hand, an innovation decision driven by imposition of standards by a voluntary industry trade group would involve a mixture of collective and authority-based decisions. From the standpoint of states and other stakeholders, development and implementation of the CES small domain program may be viewed as largely authority-based, but extensive consultation with the Current Employment Statistics Policy Council added a component of collective decision-making to the process.

Fifth, Rogers (1995) and other authors often describe the process of adoption and diffusion with schematic diagrams like the one displayed in Figure 3. Along the horizontal axis, we have the elapsed time to adoption of a given innovation. The literature partitions the population of potential users into subpopulations according to their anticipated time to adoption. The resulting diagram uses a Gaussian density curve to indicate the approximate sizes of the subpopulations.

I would view the Gaussian approximation and specific subpopulation cutoffs with a considerable amount of caution, but I believe that in a qualitative sense, this schematic device can be useful. For example, on the left side of the time scale is a group labeled "innovators," who are directly involved with research, development and very early use of a technology. For small domain estimation, this group would tend to include mathematical statisticians in academic institutions and in government statistical research and methods groups. Note especially that the values, technical sophistication and organizational dynamics within the "innovators" group tend to be quite different from those in the other groups. For instance, the next group, labeled "early adopters," tend to work outside of the initial research and development environment, but potentially have an active interest in the development effort. For small domain work, this might include stakeholders directly involved with funding allocation formula work or with state Current Employment Statistics program offices.

The general suggestion from Figure 3 is that in the course of time, additional groups (labeled "early majority," "late majority" and "very late or never") may also begin to make use of a technology. Adoption decisions by these later groups often may be attributed to a combination of communication of this technology to members of these groups, and to the maturation of the technology as such. In that process, two important factors are the observable reward/risk ratio and the degree of standardization of the technology. The innovators and early adopters may be intensely interested in the technology because they anticipate that they be able to obtain a substantial (though perhaps rather uncertain) perceived reward in exchange for a given amount of investment risk. In contrast with this, the later groups may expect a higher degree of predictability in their observable reward/risk ratio, and would expect some degree of assurance that this ratio will exceed a reasonable threshold. In parallel with this, innovators and early

adopters may tend to use highly customized technologies, while the later groups may expect a substantially higher degree of standardization.

As with any schematic description or generalization of human and organizational behavior, one should view the adoption and diffusion literature with a reasonable degree of caution. For example, some of this literature assumes a bit too readily that one should adopt the technology in question, and may not place enough emphasis on observable risk/reward ratios. Nonetheless, I believe that this literature does offer us some useful insights into management of small domain projects that do not involve a single dominant stakeholder.

First, as an agency and its stakeholders work with a large number of requests for small domain estimates, schematic diagrams like Figure 3 can help one set priorities and gain some perspective on the current state of development and adoption of a given set of small domain estimation methods. For example, if a program is in a relatively early stage of development, it may be best to focus efforts on requests from stakeholders with a realistic chance of obtaining fairly high levels of observable reward, relative to risk, and who have a relatively high tolerance for associated risks. Within this context, note that the literature on adoption and diffusion of technology often reports a sharp distinction, or "gap," between the "early adopter" group and subsequent groups. Consequently, many technologies that in principle could be used by a relatively large group do not, in practice, go beyond the "early adopter" stage.

Second, for small domain estimation, a serious assessment of relative rewards and risks tends to require balanced consideration of quantitative and qualitative components. Quantitative components include multiple components of uncertainty, e.g., sampling error, nonsampling error, model equation error and model misspecification effects. Somewhat more qualitative terms involve trade-offs among the costs of non-publication, relative costs of Type I and Type II errors, and the latter factors are often complicated by the presence of implicit multiple testing. Evaluation of these risks may require a relatively high level of customization, and thus, may require allocation of a relatively high level of resources by both the agency and by some data users. Within agencies, attempts to balance resource requirements against the abovementioned reward/risk calculations may be complicated by the fact that agencies generally do not have the same market mechanisms that are observed in some other areas of technology adoption and diffusion.

Acknowledgements

The author thanks the following persons for many helpful discussions of the Current Employment Statistics Program and small domain estimation: David Binder, Shail Butani, Moon Jung Cho, Cathryn Dippo, Wayne Fuller, Jack Galvin, Patricia Getz, Julie Gershunskaya, Rachel Harter, Steve Haslett, Larry Huff, Laurie Kujawa, Partha Lahiri, Kirk Muller, Lois Orr, Ken Shipp, Clyde Tucker, Kirk Wolter, Alan Zaslavsky and the members of the Current Employment Statistics Policy Council. In addition, the author thanks Fred Conrad, Norbert Schwarz and Roger Torangeau for comments on prospect theory; and Moon Jung Cho for preparing Figures 1 through 3. The opinions expressed in this paper are those of the author and do not necessarily represent the policies of the U.S. Bureau of Labor Statistics.

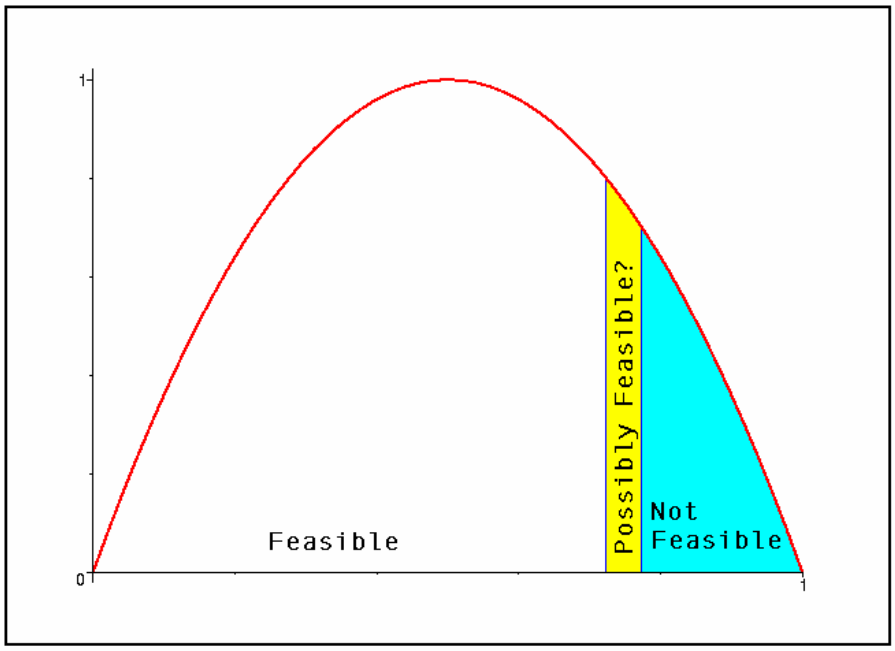


Figure 1: A Case in Which Constraints and Related Uncertainties Do Not Dominate Optimization

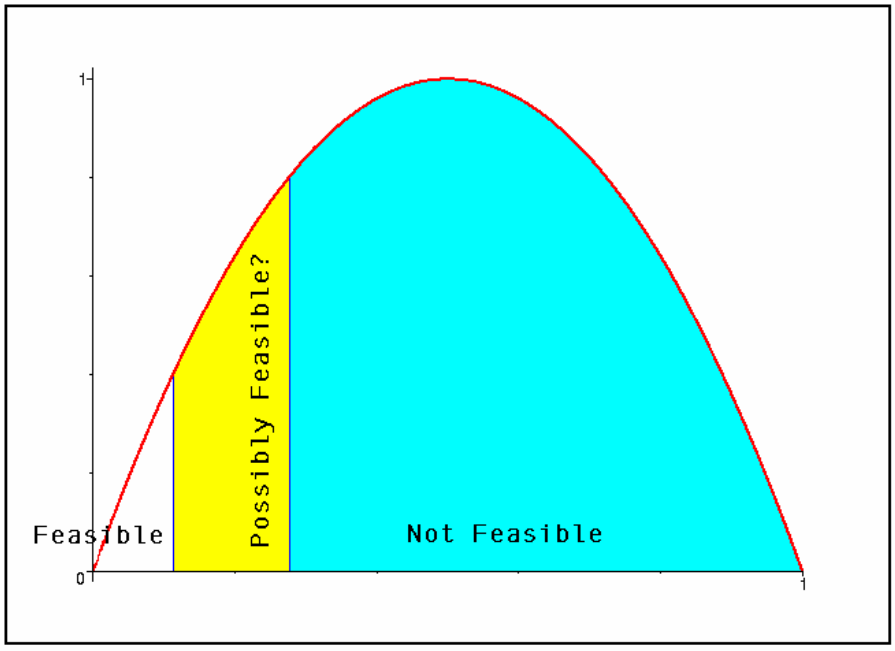


Figure 2: A Case in Which Constraints and Related Uncertainties Dominate Optimization

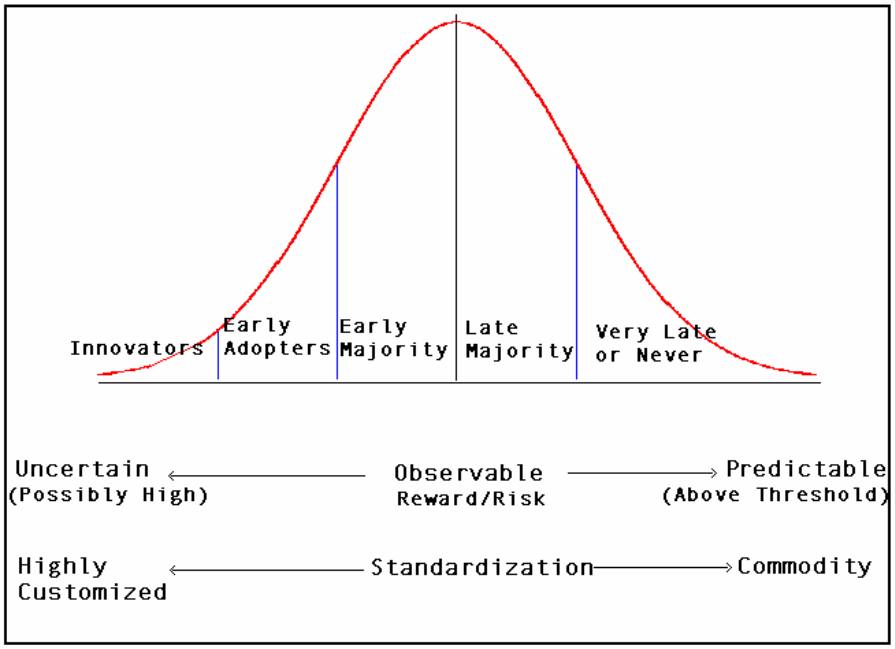


Figure 3: Customary Schematic Depiction
of the "Diffusion of Innovation"
(Adapted from Rogers, 1995 and Others)

References

Ahsan, J.J. and S.U. Khan (1982), Optimum Allocation in Multivariate Stratified Random Sampling with Overhead Cost. *Metrika* **29**, 71-78.

American Statistical Association (1994). A Research Agenda to Guide and Improve the Current Employment Statistics Survey. Report of the American Statistical Association Panel for the Bureau of Labor Statistics' Current Employment Statistics Survey.

Benjamini, Y. and Y. Hochberg (1995), Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society, Series B* **57**, 289-300.

Bonnen, J. (1988). Comment on, "Statistical Policy for State and Local Governments." *The American Statistician* **42**, 16-18.

Brackstone, G. (1999). Managing Data Quality in a Statistical Agency. *Survey Methodology* **25**, 139-149.

Butani, S., Harter, R. and Wolter, K. (1997). Estimation Procedures for the Bureau of Labor Statistics Current Employment Statistics Program. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 523-528.

Butani, S., Stamas, G. and Brick, M. (1997). Sample Redesign for the Current Employment Statistics Survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 517-522.

Cochran, W.G. (1977). *Sampling Techniques, Third Edition.* New York: Wiley.

Dillman, D.A. (1996). Why Innovation Is Difficult in Government Surveys (With Discussion). *Journal of Official Statistics* **12**, 113-197.

DuMouchel, W. (1999). Bayesian Data Mining in Large Frequency Tables, With an Application to the FDA Spontaneous Reporting System (with discussion). *The American Statistician* **53**, 177-202.

Duncan, J.W. (1988). Comment on, "Statistical Policy for State and Local Governments." *The American Statistician* **42**, 21-22.

Eckler, A.R. (1972), Some Comments on "Federal Statistics" (Report of the President's Commission on Federal Statistics – Volume I). *The American Statistician* **26**, 11-14.

Eltinge, J.L., J. Fields, R.C. Fisher, J. Gershunskaya, P. Getz, L. Huff, R. Tiller and D. Waddington (2001). Small Domain Estimation in the Current Employment Statistics Program. Paper presented to the Federal Economic Statistics Advisory Committee, June, 2001.

Fay, R.E. and R. Herriot (1979), Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association* **74**, 269-277.

Felligi, I.P. (1996). Characteristics of an Effective Statistical System (with discussion). *International Statistical Review* **64**, 165-197.

Fuller, W.A. (1999). Environmental Surveys Over Time. *Journal of Agricultural, Biological and Environmental Statistics* **4**, 331-345.

Genovese, C. and L. Wasserman (2002), Operating Characteristics and Extensions of the False Discovery Rate Procedure. *Journal of the Royal Statistical Society, Series B* 64, 499-517.

Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: an appraisal (with discussion). *Statistical Science* 9, 55-93.

Grzesiak, T.J. and C. Johnson (1989), Sample Allocation Meeting Multiple Estimate Reliability Requirements. *BLS Statistical Notes, Number 29.*

Hansen, M.H., W.G. Madow and B.J. Tepping (1983), An Evaluation of Model-Dependent and Probability-Sampling Inferences in Sample Surveys (With Discussion). *Journal of the American Statistical Association* **78**, 776-807.

Harris, L. (1972). A Note on the Decision Process of the Official Statistician. *The American Statistician* **26**, 23-24.

Harter, R.M., K.M. Wolter and M. Macaluso (1999), Small Domain Estimation of Employment Using CES and ES202 Data. *Statistical Policy Working Paper30: 1999 Federal Committee on Statistical Methodology Research Conference: Complete Proceedings*, 195-205.

Hartley, H.O. (1974), Multiple Frame Methodology and Selected Applications. *Sankhya, Series C* **36** 99-118.

Jenkins, R. and R. Chapman (1998). The Process of Adoption. *Proceedings of the 1998 Winter Simulation Conference*, 1547-1554.

Kaysen, C., C.C. Holt, R. Holton, G. Kozmetsky, H.R. Morrison and R. Ruggles (1969), Report of the Task Force on the Storage of and Access to Government Statistics. *The American Statistician* **23**, 11-19.

Kish, L. (1976). Optima and Proxima in Linear Sample Designs. *Journal of the Royal Statistical Society, Series B* **139**, 80-95.

Lehnen, R.G. (1988). Statistical Policy for State and Local Governments. *The American Statistician* **42**, 10-16.

Marker, D.A. (1999). Organization of Small Area Estimators Using a Generalized Linear Regression Framework. *Journal of Official Statistics* **15**, 1-24.

Marker, D.A. (2001). Producing Small Area Estimates From National Surveys: Methods for Minimizing Use of Indirect Estimators. *Survey Methodology* **27**, 183-188.

Moser, C. (1976). The Role of the Central Statistical Office in Assisting Public Policy Makers. *The American Statistician* **30**, 59-67.

National Research Council (2000a). *Small-Area Estimates of School-Age Children in Poverty: Evaluation of Current Methodology*. Panel on Estimates of Poverty for Small Geographic Areas, Constance F. Citro and Graham Kalton, editors. Committee on National Statistics. Washington, DC: National Academy Press.

National Research Council (2000b). *Small-Area Income and Poverty Estimates: Priorities for 2000 and Beyond.* Panel on Estimates of Poverty for Small Geographic Areas, Constance F. Citro and Graham Kalton, editors. Committee on National Statistics. Washington, DC: National Academy Press.

National Research Council (2001). *Principles and Practices for a Federal Statistical Agency, Second Edition.* Committee on National Statistics. Margaret E. Martin, Miron L. Straf and Constance F. Citro, Editors. Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Neumann, K. (1999), Applied Systems Analysis in Official Statistics. *Journal of Official Statistics* **15**, 103-114.

O'Hare, W. and K.M. Pollard (1998). Assessing the Devolution Revolution: How Accurate Are State-Level Estimates from the Current Population Survey? *Population Research and Policy Review* **17**, 21-36.

Parke, R., C. Taeuber, J.H. Aiken, A.D. Biderman, B. Clyman, D. Greenwald, J.F. Kantner, R.E. Lewis, H.C. Passer, M.D. Wann, R. Loewenstein and F.C. Leone (1976), Improving the Federal Statistical System. *The American Statistician* **30**, 154-158.

Platek, R., J.N.K. Rao, C.-E. Sarndal, and M.P. Singh (1987). *Small Area Statistics.* New York: Wiley.

Rao, J.N.K. (2002). *Small Area Estimation.* New York: Wiley.

Renner, R.M. (1979), Estimation of Company Income Statistics. *The New Zealand Statistician* **14** (3), 19-30.

Reynolds, R.T. (1988), Comment on, "Statistical Policy for State and Local Governments." *The American Statistician* **42**, 18-20.

Rogers, E.M. (1995). *Diffusion of Innovations, Fourth Edition*. New York: The Free Press.

Rosenberg, H.M. and G.C. Myers (1977), State Demographic Centers: Their Current Status. *The American Statistician* **31**, 141-146.

Ryan, B. and N.C. Gross (1943), The Diffusion of Hybrid Seed Corn in Two Iowa Communities. *Rural Sociology* **8**, 15-24.

Savage, I.R. (1985), Hard-Soft Problems. *Journal of the American Statistical Association* **80**, 1-7.

Sackrowitz, H. and E. Samuel-Cahn (1999). *P* Values as Random Variables – Expected *P* Values. *The American Statistician* **53**, 326-331.

Schaible, W.L. (2000). *Indirect Estimators in U.S. Federal Programs.* Lecture Notes in Statistics, Volume 108. New York: Springer-Verlag.

Schirm, A.L, A.M. Zaslavsky and J.L. Czajka (1999), Large Numbers of Estimates for Small Areas. *Statistical Policy Working Paper30: 1999 Federal Committee on Statistical Methodology Research Conference: Complete Proceedings*, 205-211.

Shen, W. and T.A. Louis (1998). Triple-Goal Estimates in Two-Stage Hierarchical Models. *Journal of the Royal Statistical Society, Series B* **60**, 455-471.

Shiskin, J. (1970), Strengthening Federal Statistics. *The American Statistician* **24**, 15-25.

Singh, A.C., D.M. Stukel and D. Pfeffermann (1998). Bayesian Versus Frequentist Measures of Error in Small Area Estimation. *Journal of the Royal Statistical Society, Series B* **60**, 377-396.

Smith, T.M.F. (1994), Sample Surveys 1975-1990: An Age of Reconciliation? (With Discussion). *International Statistical Review* **62**, 3-34.

Storey, J.D. (2002). A Direct Approach to False Discovery Rates. *Journal of the Royal Statistical Society, Series B* 64, 479-498.

Weiner, N.S. (1974). The Report of the President's Commission on Federal Statistics: A Summary of Recommendations, Views and Counterviews. *The American Statistician* 28, 42-46.

Werking, G. (1997). Overview of the CES Redesign. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 512-516.

West, S., Kratzke, D. and Grden, P. (1997). Estimators for Average Hourly Earnings and Average Weekly Hours for the Current Employment Statistics Survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 529-534.

Policy Considerations in the Development of State Estimates of Substance Use Rates

Doug Wright (SAMHSA)

The National Household Survey on Drug Abuse (NHSDA) is the primary source of information on the use of illicit drugs, alcohol, and tobacco by the civilian, noninstitutionalized population in the United States. The survey is sponsored by the Substance Abuse and Mental Health Services Administration (SAMHSA), and currently conducted by Research Triangle Institute under the direction of SAMHSA's Office of Applied Studies (OAS). Conducted since 1971, and annually since 1990, the survey collects data by administering questionnaires to a representative sample of the population through face-to-face interviews at their places of residence. Each year, the NHSDA interviews approximately 70,000 people age 12 years and older over a 12 month period. In 2000, SAMHSA published for the first time State estimates from the 1999 NHSDA using small area estimation techniques for the 50 states and DC. These were based on small area techniques that had been used on two different trial occasions, first using the 1991-1993 data combined, and later using the 1994-1996 data. That work demonstrated that the small area estimation methodology would work with a minimum sample size of approximately 400 persons per State. The 1999 NHSDA represented the first year in which the state sample sizes and allocation had been determined with the goal of making state-level estimates.

Prior to 1999, the NHSDA had a national design that utilized a first stage sample of approximately 120 counties (or groups of neighboring counties) and subsequent samples of block groups (either single blocks or groups of neighboring blocks), households, and persons. Sample sizes in the early 1990's ranged from about 18,000 to 28,000 respondents per year. In 1998, Congress requested SAMHSA to expand the NHSDA sample to provide State estimates. The purpose for the State estimates was to use them in conjunction with other available information to help determine those states having significant drug problems, the effectiveness of their programs, and the best allocation of block grant funds in order to reduce the national drug problem.[1]

The goals of the design for 1999 and later years were to provide state-level estimates of prevalence rates for approximately 20 measures for all persons 12 and older and separately for three age groups: 12-17, 18-25, and 26 and older. Approximately 900 cases were allocated evenly to the three age groups within each of the 42 states with smaller populations and the District of Columbia. In order to improve the estimates of precision at the national level, the remaining eight largest states were each allocated samples of approximately 3,600 cases. The total sample size was approximately 70,000 persons.

The conceptualization of the small area estimation is unique in many ways. Estimates are made at the block group level for every block group in the nation and summed up to the state level. The estimation method employs a full hierarchical Bayes approach that utilizes the sample weights, which are callibrated to approach the design-based estimate as the sample size increases. The model

[1]SAMHSA's Report to Congress: The Expansion of the NHSDA and SAMHSA's Plans to Improve Substance Abuse Services, Section III, page 3.

includes fixed and random effects which are estimated using Markov Chain Monte Carlo (MCMC) processes:

$$\lambda = \mathbf{X}\beta + \mathbf{Z}U$$

where $X\beta$ are the fixed effects and **ZU** are random effects.

The fixed effects include demographic variables from the Decennial Census at the block and tract level and county-level data related to substance use from a variety of other federal agencies. The random effects are estimated at the State and Field Interview (FI) region group levels. Each state has either 12 or 48 FI regions (geographic strata) depending on whether the State sample size is 900 or 3600. These regions are grouped into larger regions for estimating the random effects such that the small states (in terms of sample allocation) have 4 FI region groups and the large states have 16 such groups.

In order to obtain the best advice, an advisory panel of experts in small area estimation was constituted to advise the project by helping inform our review of the procedures and results, and the presentation and interpretation of those results.[2] This panel met once before the first release of the 1999 State estimates. Subsequently, it met a number of times to review and recommend how to publish the 2000 State estimates and what research to pursue to improve the precision of estimates of annual change at the State level.

The primary goal of state estimation was to rank the states from highest to lowest on a number of licit and illicit substance use measures and to measure the annual change at the state level. One of the crucial aspects of this project was to provide estimates of the precision of the small area estimates. Another was to validate the State estimates. Given the large sample sizes in the eight largest states, their small area estimates were heavily weighted toward the design-based estimates, and little weight was placed on the national model. However, for the other 42 states and the District of Columbia, in which the sample sizes were approximately 900, their precision rested more heavily on the fixed national model and validation was especially important for them.

The validation was based on combining large states so as to form four pairs of estimates. Thus, each of the four pairs of states had sample sizes of approximately 8,000 cases that could provide design-based estimates that were very accurate. The direct estimates for these quasi-states that were based on all 8,000 cases became the "true" values that would be compared to the small area estimates that were generated. To generate estimates that were comparable in design and process to our small area estimates for the smaller states, we exactly replicated the sampling process to produce "pseudo-

[2] The panel included William Bell of the U.S. Bureau of the Census, Partha Lahiri of the University of Nebraska, Balgobin Nandram of Worcester Polytechnic Institute and NCHS, Wesley Schaible formerly Associate Commissioner for Research and Evaluation at the Bureau of Labor Statistics, J.N.K. Rao of Carleton University, and Alan Zaslavsky of Harvard University.

states" with sample sizes of approximately 900 cases. Then the HB procedures were used to fit models and generate estimates for four substances for all persons age 12 and older, and for three age subgroups. We estimated the relative absolute bias for each measure. We also generated the 95% prediction intervals from the MCMC process and compared these to the 95% confidence intervals of the design-based estimates based on samples of the same size.

Results

Year 1. The results were released in two formats: tables of estimates alphabetically by state with their associated 95% prediction intervals and maps that reflected the ranking of states into quintiles. The states and DC were not ranked from 1 to 51 because there was significant bunching in the center of the distribution and prediction intervals were quite large (more so in 1999 when samples were based on a single year's data.). Quintiles were chosen to present the results. Since there were 51 estimates, 10 states were assigned to every quintile except the middle one which was assigned 11. On occasion, two or more States had identical estimates for a measure to the third decimal place or more. When this occurred at a boundary line between quintiles, those States were assigned to the lower quintile because it was desirable to "err" on the conservative side by assigning the states to a lower quintile.

The discussion in the report mainly focused on the highest and lowest quintiles: the States that had the highest or lowest prevalence rates. The maps of the states reflected the quintiles, assigning States in the highest quintile the color red and States in the lowest quintile the color white. Since States often had similar prevalence rates that were not statistically different, data users were encouraged to focus more on the prediction intervals for their estimates rather than the rankings themselves. Also, because the national average was more precise than any state estimate, the emphasis was placed on comparing States to the national average rather than to another state. An extensive technical appendix provided complete information on the validation, prediction intervals, interpretation of results, and sources of potential bias. Simultaneous with the data release, the governor's office in every State and the District of Columbia were sent the results in order to provide time to prepare for press inquiries.

In addition to discussing the states with the highest prevalence rates and the distribution of the top and bottom quintiles across the four Census regions, the discussion also covered the similarity of rankings for similar drugs, e.g. any illicit use and marijuana use, and the similarity across age groups. The 1999 HB estimates were quite good. Based on the validation, the relative absolute biases for the 12 and older age group ranged from a low of 1 percent for past month cigarette use to a high of 6 percent for past year cocaine use. In addition, the model-based prediction intervals were approximately 35% shorter than the corresponding design-based confidence intervals.

The NHSDA estimates have been useful in identifying states that have the highest and lowest prevalence rates, and raising questions about the possible reasons for those differences. For those states with low estimates, policy officials and researchers can ask what protective factors do those states possess and can their experience be transferred to the states with high prevalence rates. The

rankings are also useful in identifying classes of drugs that have similar state rankings. The maps are useful in seeing the extent of regional similarity in the use of licit and illicit substances.

States are responsible for allocation of funds within the state. At this time, the only substate estimates that can be estimated from the HB process are those for the FI region groups where we estimate random effects. The FI region groups, however, don't necessarily match areas of interest to the States. At the state level, we have provided design-based estimates for a single year for those states with the large annual samples of approximately 3600 persons. For small states, we have combined two or more years to provide special design-based estimates.

Year 2. In the second year (2000), it was possible for the first time to estimate annual change at the state level using the HB methodology. However, evaluation of those estimates revealed that the changes were so small, and the prediction intervals so large, that there were almost no significant differences for any of the states for any of the measures. After meeting with our expert panel, it was agreed that we should not publish estimates of change based on just two years' data. Rather, we should combine two years' of data together in order to estimate a moving average and thus improve the precision of State estimates. Various options were considered for improving the estimates of change in the future, including estimating the difference of two consecutive moving averages, using retrospective estimates to improve precision of change, and simultaneously modeling two or more related variables. Another option discussed was to obtain better predictors of change in the HB model. Unfortunately, there do not appear to be any current national sources for county-level (or lower) data that reflect programmatic activity in the area of substance use prevention or treatment, or in other programs aimed at reducing substance use.

The 2-year estimates have been very precise with much less shrinkage toward the national model component, especially for the states with annual samples of about 900 persons. The rankings based on 2 years' data are very similar to those based on just the 1999 data. The relative absolute biases remained similar to those for 1999 for the 12 and older age group, ranging from about 1 percent for past month cigarette use and past month "binge" alcohol use, to about 8 percent for past year cocaine use. The prediction intervals based on 2 years' data were smaller than the corresponding prediction intervals based on a single year's data. For example, the prediction interval width for past month use of marijuana (persons age 12 and older) was 2.40 percent in 1999, but only 1.98 percent for 1999 and 2000 combined.

A public use file for the NHSDA has been developed; however, it does not include state identifiers for reasons of confidentiality - nor does it contain a linkage between sample respondents from the same household. The public use file is based on a complex disclosure method that utilizes subsampling and substitution subject to constraints that minimize the decrease in precision relative to the full file. At this time, we discussing the possibility of a license that would permit qualified users to conduct analysis with the full confidential file. We are also considering a data analysis system that might permit estimation of State-level crosstabulations, but provide no access to download the State identifiers.

Now in our third year of State estimation, we are obtaining our first estimates of change based on the difference of the 1999&2000 estimate from the 2000&2001 estimate. We believe that we may be starting to witness the emerging effects of prevention, treatment, and other programs aimed at reducing substance use, in that a few States that had high estimates in 1999 are beginning to show slight decreases in prevalence rates.

Response and nonresponse bias have always been a concern when collecting sensitive information. This is especially so given the varied response rates among states and the observation that there is a negative correlation across states between response rates and reported prevalence levels. Also, changes in the methodology, such as new field interviewer training to improve adherence to data collection standards and the use of monetary incentives to improve response rates, have apparently caused significant increases in prevalence rates, making it difficult or impossible to measure true year-to-year change net of any "field effects." We are currently studying these issues and the impact on our State small area estimation program.

Providing Small Area Estimates Discussion

Graham Kalton
Westat

Since the end of World War II the demand for survey data has experienced a continuous and ongoing expansion. In part, the expansion has been in the range of topics for which survey data are needed, and that has stimulated a number of methodological developments. Using newer data collection techniques, surveys are now used to collect data on topics—particularly sensitive topics—that in earlier years would have been considered beyond the realm of survey research. The National Household Survey of Drug Abuse (NHSDA)—the subject of Doug Wright's paper—is an example. In part, the expansion has also been in the sophistication of the demand for survey data. Whereas in the past policymakers would make do with often somewhat dated national estimates and estimates for a few large domains, their current demands are for timely data and for estimates for small domains. Some small domains are nongeographic subgroups, such as demographic domains (e.g., domains based on combinations of age, race, and sex) in population surveys and industrial division and size class in establishment surveys. For such domains, the production of small domain estimates of adequate precision may be achieved through increasing the survey's sample size and using methods to oversample the smaller domains. Another approach is to accumulate the sample over time, as is planned for the American Community Survey and is done in the NHSDA.

Other small domains are small geographic areas, such as states, metropolitan statistical areas, counties, and school districts. The expansion of sample required to produce reliable estimates even for states is often greater than resources can support. Moreover, the small area estimates of interest often relate to only a subdomain of the total population (such as the estimates for 12- to 17-year-olds in the NHSDA), in which case even larger sample sizes are needed. In such a situation the solution of accumulating sample over time may require too long a time period to satisfy the need for up-to-date estimates. Thus, alternative methods are needed.

The standard model for statistical inference in survey sampling is design-based inference. Design-based, or direct, estimates are not model-dependent, although they may be model-assisted. When a survey's sample size is inadequate to produce reliable direct estimates for small areas, it becomes necessary to employ indirect estimates. These indirect estimates are model-dependent, and there must therefore be concerns about model misspecification.

The essence of small area estimation is the use of auxiliary data available at the small-area level in a statistical model to predict the small area survey statistics of interest. The key requirement for this approach to be effective is the availability of good predictive auxiliary data. Such data can come from administrative records, a past census, or some other source. These data then need to be used in the careful development of a predictive model from which the small area estimates can be produced. An essential component of the approach is a thorough evaluation of the model and the estimates. Finally, valid measures of precision for the small area estimates need to be produced.

The idea of model-dependent estimation for producing small area estimates has a long history. An early example is to be found in the text by Hansen, Hurwitz, and Madow (1953, Volume I, pp. 483-486). Indirect estimates are now published on a regular basis by several Federal statistical agencies. A valuable review of eight Federal small area estimation programs is provided in the report of the Federal Committee on Statistical Methodology (FCSM) subcommittee on the subject (Schaible, 1993, 1996). Since that report was prepared, the U.S. Census Bureau has established its Small Area Income and Poverty Estimates (SAIPE) program and the Substance Abuse and Mental Health Services Administration (SAMHSA) is producing the state estimates of substance use described in Wright's paper. The Bureau of Labor Statistics is considering the introduction of small area estimates with its Current Employment Statistics (CES) program, and John Eltinge's paper discusses the issues involved.

The past decade has seen an explosion of theoretical research on small area estimation models and estimation methods, using in particular empirical best linear unbiased prediction, empirical Bayes, and hierarchical Bayes methods. A range of different models has been developed to cover dependent variables measured as categorical, continuous, or count variables and auxiliary variables measured at the area or unit levels. Also multivariate models have been developed to borrow strength for small area estimates for one subdomain from data for other subdomains and to borrow strength over time. Such methods are often computer intensive, using, for example, Markov Chain Monte Carlo methods. Current computing power and the availability of software, however, now make the application of these methods feasible (see Wright's paper for an example). Rao (2003) provides an excellent account of the current state of small area estimation methods.

While recognizing the importance of the recent theoretical developments, it remains the case that the model estimates can be no better than the auxiliary data on which they are based. Any small area estimation program should give a great deal of attention to finding auxiliary data sources, to checking the suitability of the data for use in models, and to constructing effective indices from the data for use in the models. The auxiliary data need to be measured uniformly across all small areas; alternatively, appropriate adjustments must to be made. The indices formed from those data need to be carefully constructed and thoroughly examined.

As an illustration, consider the estimation of poor school-age children for states and counties in the Census Bureau's SAIPE program. The total numbers of food stamp recipients, which are available monthly for states and annually for counties, are valuable auxiliary data, given that only poor households are eligible to receive food stamps. However, Alaska and Hawaii have higher income eligibility thresholds for food stamps than other states because of their higher costs of living. Thus adjustments need to be made to produce state and county numbers for Alaska and Hawaii that are comparable to the numbers of food stamp recipients in other states and counties in order to avoid distortion in the model-dependent small area estimates.

The indices constructed from the auxiliary data for use in a predictive small area model can significantly affect the quality of the small area estimates. In the initial formulation of the SAIPE state-level model, the food stamp index used in the model was based on the count of food stamp recipients in July of the reference year. Subsequent research led to a change to an index based on the monthly counts over a 12-month period centered on January 1 of the calendar year

subsequent to the reference year. Also the counts were refined to remove persons who received food stamps due to specific natural disasters, and outliers were smoothed (Citro and Kalton, 2000, p. 28). As another example from SAIPE, evaluation of the initial county model estimates identified some distortion in the estimates in counties with high proportions of group quarters residents when an index based on the estimated number of persons aged under 21 was used in the model. This distortion was removed by changing the index to one based on the number of persons aged under 18, and this change also improved the model estimates in other respects (Citro and Kalton, 2000, p. 86).

This discussion draws attention to the importance of including a thorough evaluation of the model and the small area estimates in a small area estimation program. The SAIPE program provides a good illustration. In that program, the estimates of poor school age children have been evaluated by analyzing the regression residuals associated with some alternative models; by applying the models to the 1990 Census year and comparing the model estimates with Census estimates; by grouping counties on the basis of a variety of characteristics and comparing the SAIPE estimates with the direct estimates from the Current Population Survey for these groups; and by examining the stability of the models over time. In any such program imaginative ways to test the quality of the estimates should be sought. In planning a small area estimation program, considerable resources should be allocated to evaluation.

An important feature of small area estimation programs brought out in both Eltinge's and Wright's papers is that many different estimates may be needed for the small areas, not just the single estimate that is the focus of most theoretical work. Both papers point out the demand for estimates of both level and change, and that demand needs to be reflected in the modeling. To serve both these demands adds to complexity; a multivariate model is likely needed to produce valid estimates of change.

In the terms used by Rogers with regard to the diffusion of innovations, as discussed in Eltinge's paper, I would classify the present state of small area estimation in Federal statistics as being "early majority." The use of small area estimation methods is fairly well established in a number of areas, but I think that much greater use will likely be made of these methods in the future. Those who plan to develop new small area estimation programs need to appreciate the resources required and the properties of the resultant estimates. In particular, specialist skills are needed in the high-powered statistical methods that are used and expertise is needed in the auxiliary data sources and their properties. (The FCSM might play an important role in this area by facilitating the exchange of expertise in methods and data sources between Federal statistical agencies with small area estimation programs.) Careful model development and thorough testing are needed, and these are labor-intensive activities. Furthermore, the acquisition and checking of auxiliary data, model development, and testing are time-consuming activities that may seriously affect the ability to produce timely estimates.

The producers of small area indirect estimates need to make users aware of the model-dependent nature of the estimates and of the distinction between such estimates and the usual direct survey estimates. Small area estimation methods can be extremely valuable in addressing users' needs for such estimates. However, these methods should not be viewed as quick, easy, and inexpensive. Small area estimates need to be produced with great care and assessed with caution.

References

Citro, C.F. and Kalton, G. eds. (2000). *Small-Area Estimates of School-Age Children in Poverty. Evaluation of Current Methodology*. Panel on Estimates of Poverty for Small Geographic Areas, National Research Council. National Academy Press: Washington, D. C.

Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953). *Sample Survey Methods and Theory*. Wiley: New York.

Rao, J. N. K. (2003). *Small Area Estimation*. Wiley: New York.

Schaible, W. L. ed. (1993). *Indirect Estimators in U.S. Federal Programs.* Statistical Policy Working Paper 21. National Technical Information Service: Springfield, Virginia.

Schaible, W. L. ed. (1996). *Indirect Estimators in U.S. Federal Programs.* Springer-Verlag: New York.

Session 11

Ensuring Citizen Privacy

Data Privacy and Confidentiality Issues and the Role of the IRB

Lawrence A. Greenfeld, Director
Bureau of Justice Statistics

The collection of data on crime and the administration of justice nationwide, particularly data from individuals on their victimization experiences, is often sensitive and controversial because of the fear of disclosure. A sheriff may fear that disclosure of data will reflect poorly on the operations of his/her jail or that the information may be used by DOJ's Civil Rights Division to challenge the conditions within the jail. Knowing that about 45% of the violent victimization of males and 68% of the violent victimizations of women occur between offenders and victims with a prior relationship, victims may fear disclosure of information about the victimizer but they also may fear the release of what they consider to be intimate, personal information. Many victims, for example, tell us that they did not conceive of the violence they experienced as a criminal act.

BJS data series entail more than 200,000 interviews with the public annually and the collection of administrative data from some 50,000 agencies, offices, and institutions that do something called criminal justice. In addition, every few years we interview tens of thousands of prisoners and jail inmates to learn about their backgrounds and the contingencies of their offense. The most obvious problem we face at BJS is insuring that respondents feel free to share with us their experiences and information without fear of exposure to legal process or authorities. This expectation must never be violated though DOJ litigating branches often express interest in what we collect. This tension between operating agencies who want to use detailed data to make decisions and statistics agencies who want to protect data from disclosure so as to insure the continued ability to obtain the data is, I think, the core of what this session is about. The IRB is thought to represent the best vehicle for insuring that respondents are protected as this drama plays out between stats agencies and operating agencies.

Perhaps the most important element of statistical collections is the maintenance of stable ways to collect information and stability in what is collected over time and across geographic locations so that time series are possible—everyone wants to know whether domestic violence is increasing or decreasing and whether prisons are more crowded this year compared to last year. In my experience, the IRB generally does not demonstrate a strong commitment to this core principle of stability. Rather, the potential discomfort of the respondent may be exaggerated by IRB members in what I have seen as undisciplined meandering through data collection instruments, often instruments which have a long track-record of successful use without complaint. In addition, many instruments reflect a consensus as to what is actually collectible from literally thousands of agencies within a framework of common counting rules and units.

Much of our thinking about privacy and confidentiality comes from various protections afforded respondents who cooperate in Federal statistical programs. The Federal guidelines for the protection of human subjects have been incorporated into the regulatory structure of many Federal agencies. 45 CFR 46, Subpart A, which is now known as the Common Rule defines the criteria for IRB review, the elements of research protocols which are relevant for concern such as informed consent, and it describes the role and operation of the IRB. The DOJ regulation is found at 28 CFR Part 46.

The Common Rule states that "When appropriate, there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data."

For most of what we do as Federal statistics agencies, risks to privacy and confidentiality are the primary risks associated with our work. However there are other risks as well for respondents for which we may or may not be concerned---inconvenience, emotional upset or worry, legal risk if illegal activities are disclosed, etc. Most of us do not prepare a detailed list of every conceivable risk a respondent could face as a result of participation in a self-report survey, for example, since we consider privacy and confidentiality the main concerns for risk of our intrusion into the life of the respondent. But we should always be mindful that a sensitive and private matter is often a troubling matter for the respondent. In addition, a victim-respondent's disclosure to data collectors about certain behaviors by others---such as domestic violence committed by a husband or boyfriend---also raises privacy and confidentiality issues relative to the offender. I think we always need to carefully review our procedures to insure that respondents fully understand the nature of the consent they give---especially children, vulnerable populations, non-English speaking respondents, and even proxy respondents.

The Common Rule really does not clearly tell us objectively what to do to minimize the risk of privacy or confidentiality problems. Essentially, the Common Rule advises us only to use procedures "which are consistent with sound research design and which do not unnecessarily expose subjects to risk." At BJS we often spend a great deal of time, often together with the Census Bureau's Disclosure Review Board, thinking through what anonymity really means---birthdates, occupations, race, geographic locations, gender, type of offense, etc.---may make someone identifiable even when no name or address is given.

Recently, for example, I was utilizing the FBI's Supplementary Homicide Report file covering the years 1976-99; this file contains over a half million descriptions of individual homicides provided by investigating officers occurring over that reference period without names or addresses for any of the victims or offenders. I wanted to see if I could find the actual statistical information filed on the OJ Simpson case. First, I selected cases with multiple victims; among these I selected cases in cities over 1 million population; among these I selected cases where the victims were white and the offender was black. I then selected cases where the offense had been cleared by the arrest of a black male. I then selected cases where the offender and at least one of the victims had a prior relationship as spouse or ex-spouse. I came up with one case which met all the criteria and it occurred in June 1994 the same month as the murders in the Simpson case. The police had coded the circumstance of the case as a "lover's triangle" which was new information to me. The point is to illustrate that even with a very large, "anonymous' dataset, knowing only a few pieces of information can reveal identities.

Often at BJS and at other agencies as well, I am sure, we use privacy and confidentiality as interchangeable terms. Privacy generally refers to how we wish to control what access we give others to ourselves. Confidentiality refers to procedures we use to keep identifiable data from being disclosed and what agreements we make with a data supplier, such as a respondent, about the handling of the information they give us. Privacy is more of a conceptual concern while confidentiality is a procedural concern. It is assumed that if we have the proper confidentiality procedures, privacy will be protected. As the Simpson case illustrates, I think, even the best

confidentiality procedures may not fully protect privacy, if the contingencies or characteristics of an event are sufficiently rare.

At BJS, there are really 6 key confidentiality actions we take to try to make privacy a primary goal:

1) Eliminate linkage of data to unique identifiers---our surveys are conducted with an assurance of anonymity. Different procedures may be needed depending upon whether the survey is being conducted cross-sectionally or longitudinally.

2) Minimize identifiable information from linked datasets---oftentimes we may try to match individuals from one dataset to another---a frequent practice of ours to study recidivism, for example, where we match correctional discharges with criminal records. Scrambling of identifying numbers or other information or their elimination from public use datasets is essential.

3) Utilize statistical methods to protect privacy---a very frequent practice at BJS is to group or aggregate data likely to result in identifiability, such as age, employment, education, income, number of offenders (i.e. 2 or more), etc.

4) Collect data under legal authority guaranteeing confidentiality---we have a statute governing the operation of BJS which provides for protection against disclosure and we often utilize Census Bureau's statute as well. We also have grantees and others who use our data enter into confidentiality agreements (privacy certificates) with us. Such certificates create the kind of intermediate data that are not really fully public but are made available for limited, specified research purposes.

5) Test the dataset to see if small cells can be produced---as a part of the process for preparing a dataset for public release, we try to run numerous cross-tabulations to see if cells can reasonably be generated containing only 1 respondent. If such is the case, we will impose greater aggregation to eliminate the likelihood of creating identifiable cases. A good example, which often occurs, is the use of "other race" to cover persons who may have identified themselves as Asian, Pacific Islander, American Indian, Alaska Native, etc.

For statisticians, aggregation is a very tough problem as we are always curious about subgroup differences and fear that aggregation will introduce bias into an analysis. The problem of aggregation bias is a very real one as revealed in the following example: Suppose I was studying racial bias in the implementation of the death penalty in a State which has only two counties---County A and County B. Suppose County A sentences 10/20 black murderers to death and 50/100 white murderers to death. In each case the probability for white and black murderers is identical . In County B, 1/5 black murderers is sentenced to death and 100 of 500 white murderers is sentenced to death---again, within this county the probability of a death sentence by race is exactly the same. When we aggregate to produce "Statewide" statistics we find that for blacks the probability was 11/25 or 44% while for white murderers the probability was 150/600 or 25%. We might erroneously conclude that there is a huge racial disparity yet in no

jurisdiction within the State was that the case. While aggregation is an obvious solution to problems of confidentiality, it may also introduce other problems.

> 6) Use IRB review as a last resort after seeking exemption—BJS has consistently sought exemption from IRB review for our major statistical series, a strategy which has been based upon our view that the review procedures within BJS and Census Bureau are sufficient and that our governing legislation is clear with respect to confidentiality. In some cases, exemptions have not been granted because juveniles are respondents. However, in these cases, the IRB found minimal to no risk involved and the collections cleared review.

Based on our experience, BJS believes that for IRBs to effectively work with statistical agencies, they must:

> (1) recognize the need to measure the same thing over time and consider the past success of data collections in protecting human subjects (for instance, we have been collecting data from jail inmates for 25 years and have never received a complaint related to the survey);
>
> (2) keep panel discussions focused on human subject protection issues and not stray into areas which go beyond their purview, such as survey administration;
>
> (3) give adequate regard for the procedures already in place to protect human subjects, such as data confidentiality statutes and procedures; and
>
> (4) ensure that the IRB members possesses certain relevant knowledges, particularly in the area of survey methodology.

When IRBs fail to instill these elements, the result is much more paperwork for the statistical agency in responding to questions that oftentimes has already been provided (and not clearly understood by the IRB) and at times, irrelevant to the matter at hand. The result is a poor relationship between the agency and the IRB.

BJS does however, acknowledge that the IRB process impels us to think through our procedures to make certain that interviewers and others who come into contact with respondents are adequately prepared to handle upsetting or troubling matters about which we want to know—i.e. domestic violence, sexual assault, child abuse, etc.

Another related issue that BJS, and possibly other research agencies, is experiencing is how to handle situations in which the respondent reveals information that they are being abused or in some way endangered -- particularly if they are under 18 years old. One of our most important surveys at BJS is the National Crime Victimization Survey or NCVS. The NCVS utilizes a nationally representative sample of 50,000 households who participate under a rotating panel design---each household is in sample for 3 years before rotating out. The first interview from the 7 which take place during the 3-year period is a bounding interview used only for cueing respondents temporally. Respondents can be as young as 12 years old and are asked very

sensitive questions about assaults, sexual assaults, and other crimes they may be experiencing, sometimes on a serial basis. Oftentimes, the offender described by the victim is another household member, neighbor, or family acquaintance. I have struggled with the issue of what to do when young or other vulnerable respondents tell our interviewers about a continuing pattern of abuse and victimization which has never been reported to police or other agencies. Obviously "doing something" about the continuing victimization will be rate-affecting given the longitudinal nature of the design and our promise of strict confidentiality. Various suggestions have included replacing the household in the sample with another similar household and providing social service information to the victim. I would certainly appreciate any thoughts on this.

Basically, each agency needs to consider a set of procedures which may be tailored to the individual series but which follows a general set of principles with respect to both respondent concerns and disclosure concerns. In my mind, methods for addressing the respondent concerns are far less developed than the methods for protecting against disclosure of data. In our case, providing assistance or intervention to a vulnerable respondent who is experiencing emotional upset as a consequence of our questions could affect what we are trying to measure over time and I suspect among many of you this is also a great concern. But the single most important challenge, both statistically and morally, is that once we know the respondent is experiencing a continuing exposure to victimization, what should we do about it?

Another matter of growing concern with respect to privacy among statistical agencies has nothing to do with IRB's. One of the principal challenges to the protection of privacy is the loss of control over computer networks and the increasing centralization of information management in the hands of a CIO. That is, CIO's for the departments in which stats agencies are often housed are assuming increasing oversight for all computer functions. This has the effect of potentially reducing the firewall between policy-making and policy-advocating branches and the statistical agencies. For example, BJS has been aggressively moving toward the collection of administrative data through web-based collection protocols. The CIO is also asserting control over our software purchases and maintenance contracts as our administrative funds are given to the CIO to manage these computer-related functions. There is obviously much that we will need to negotiate with the CIO about the architecture of our computing facilities to insure the protection of respondent privacy and the security of our statistical data—but it is a challenge which will probably be similarly experienced by many in the Federal statistical community.

To conclude, I guess my bottom line is that while I understand the need for IRB's, social science surveys and data collection conducted by Federal statistical agencies require a very different perspective than other types of statistical activity which involves intrusive methods for collecting data or research conducted for the government by outside entities. For the most part, respondent risks associated with Federal social science surveys primarily involve the risk of disclosure, a concern which has always been of paramount concern to Federal statistical agencies and one for which we have all put in place procedures and statutes to protect against such breaches of standards and practice.

In the most recent IRB in which we participated, the review of the our Survey of Inmates in Local Jails, a survey which we have conducted since 1972, the IRB requested that we prepare

some hundreds of pages of material including about 40 pages responding to their questions about why we ask certain questions. The need to maintain stability over time in what is collected should not be subject to the whims of an IRB made up of persons with little or no familiarity with the long-term need for particular data; for many of our statistical series it creates a paperwork burden to simply, in the final analysis, receive approval for the things we have been doing successfully for 30 years without ever having received a single respondent complaint.

Oval Pegs in Round Holes: Health Surveys and the Common Rule

Jennifer Madans, PhD
National Center for Health Statistics

The application of the Common Rule to the federal statistical system is not done uniformly and has not been without some challenges where it has been implemented. NCHS is a good case study of how these regulations have been applied because its data collection systems bridge social, behavioral and health sciences and even encroach into the fringes of biomedical research (eg. NHANES).

My comments are based on my personal experiences as the NCHS Human Subjects Contact, as an observer and presenter at IRB meetings, and as a reader of multiple IRB submissions and reports. Others are more expert in the regulations themselves; the perspective I hope to bring is how these regulations have been applied in one federal statistical system. Primarily, I will talk about survey research and mostly, but not exclusively, from the perspective of using questionnaires to collect data.

History of NCHS's Experience With IRBs:

The manner in which human subject's protection activities and IRB operations have been conducted at NCHS has matured over time and is reflective of changes in the agency's organizational location, in the type of data being collected and in the IRB environment itself. NCHS has a long history of conducting human subjects review (starting in xxx) and currently, the IRB is monitoring about 45 active protocols. Since January 2002, the IRB has reviewed and processed 16 new protocols (9 full Board, 7 expedited), 27 continuations (6 full board, 21 expedited), and 52 amendments (14 full board, 38 expedited).

The NCHS IRB may be unique in that it reviews a relatively small number of protocols but the protocols are for relatively large, repeating studies--the basic data collection systems of the agency. This would likely be the case for other statistical agencies. Some of our data systems submit a single protocol and undergo a series of continuations and amendments; one study has submitted 71 amendments and is still in operation. The oldest protocol still running was initiated 12/06/96. Other data systems submit a new protocol for each data collection cycle, usually a year. However, all NCHS protocols will require at least one continuation as data collection and processing go beyond the 12 months from when the protocol is first approved. So, in any given year, a data collection system would submit a new protocol, a continuation of the protocol submitted the year before and a termination for a protocol submitted in a year prior to that.

Application of 45 CFR 46:

To determine if NCHS surveys fall under 45 CFR 46, it would be necessary to review the requirements for inclusion (do the surveys meet the definition of research involving human subjects as used in the regulation) and exemption. Of note is that most data collections that are done by statistical and other agencies do not obtain IRB approval. One could argue that the Common Rule does not apply or that all protocols would be exempt given NCHS's authorizing legislation and requirements for confidentiality. However, even if this argument could be upheld there would still be the need to assure that the rights of subjects were being protected. If NCHS did not review protocols under 45 CFR 46, we would need to develop another process for review but that process would have no external validity and would not be held in as high a regard as the IRB process. As a result, we have chosen not to make the argument that 45 CFR 46 does not apply. While this had lead to implementation problems, there have also been benefits from under going review under 45 CFR 46 including improved protocols and documentation of survey procedures and more careful consideration of human subjects issues, specifically informed consent.

Although intended to apply more broadly, 45 CFR 46 was modeled for use in biomedical research involving some risk and this model doesn't always fit the survey situation. If research isn't exempt, it follows the same rules as clinical trials sometimes making it difficult to implement the regulations in a reasonable way. This causes problems for investigators and for the IRB. The unintended consequence is that there is an appearance that regulations are not meaningful but just a bureaucratic hurdle to be met.

What Makes For Oval Pegs:

There are several characteristics of survey research that can make applying the regulations challenging.

The evaluation of risks and benefits--If one considers the entire risk/benefit continuum, NCHS surveys are definitely at the low risk end with the main risk being a breech of confidentiality that would lead to the damage to the subject's reputation, financial standing or employability not an adverse effect of a medical intervention. Other risks are difficult to evaluate. For example, how can one evaluate hypothetical risks such as the possible psychological discomfort of asking people about illnesses they may have. Who can determine how likely such a risk is or how severe it would be? It is also more difficult to evaluate the benefits, as they tend to be indirect and non-specific. How should the IRB deal with these potential risks? NCHS surveys are generally national in scope but IRBs are meant to reflect community attitudes and adhere to state law. There is a general lack of guidance in how to adjudicate local vs. national requirements.

Meaning of informed consent in the survey situation-- A major component of IRB review is the evaluation of the informed consent process. The nature of the survey process raises some unique

questions about critical aspects of the process. For example, the identity of the subject is not always clear at the onset of the survey process. This raises the question of who needs to be informed and to consent to the research and when does the consent process start. The consent process for surveys also takes different forms. Advance letters are often sent prior to actual contact. Many subjects don't receive or read these letters. Are the letters considered part of the process and must they include all required items? The primary risk associated with participating in most surveys is the adverse effects of disclosure and many statistical agencies have the legal ability to protect confidentiality. However, this is different than being able to protect against all disclosure risk either during data collection or after release especially in an evolving IT environment. What must respondents be told about the possibility of risks to confidentiality that are either very small or that might occur in the future?

Informed consent makes most sense in situations where the components of the research can be clearly described, as is the case for most clinical research. This is harder in the survey situation where it is difficult to convey comprehensive information about the information that will be collected. In cases where the data collection is multi-purpose both the nature of commitment (generally in terms of time) and the exact content of the survey varies significantly across subjects. Can respondents truly consent to participation before hearing all that will be asked of them but how can this be done prior to actually asking the questions? Is has be suggested that it makes more sense to obtain consent after an interview than before it and that only some aspects of consent need to be mentioned prior to beginning the data collection.

Nature of the information obtained—The notion of private information is used to determine whether an individual is in fact a "human subject" but the distinction between private and public information is not always clear. Many times, the information obtained in surveys might be considered shared information or quasi-public information. For example, if a family member can report information about another family member or if an informant can provide information about the observable characteristics of another person, is the information private? These gradations of "private" are not dealt with in the IRB regulations but they can be important in survey research.

Recognition of investigator expertise— Questionnaires are designed to be understandable to respondents and since it is easy for IRB members to think of themselves as potential respondents (they would be more likely to be survey respondents of some kind than the subjects of a clinical trial), it is not unusual for IRB members to devote considerable attention to question wording and to make the leap from respondent to survey designer. However, potential respondents (including IRB members) are not qualified by virtue of their IRB membership to require changes in research instruments. Identifying the appropriate responsibility of the IRB in the area of scientific quality is always difficult but it is more difficult in survey as opposed to clinical research where IRB members are more likely to erroneously believe that they have more technical/scientific expertise than the investigators. IRBs that are over zealous in their requirements regarding technical issues can have the effect of reducing rather than enhancing quality and this is more likely to occur in the case of survey research. It is important to find a

way to obtain appropriate outside scientific review of the technical aspects of protocols so that IRBs can be comfortable with the scientific validity.

Results of putting oval pegs in round holes:

In trying to apply the requirements of 45 CFR 46 without acknowledging basic differences in research characteristics can have unintended consequences that serve neither to protect subjects nor to improve the science. Often, informed consent processes can overstate the risks of the research thus making it difficult for the respondent to appropriately way the risks and benefits. Perhaps more problematic is that overly strict requirements can give the appearance that the agency sponsoring the research is trying to avoid liability by transferring the burden to the respondent. For example, some IRBs have required signed informed consent prior to participation in survey research. Studies have shown that some respondents find this inappropriate given the nature of the research and question why a signature is needed.

A lack of flexibility can lead to large workloads for the IRB as they have to work very hard to figure out how to make the pegs fit. This can also lead to different standards being used across IRBs, which diminishes credibility of the system especially when data collection is national in scope.

Do we change the pegs or the holes?

To maximize the protection of survey subjects as well as scientific quality we need procedures that will provide human subjects protection appropriate to the level and nature of the risk and that respect the rights of the subject without unnecessary and seemingly irrelevant requirements. One option is to create regulations that apply to survey research. Such regulations might focus on the following aspects of the research activity: providing a general description of survey (rather than mentioning research which is less meaningful in this context); stressing that participation is voluntary and can that the subject can cease participation at any time; identifying who will see the data and perhaps allowing for the final consent to be obtained at the end of the survey rather than at the beginning. However, is this option really viable and would such regulation carry the same weight as the Common Rule. While the current regulations were not written primarily for the survey situation, they do contain enough flexibility so that they can be successfully applied to the work of statistical agencies, as they have been at NCHS. The use of waivers as well as the ability of the IRB to interpret the regulations provides this flexibility.

It is often suggested that the use of the expedited review can solve some of the problems listed above. However, this really is not the case. The use of an expedited review process does not alter any of the stated, potentially problematic requirements in the regulations and does not always result in a quicker review. The regulations do provide for waivers. However, IRBs do not necessarily grant waivers when the requirements for a waiver are met. In some cases, IRBs also require that investigators demonstrate a positive argument for granting the waiver.

Confusion about how to evaluate waiver requests introduces inconsistencies into the system and can actually decrease scientific value without increasing subject protection.

The IRB is itself a social group that responds to the social context within which it exits and its behavior is, to some extent, conditioned by its environment. Recent events in the biomedical/clinical areas are putting more pressure on IRBs to be more conservative and to use literal interpretations of the regulations. Since this can have adverse consequences in some settings, it is important to also provide mechanisms for discussion of all aspects of the review process and to support research that can shed light on how to best achieve the dual objectives of human subjects protection and quality research. These activities could be directed to providing more guidance to IRBs in how to applying 45 CFR 46 to survey research particularly regarding the granting of waivers. Given that survey research involves no more than minimal risk, IRB might benefit from guidance in what standards to use in determining whether the rights and welfare of subjects would be adversely affected by the waiver and whether the research could not be practicably carried out without the waiver particularly in determining what is practicable. Such guidance would have to come from a respected source with standing in the IRB community and should be the result of serious deliberations. If such further exploration of the issue related to the protection of human subjects in survey research could be conducted, we could likely make oval pegs fit nicely in round holes.

Ensuring Citizen Privacy Discussion

Wendy Visscher
RTI International

I'm pleased to be here from Research Triangle Institute. I enjoyed both presentations and will comment on the issues raised by Larry Greenfield and Jennifer Madans, particularly those related to the role of the IRB in ensuring citizen privacy.

Our collective goal is to be able to do important social research while protecting the people who provide the information we need. This means we need to consider how to protect privacy and confidentiality, how to assess risks, and how to design studies that are compliant with the human subjects regulations and acceptable to the IRB. This is a lot to think about and often very creative procedures are needed to accomplish the scientific goal while assuring the welfare of the participants. And the system for protecting human subjects in social research could be improved.

As Larry Greenfield reminded us, privacy and confidentiality are two closely related, but different things. Privacy is how a person protects his or her personal information, while confidentiality refers to how we – as researchers – protect this information once we have been granted access to it. In research studies, we are concerned about both privacy and confidentiality.

Both speakers pointed out that violations of privacy and confidentiality are generally considered to be the major risks associated with social research. People decide to participate in research studies for a variety of reasons. We make a promise to them to keep their information confidential as part of the informed consent process. Different people consider different types of information to be personal and private. This judgment will affect whether they decide to accept our promise and the associated risk of an inadvertent breach. I think that if we could maximize our ability to protect confidentiality - and the public's confidence in this ability - we might also be able to increase our survey response rates. Larry Greenfield also noted that this could affect data quality if respondents are not convinced that their data will be confidential.

A researcher's plan for protecting privacy and confidentiality is one element that must be considered by an IRB. Unfortunately, the human subjects regulations give very little guidance about what constitutes adequate protection. In fact, all the regulations say is that the provisions for this must be adequate. Thus the IRB must evaluate the proposed procedures for protecting study data, in conjunction with the promises made to respondents in the consent form. Evaluating data security plans is becoming increasingly difficult for IRBs as information technology advances. At RTI, we have done a lot of web-based data collection. Some of the data collected using this technology are quite sensitive. For example, we just completed a study of post-traumatic stress in New York and DC following the September 11 attacks. When our IRB initially reviewed this data collection method, it had to assess very complicated computer systems to determine if the study data would be well protected.

Another issue mentioned by Larry Greenfield is that study respondents must be told if there are exceptions to confidentiality. For example, if minor respondents report abuse, the researchers and the IRB must develop procedures for the mandatory reporting of this information to the proper authorities. RTI is currently conducting a very sensitive study on child welfare that requires elaborate reporting procedures for this.

In addition to the human subjects regulations, the new medical privacy law HIPAA (the Health Insurance Portability and Accountability Act), will afford additional protections to personal health information. This law will restrict how clinics and hospitals can release identifiable medical information to persons outside their institution, including researchers. Fortunately, HIPAA gives slightly more guidance than the human subject regulations regarding what is needed to protect this personal information. In order for a researcher to receive this type of information from a health provider, he or she must satisfy the three criteria: (1) an adequate plan for protecting identifiers, (2) a plan for destroying identifiers at the earliest possible time, and (3) an assurance that they will not release the information to anyone else.

If we acknowledge that potential breaches of privacy or confidentiality are the major risks of social research, what type of risks are they? As the speakers said, risks are hard to assess and the IRB may make a different assessment than do the researchers. Two components of risks must be considered – the probability of harm and the magnitude of harm. We can think of these in a 2 by 2 table, high and low probability by large and small harm. Hopefully, neither a social study nor a medical study falls into the quadrant of high probability of a large harm. It would be nice if all studies were low probability of small harms, but in reality either type of study can fall into any of the other three quadrants. Another thing to keep in mind, from an ethical standpoint, is that a person can be wronged even if he or she is not actually harmed.

So what are the risks that are possible from participation in a social research study. As mentioned, most are related to breach of confidentiality. If private information is released inappropriately, it could adversely affect a person's legal status, financial standing, reputation, job, or insurability. Larry Greenfield noted that these risks would vary for vulnerable populations such as inmates. Another risk to be considered for social studies is the emotional risk associated with recalling, in the course of an interview, past or present events that are upsetting. These are very real risks. Note that the types of information that could be damaging if released – even in a social research study – are not limited to social, demographic, or behavioral data. A definite trend is that these studies are starting to collect biospecimens so that genetic indicators can be studied. DNA information is some of the most identifiable and most personal data and its release could damage not only the study respondent but also his or her family. Jennifer Madans discussed this issue of shared information and third parties in her remarks.

How does the IRB consider risks during its review? The regulations define a risk level called "minimal risk". This definition takes into account both the probability and magnitude of harm, and compares these to those that a person would experience in his or her everyday life. Some

types of social research - that impose no more than minimal risk - may qualify as exempt from IRB review, or can undergo expedited IRB review. Both speakers mentioned IRB exemptions. It appears that exemptions are not applied in a standard fashion across institutions. Jennifer Madans also noted that if a study is exempted, there are often no standard procedures for assuring protections for respondents. She also reiterated that the level of risk should determine what types of protections, and what level of review, the IRB should require for a given study.

Who is responsible for protecting the information our respondents entrust us with and making sure the risks we subject them to are as low as possible? This responsibility falls on all of us who do research with human subjects. As we've all described, assessing possible risks and appropriate protections are not straightforward and are not clearly defined in any regulation. Thus, the client, the research team, and the IRB must work together to protect research participants. And all of us have a vested interest in doing this right. Not only is it the right thing to do, but it increases the public's confidence in research, which increases the chance that they will participate in our studies. The IRB at RTI tries very hard to be collaborative, and not adversarial, with our researchers. Many of our IRB members are researchers themselves so really can help the researchers think through these issues. Larry Greenfield mentioned that some IRB members may not fully understand the needs of longitudinal research. It may be that adding social researchers to the IRB membership could help alleviate this problem.

Both speakers mentioned their experiences with IRB review and the focus of some of these reviews. IRBs are tasked with reviewing all studies which involve human subjects (or data from human subjects) for compliance with human subjects regulations -- either the "Common Rule", or the corresponding FDA regulations. These regulations are based on three ethical principles that were delineated in Belmont Report -- beneficence, justice, and respect for persons. If they focus on these principles, the IRB can best achieve its main purpose of safeguarding the rights and welfare of the participants. As such, the IRB can be more objective than the researchers about the true level of risks possible and what are reasonable protections. Larry Greenfield acknowledged that one area in which researchers can benefit from the IRB's perspective is in developing effective ways for interviewers to handle upset respondents.

I'd like to take a closer look at the DHHS human subjects regulations – or the "Common Rule". It is often said that this regulation was written with only biomedical studies in mind. This is not entirely true. The Common Rule does not distinguish between social, behavioral, and biomedical research and it was problems in both medical and behavioral research that prompted the regulations in the first place. In fact, the regulations were written to be intentionally vague – for two reasons. First, they allow the IRB to be flexible and to make judgments based on the specifics of an individual study. An example of flexibility is that an IRB can waive some elements of informed consent for some studies. Second, they allow the IRB to apply the Common Rule to any study, whether it is clinical trial or a household survey about health habits. Jennifer Madans noted that CDC does both types of studies, which further supports the need for a single set of regulations and system of protections. As she said, it is possible to fit oval pegs

into round holes, but more guidance is needed to help the IRBs apply the Common Rule to social and behavioral research.

The speakers have pointed to deficiencies in our system for ensuring citizen privacy in research. I recognize these issues and have also heard them from RTI researchers. How can we improve it? First, we must affirm our shared commitment to do ethical research and to protect our respondents. We need to design good studies and assess the risks and threats to data security imposed by increasingly sophisticated computerized data collection and management. But most importantly, we need to find ways to encourage more collaboration and openness between IRBs and researchers. I think that researchers can realize real value from IRB review of their studies. An IRB review does not need to be mysterious and it should not be arbitrary. But the review of a study by people with different perspectives really can strengthen the protections that are given to respondents. There are regulations that need to be followed by the IRB, and although they leave some room for interpretation, their purpose is clear. I agree that, although the IRB system is not perfect, if does offer some structure and validity to the review of studies that involve human subjects. Since the IRB and the researchers have the same overriding goal of doing good research that respects the rights of participants; it makes sense for them to work collaboratively towards this goal.

Session 12

Obtaining Respondent Cooperation

Response Rate Achieved in Government Surveys:
Results from an OMB Study

Ruey-Pyng Lu
Energy Information Administration

Purpose of Report

The purpose of this report is to make recommendations to OMB for processing information collection Requests (ICRs) from the Agencies. We recommend that OMB adopt a formula based on historical experience for calculating expected response rates. When a proposed collection has a predicted response rate of 80% or less, the sponsor must provide detailed information showing why it expects a higher response rate and must identify specific steps it will take to maximize the response rate, also a plan of assessing the nonresponse bias.

Introduction

This study is to gain an understanding of the achieved response rate in federal government surveys. A proper response rate form is used to calculate actual response rate of surveys approved by Office of Management and Budget (OMB) in 1998. After analyzing the response rates characterized and tabulated by various factors, a standard for OMB use in reviewing information collection requests (ICRs) is recommended. These analyses also provide recommendations to agencies predicting response rates and how to assess nonresponse bias.

Predicting the Response Rate

This section estimates a rule for predicting a proposed survey's response rate based on the response rate of similar surveys conducted in the past. The rule is restricted to general-purpose statistics for individuals and organizations within the United States. The basic idea is the predicted response rate depends upon the kind of organizations collecting the data, whether it is a survey or Census, voluntary or mandatory and the target population (individuals/households, establishments or government), how the data is collected (personal interviews, mail, telephone, or multimode).

The Sample

This project studies the "unit response rates" of statistical Information Collection Requests (ICRs) approved in 1998. When an agency submits an ICR to OMB for review, there are several items in the OMB Form 83-I (see Appendix A-1) to be completed by the agency to fulfill the requirements of Paperwork Reduction Act. The ICRs, except customer surveys, with following three characteristics were selected to make up the universe of this study:

1. "**General purpose statistics**" was marked as "primary" in item 15 - Purpose of information collection.

2. "**Individuals or households**", "**Establishment** (Business or other for-profit, Not-for-profit institutions, Farms)" or "State, Local or Tribal **Government**" was marked as "primary" in item 11 - Affected public.

3. "**Voluntary**" or "**Mandatory**" was marked as "primary" in item 12 - Obligation to respond.

There were 130 Information Collection Requests (ICRs) that met this criteria, see table 1.

Table 1.	**Item 12:** Obligation to respond		
Item 11: **Affected public**	**Voluntary**	**Mandatory**	**Total**
Individuals or households	26	5	31
Establishments (Business or other for-profit, Not-for-profit institutions, Farms)	38	40	78
State, Local or Tribal Government	19	2	21
Total	83	47	130

The agency contact of all these ICRs were asked to provide the information in the OMB Response Evaluation Form (see Appendix A-2) and Data Collection Mode Form (see Appendix A-3). Several agencies (Census Bureau, NCHS, NCES, and EIA) bundled similar collections together in one ICR, in fact we have a final 216 potential information collections and they are used to evaluate response rates. See table 2.

Table 2:	**Information collections**
Eligible and returned	199
Ineligible	17
Total	216

Among these 216 collections, 10 were classified as ineligible by the agency, because they did not have the characteristics specified in this study; And the analyst excluded another 7 collections from this study. Detailed information about these ineligible collections is provided in Appendix B. Finally, a total of 199 information collections were analyzed in this study.

Methods

Factors impact the Response rate

The following six items were collected for each survey,

1. c = Number of respondents completing the survey.
2. e = Number of potential respondents that were eligible but did not respond.

3. i = Number of potential respondents that were identified as being ineligible.
4. u = Number of potential respondents of unknown eligibility.
5. x = Estimated proportion of potential respondents of unknown eligibility that are eligible.
6. n = Total number of potential respondents in the survey (or population, if for census)

where n = c + e + i + u ; If x is not provided, then the best estimate of x = (c + e) / (c + e + i) will be used to calculate response rate, and our target variable, the response rate, RR, is calculated as c / (c + e + xu) .

If a weighted response rate was used in the information collection, the weighted response rate formula was provided for further research. Response rates were computed separately for several different characteristics:

1. **Primary function of the agency submitting the ICR** (ICSP: Interagency Council of Statistical Policy vs Non-ICSP);
2. **ICR type** (Census vs Survey);
3. **Obligation to respond** (Voluntary vs Mandatory);
4. **Affected public** (Households vs Establishments vs Government).
5. **Collection mode:** (Self-administered questionnaires; Personal or group interviews [including CAPI]; Mail survey; Telephone interview [including CATI]; Multi-mode [combination of the above and other methods].)

The combination of these factors will also be used to investigate their impact on response rate.

Across these 199 surveys, the mean unweighted response rate is 82.2%; and the median unweighted response rate is 84.7%.

The distribution of calculated unweighted response rate is displayed as

Table 3: Calculated unweighted response rate	Percentage achieved
Above 90%	37.7%
Above 80%	65.8%
Above 75%	73.4%
Above 70%	80.4%
Above 60%	90.0%
Above 50%	95.5%

There are 68 Surveys have the calculated response rate below 80 percentage. The above tables show that
65.8% (131/199) of ICRs have the calculated response rate above 80 percent;
73.4% (146/199) of ICRs have the calculated response rate above 75 percent; and
80.4% (160/199) of ICRs have the calculated response rate above 70 percent.

Single factor analysis
Six factors were characterized to evaluate their impacts on response rates. They are

I. Primary function of the agency submitting the ICR: ICSP (Agency is a member of Interagency Council of Statistical Policy) or Non-ICSP. The Members of the Interagency Council on Statistical Policy are: Economic Research Service (ERS); National Agricultural Statistical Service (NASS); Bureau of Economic Analysis (BEA); Bureau of Census (BOC); National Center of Education Statistics (NCES); Energy Information Administration (EIA); National Center of Health Statistics (NCHS); Bureau of Justice Statistics (BJS); Bureau of Labor Statistics (BLS); Bureau of Transportation Statistics (BTS); Statistics of Income (SOI); Environmental Protection Agency (EPA); National Science Foundation (NSF); and Social Security Administration (SSA).

The distribution of the Response Rate by "Primary function of the agency submitting the ICR" is

Table 4: **Primary function the agency**			
RR (no decimal)	ICSP	Non-ICSP	Total
[30%, 40%)	2	0	2
[40%, 50%)	5	2	7
[50%, 60%)	6	5	11
[60%, 70%)	17	2	19
[70%, 75%)	11	3	14
[75%, 80%)	13	2	15
[80%, 90%)	49	7	56
[90%, 100%]	56	19	75
Total	159	40	199
Average response rate	82%	82.8%	82.2%

II. ICR type: Census vs Survey; The distribution of the Response Rate by (Census or Survey) is

Table 5. **ICR type**			
RR (no decimal)	Census	Sample	Total
[30%, 40%)	0	2	2
[40%, 50%)	0	7	7
[50%, 60%)	0	11	11
[60%, 70%)	0	19	19
[70%, 75%)	2	12	14
[75%, 80%)	1	14	15
[80%, 90%)	2	54	56
[90%, 100%]	13	62	75
Total	18	181	199
Average response rate	92.5%	81.8%	82.2%

III. Obligation to respond: Voluntary vs Mandatory; the distribution of the Response Rate is

Table 6. **Obligation to respond**			
RR (no decimal)	Mandatory	Voluntary	Total
[30%, 40%)	2	0	2
[40%, 50%)	5	2	7
[50%, 60%)	2	9	11
[60%, 70%)	10	9	19
[70%, 75%)	4	10	14
[75%, 80%)	7	8	15
[80%, 90%)	18	38	56
[90%, 100%]	39	36	75
Total	87	112	199
Average response rate	82.8%	81.7%	82.2%

IV. Affected public: Households vs Establishments vs Government; the distribution of the Response Rate is

Table 7. **Affected public**				
RR (no decimal)	Individuals or households	Establishments	Government	Total
[30%, 40%)	0	2	0	2
[40%, 50%)	1	5	1	7
[50%, 60%)	6	5	0	11
[60%, 70%)	4	15	0	19
[70%, 75%)	5	9	0	14
[75%, 80%)	4	11	0	15
[80%, 90%)	13	34	9	56
[90%, 100%]	16	51	8	75
Total	49	132	18	199
Average response rate	79.6%	82.2%	88.9%	82.2%

V. Data collection mode:

The categories of "Data collection mode" are:

- Self-administered questionnaires;
- Personal or group interviews (including CAPI);
- Mail survey;
- Telephone interview (including CATI);
- Multi-mode (combination of the above and other methods)

The distribution of the Response Rate by "Data collection mode" is

Table 8:	**Data collection mode**					
RR (no decimal)	Self-administered questionnaire	Personal Interview	Mail Survey	Telephone interview	Multi-mode	Total
[30%, 40%)	0	0	2	0	0	2
[40%, 50%)	0	0	7	0	0	7
[50%, 60%)	0	1	4	2	4	11
[60%, 70%)	0	0	14	0	5	19
[70%, 75%)	0	3	3	1	7	14
[75%, 80%)	0	3	6	3	3	15
[80%, 90%)	4	3	29	2	18	56
[90%, 100%]	0	10	26	3	36	75
Total	4	20	91	11	73	199
Average response rate	85.3%	84.7%	78.2%	80.2%	86.6%	82.2%

Conclusion

Recommendation to OMB about Response Rates

From the above analyses, the following statement about response rate is recommended to OMB in reviewing information collection requests:

1. Agencies that submit ICRs with expected response rate of 80% or more should provide a complete description of how they arrived at the expected response rate.

2. Agencies that submit ICRs with expected response rates between 60% and 79% should provide a complete description of how they arrived at the expected response rate, a detailed description on steps they will take to maximize the response rate; and a discussion of how they plan to evaluate nonresponse bias.

3. Agencies that submit ICRs with expected response rate of less than 60% should generally not expected the ICR to be approved. However, agencies can, on occasion, justify conducting the information collection depending on the purpose of the study, the population being studied, past experience with response rates when studying this population, plans to evaluate nonresponse bias, and plans for an aggressive survey methodology to achieve at least 60%.

Maximizes response rates

Sponsors of surveys with predicted response rates of less than 80% should describe what they are doing to increase response by specific reference to the listed attributes of high response surveys. A sample of efforts is summarized from those information collectors achieved RR above 90%. Some tips for accomplishing this include:

For any survey:

(1) Send a letter in advance to inform respondents of what, why, who, and how.
(2) Provide a toll-free phone number for respondents to verify legitimacy of survey.
(3) Provide information about the survey on the agency's website.
(4) Address confidentiality and anonymity.
(5) Use more than 1 collection mode: e.g. face-to-face interviews in non-telephone households.
(6) A threat of fines for noncompliance is mentioned in the mailout letter (for mandatory surveys only).
(7) Keep the survey brief.
(8) Make sure the survey is easily understood by respondents.
(9) Try to make the content of the survey relevant to respondents.
(10) Offer small incentives up front or a prize for responding (see Incentives FASQ).
(11) Identify strategies for contacting hard-to-reach populations.
(12) Obtain endorsement of the survey by relevant organizations.
(13) Conduct outreach sessions with presentations in several cities.
(14) Provide news releases to trade journals, state associations, and other interested parties.
(15) Maintain contact with respondents between waves in a longitudinal survey through birthday cards and postcards to be used in case of address changes.

For personal visit and telephone surveys:

(16) Increase the number and timing of contact attempts.
(17) Increase the length of the field period.
(18) Enhance interviewer training.
(19) Use senior, experienced interviewers to do "refusal conversion."

For mail surveys:

(20) Follow-up the first mailout with a phone contact (or a Fax if a business).
(21) Use Priority mail.
(22) Address letters to specific individuals.
(23) Send reminder/Thank you cards.
(24) Send replacement questionnaires as part of nonresponse follow-up.
(25) Follow-up phone calls to second mailing (or a Fax if a business).
(26) Allow respondents to complete the survey on the web or via phone.

For internet surveys:

(27) Use e-mail for advance notification, reminders, and follow-ups.
(28) Allow respondents to complete the survey on a hardcopy (to mail in) or via phone.
(29) Follow-up nonresponses with phone contact (or Fax if a business).

The evaluations conducted to assess the impact of possible non-response bias were requested in the OMB response rate evaluation form. A number of statements were reported and summarized. These are recommendations to agencies how to assess nonresponse bias.

Evaluations conducted to assess the impact of possible non-response bias

Examples of activities for information collectors that failed to meet agency's target response rate:
A large-scale evaluation was performed to examine nonresponse bias in the 1998 NAMCS. It used information from the master files to compare respondents with nonrespondents and it used information from a nonresponse mail back follow-up to compare respondents and nonrespondents. The study found several interesting results. (1) Break off was most likely to occur at the stage of the telephone screener (43 percent) and that often the refusal is from the office staff rather than the physician. This is consistent with information from the nonresponse follow-up that shows that a majority of nonresponding physicians do not remember being contacted about NAMCS. (2) A comparison of cooperation rates for many variables including physician specialty, gender, age, geographical region, board certification, MSA status, and type of practice found that only type of practice had varying cooperation rates. Physicians in group practices where the physician was part-owner were less likely to participate compared with solo practice physicians or physicians in group practices where they were an employee or contractor. (3) Effects of the observed differential nonresponse were not found to have much effect on the visit statistics produced by the survey as the nonresponse adjustment factor which takes physician specialty, region, and MSA status into account, reduced the nonresponse bias for most of the visit statistics examined.

Considerable resources were committed to studying potential nonresponse bias. A summary has been provided. 1. A comparison of estimates to extant data sources was conducted. 2. A Chi-square automatic Interaction Detection (CHAID) analysis was conducted on an extensive set of variables to determine which, if any, were significantly related to nonresponse.

(1) Use of the nonresponse methods discussed in item 7 (i.e., prioritizing nonresponse to follow-up on larger companies) of Appendix A-2 results in a weighted response rate much higher than the unweighted response rate. While the unweighted response rate is approximately 80% of the eligible companies, the respondents are estimated to account for over 95% of the data of interest; (2) For the nonrespondents, Agency imputes using previously reported data adjusted for changes and general trends in the industry.

Two types of analyses were done on the faculty data to assess the impact of possible nonresponse bias in the faculty data. First we compared respondents to non-respondents by various frame variables such as sex, race, and employment status. Here we found that there was no significant difference in the unweighted response rates of men (70%) and women (72%). There were small differences by race: the highest response rate was for Asian/Pacific Islander (68%) and the lowest was for Black, non-Hispanic (64%). The largest difference in response rates was between full time (76%) and part time faculty (60%). In addition to this analysis we also compared the survey responses for early responders and late responders, using late responders as proxy for non-respondents. We picked eight variables to analyze. Out of the eight, six had consistent responses regardless of how many days after the initial mailing they were completed. For two questions (mean age, and tenure status) there were slight declines in the mean age and the likelihood of having tenure status over the response period. The declines were substantially small

and concentrated in respondents from private doctoral school. In this stratum, late responders were on average about a year younger that early responders and 3% less likely to have tenure.

Usage of weighted response rate:

The distribution of the Response Rate by "Usage of weighted response rate" is

Table 9	**Usage of weighted response rate**			
RR (no decimal)	Weighted RR used	Weighted RR not used	Unknown	Total
[30%, 40%)	0	0	2	2
[40%, 50%)	0	1	6	7
[50%, 60%)	2	6	3	11
[60%, 70%)	0	5	14	19
[70%, 75%)	4	6	4	14
[75%, 80%)	1	8	6	15
[80%, 90%)	14	20	22	56
[90%, 100%]	3	47	25	75
Total	24	93	82	199
Average response rate	80.9%	85.7%	78.5%	82.2%

"Used" means this type of response rate was used in the ICR or in publication.

Recommendation to OMB on multi-stage and longitudinal surveys

When agencies submit ICR with multi-stage or longitudinal surveys, the following statement is recommended to OMB in reviewing those information collection requests:

1. Ask agency to provide expected response rate at each stage of sample.
2. Ask agency the expected total response rate taking into account of all stages.
3. Agency should always describe the steps to improve response rate.
4. Agency should include a discussion of how they plan to evaluate nonresponse bias.

The final recommendation to Form 83-I revision is to add one item for expected response rate, and ask agency to provide supporting material when submitting Information Collection Requests.

References

American Association for Public Opinion Research (May 1998). *Standard Definitions*.

Couper, M. & Groves, R. (1996). Household-Level Determinants of Survey Nonresponse. . *New Directions for Evaluation: Advances in Survey Research*, 70, 63-80.

Appendix A-1

PAPERWORK REDUCTION ACT SUBMISSION

Please read the instructions before completing this form. For additional forms or assistance in completing this form, contact your agency's Paperwork Clearance Officer. Send two copies of this form, the collection instrument to be reviewed, the Supporting Statement, and any additional documentation to: Office of Information and Regulatory Affairs, Office of Management and Budget, Docket Library, Room 10102, 725 17th Street NW Washington, DC 20503.

1. Agency/Subagency originating request

2. OMB control number ? None

3. Type of information collection (check one)

a. ? New Collection
b. ? Revision of a currently approved collection
c. ? Extension of a currently approved collection
d. ? Reinstatement, without change, of a previously approved collection for which approval has expired.
e. ? Reinstatement, with change, of a previously approved collection for which approval has expired
f. ? Existing collection in use without OMB control number

For b-f, note item A2 of Supporting Statement instructions

4. Type of review requested (check one)

a. ? Regular
b. ? Emergency – Approval requested by __/__/__
c. ? Delegated

5. Small entities
Will this information collection have a significant economic impact on a substantial number of small entities? ? Yes ? No

6. Requested expiration date

a. ? Three years from approval date b. ☐ Other specify __/__

7. Title

8. Agency form number(s) (if applicable)

9. Keywords

10. Abstract

11. Affected public (Mark primary with "P" and all others that apply with "X")

a. ___ Individuals or households
b. ___ Business or other for Profit
c. ___ Not-for-profit institutions
d. __ Farms
d. __ Federal Government
e. __ State, Local or Tribal Government

12. Obligation to respond (Mark primary with "P" and all others that apply with "X")

a. ___ Voluntary
b. ___ Required to obtain or retain benefits
c. ___ Mandatory

13. Annual reporting and recordkeeping hour burden

a. Number of respondents _____
b. Total annual responses _____
1. Percentage of these responses collected electronically _____
c. Total annual hours requested _____
d. Current OMB inventory _____
e. Difference _____
f. Explanation of difference
1. Program Change _____
2. Adjustment _____

14. Annual reporting and recordkeeping cost burden (in thousands of dollars)

a. Total annualized capital/startup costs _____
b. Total annual costs (0&M) _____
c. Total annualized cost requested _____
d. Current OMB inventory _____
e. Difference _____
f. Explanation of difference
1. Program change _____
2. Adjustment _____

15. Purpose of information collection (Mark primary with "P" and all others that apply with "X")

a. __ Application for benefits
b. __ Program Evaluation
c. __ General purpose statistics
d. __ Audit
e. __ Program planning or Mgmt.
f. __ Research
g. ___ Regulatory or compliance

16. Frequency of recordkeeping or reporting (check all that apply)

a. ☐ Recordkeeping b. ☐ Third party disclosure
c. ? Reporting
1. ☐ On Occasion 2. ☐ Weekly 3. ☐ Monthly
4. ☐ Quarterly 5. ☐ Semi-annually 6. ? Annually
7. ☐ Biennially 8. ☐ Other (describe) _____

17. Statistical Methods

Does this information collection employ statistical methods?

☐ Yes ? No

18. Agency contact (person who can best answer questions regarding the content of the submission)

Name:

Phone:

19. Certification for Paperwork Reduction Act Submissions

On behalf of this Federal agency, I certify that the collection of information encompassed by this request complies with 5 CFR 1320.9

Note: The text of 5 CFR 1320.9, and the related provisions of 5 CFR 1320.8 (b) (3), appear at the end of the instructions. Their certification is to be made with reference to those regulatory provisions as set forth in the instructions.

The following is a summary of the topics, regarding the proposed collection of information, that the certification covers:

(a) It is necessary for the proper performance of agency functions;

(b) It avoids unnecessary duplication;

(c) It reduces burden on small entities;

(d) It uses plain, coherent, and unambiguous terminology that is understandable to respondents;

(e) Its implementation will be consistent and compatible with current reporting and recordkeeping practices;

(f) It indicates the retention periods for recordkeeping requirements;

(g) It informs respondents of the information called for under 5 CFR 1320.8 (b) (3):

- (i) Why the information is being collected;
- (ii) Use of information;
- (iii) Burden estimate;
- (iv) Nature of response (voluntary, required for a benefit, or mandatory);
- (v) Nature and extent of confidentiality; and
- (vi) Need to display currently valid OMB control number;

(h) It was developed by an office that has planned and allocated resources for the efficient and effective management and use of the information to be collected (see note in Item 19 of the instructions);

(i) It uses effective and efficient statistical survey methodology; and

(j) It makes appropriate use of information technology.

If you are unable to certify compliance with any of these provisions, identify the items below and explain the reason in Item 18 of the Supporting Statement.

Signature of Program Official	**Date**
Signature of Senior Official or Designee	**Date**

Appendix A-2
OMB Response Evaluation Form

OMB control number _ _ _ _ - _ _ _ _

Survey ID

This survey was (mark one)

A. Completed

B. Not conducted

C. Other – please describe

If you marked "A" above, please provide the following information for the completed survey.

1. Number of respondents completing the survey.
2. Number of potential respondents that were eligible but did not respond.
3. Number of potential respondents that were identified as being ineligible.
4. Number of potential respondents of unknown eligibility.
5. Estimated proportion of potential respondents of unknown eligibility that are eligible.
6. Total number of potential respondents in the survey (or population, if for census)
7. Describe all efforts to maximize the response rate.

8. Describe evaluations conducted to assess the impact of possible nonresponse bias.

Appendix A-3
OMB Response Evaluation Form - data collection mode

OMB control number _ _ _ _ - _ _ _ _

Survey ID

1. The data collection mode in this ICR:

The data collection mode was (mark one)

___ A. Personal or group interview (including CAPI)

___ B. Mail Survey

___ C. Telephone interview (including CATI)

___ D. Some other methods (e.g. Fax, ...) , please specify

_____

___ E. Multi-mode (combination of the above)

2. Weighted response rate:

___ Yes, a weighted response rate was used in this ICR and the weighted response rate formula was _____ .

and the calculated weighted response rate in 1998 was _____ .

___ No, we did not use weighted response rate in this ICR.

Appendix B

I. Ineligibles as identified by agency

0581-0033: The program inadvertently checked the wrong purpose in box 15 on the OMB 83-1 form. This collection is not for "general purpose statistics" but "program evaluation". This collection is not a survey. The collection has only one form and the form is used to collect info from commercial dried egg products plants. The information is then assembled and summarized and distributed monthly to interested parties. This information is used by industry as an aid in making knowledgeable marketing decisions.

0607-0350: This form is just used on an "as needed" basis to collect information about new building permit issuing jurisdictions, places that may have changed their permit issuing requirements, or potential nonpermit places that may now be issuing permits.

0607-0809: The activities that were covered under the OMB generic clearance from 1998 were listing activities for the 2000 decennial census. Activities such as Address Listing and Block Canvassing were operations undertaken with the objective of building a complete address list of the United States for Census 2000. Since the activities performed under the generic clearance were listing activities and not survey related activities, there is no pertinent information to deliver regarding survey response rates.

0938-0741: This collection is a form, not a survey. Beneficiaries fill it out when they "disenroll" from a Medicare+Choice plan.

0970-0183: This is a voluntary program which provides States' Child Support Enforcement agencies upon their request access to all of the earned and unearned income information reported to IRS by employers and financial institutions. The IRS 1099 information is used to locate noncustodial parents and to verify income and employment. It is not a general purpose survey.

1110-0002: Voluntary; The Supplemental Homicide Report (SHR) is a form disseminated to the Nation's law enforcement community for the voluntary submission of detailed homicide information pertaining to their monthly homicide counts. The document is not designed to enhance the characteristics associated with each homicide reported. Response to the request for information on this form is voluntary on the part of the law enforcement community. Block 15 on OMB Form 83-I, was identified by the FBI in the option of "general purpose statistics" because the information collected is used by criminologists, sociologists, and academia for this purpose. At no time has this form been used to survey law enforcement's needs or interests in participating in the national Uniform Crime Reporting Program.

1875-0134: This is not a "survey"; it is the performance report form for the Safe and Drug-Free Schools (SDFS) State Grants program, which includes the State Education Agency and Governor's programs. The form has been used by states to provide their annual performance reports to the U.S. Department of Education for their SDFS programs.

1905-0129(3): EIA-417R; Electric Power System Emergency Reporting Procedures. This is not a statistical survey per se. Reporting is required when an electric power system undergoes one of the following types of incidents: interruption in firm power, voltage reductions, public appeals to reduce use of electricity, vulnerability actions, and fuel supply emergencies.

2040-0180: Part 132 requirements are part of already existing State Water Quality Programs. Thus, the ICR that was prepared tried to estimate any increase in burden to respondents regulated by existing State programs that may result from Part 132. As such, there is no way to determine or identify if the Part 132 burden estimated by the ICR actually resulted from Part 132 because Part 132 did not have any specific information or data collection requirements.

3045-0043: *Note that this is not a "survey". It is a Project Progress Report (PPR). AmeriCorps*VISTA sponsors (approximately 1,200) submit a PPR on a regular and ongoing basis. Required of all Sponsors, the PPR is an opportunity to document progress towards meeting the goals and objectives of their Project Work Plan. Only active AmeriCorps*VISTA project Sponsors submit the report.

II. Ineligibles identified by analyst

1505-0010, RR=0%; The purpose of the form is to provide an opportunity to comment on proposed revisions to the Foreign Currency Form FC-2.

1505-0012, RR=0%; The purpose of the form is to provide an opportunity to comment on proposed revisions to the Foreign Currency Form FC-1.

1505-0014, RR=0%; The purpose of the form is to provide an opportunity to comment on proposed revisions to the Foreign Currency Form FC-3.

1505-0088, RR=0%; The purpose of the form is to strengthen compliance with other TIC reporting requirements.

0420-0513 (1), RR=1%; This is a rolling continuous use enrollment form and is not a statistical survey. U.S. Teacher and Peace Corps Volunteer Enrollment Form.

0420-0513 (2), RR=0.46%; This is a rolling continuous use enrollment form and is not a statistical survey.

0607-0850, RR= 30.2%, United States Census 2000 Dress Rehearsal Large Household Followup; Public awareness campaign in targeted Dress Rehearsal sites.

The 2002 Response Rate Summit: Recommendations from an Expert Panel[32]

Nancy Bates
U.S. Census Bureau

Introduction

In the Spring of 1997, the Census Bureau and several of its sponsoring agencies formed a new interagency committee to address current nonresponse issues. The committee (the Interagency Household Survey Nonresponse Group or IHSNG) is a subcommittee of the Federal Committee on Statistical Methodology. The group maintains members from the Bureau of Transportation Statistics, the Energy Information Administration, the Bureau of Labor Statistics, the U.S. Census Bureau, the Office of Management and Budget, the National Center for Health Statistics, the National Center for Education Statistics, and the Bureau of Justice Statistics (see www.fcsm.gov/committees/ihsng/ihsng.htm for more information).

As a means of stimulating new ideas for how best to research and evaluate nonresponse in government surveys, the IHSNG sponsored an expert panel. A Response Rate Summit convened on February 21st and 22nd in Arlington, Virginia. The purpose of the summit was to provide a forum for discussion among experts in the field about how to address concerns related to the decreasing response rate trend in household surveys. The subcommittee chose two national surveys, the National Health Interview Survey (NHIS) and the Consumer Expenditure Quarterly (CEQ) survey, as the basis for discussion during the summit because they represent a cross-sectional and a longitudinal household survey, respectively.

In this paper, we summarize the discussions and recommendations of the expert panel and conclude with a preliminary follow-up of the panel's number one recommendation: the collection and analysis of detailed contact history/call record data.

The Summit

The Summit lasted two days and included eleven panel members with expertise in survey methodology, survey sampling, and survey operations[33]. In addition, staff from the Census Bureau, Bureau of Labor Statistics, and National Center for Health Statistics also attended and

---

[32]This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau publications. This report is released to inform interested parties of research and to encourage discussion.

[33] Expert panel members included: Roger Tourangeau (JPSM), Paul Biemer (Research Triangle Institute), David Cantor (Westat), Ed Cohen (Arbitron), Robert Groves (Univ. of Michigan), Graham Kaltom (Westat), Daniel Kasprzyk (Mathematica Policy Research), Paul Lavrakas (Nielsen Media Research), Stanley Presser (Univ. of Maryland), Eleanor Singer (Univ. of Michigan) and Brian Williams (Statistics Canada).

participated in the discussions. On the first day, representatives from the data collection and sponsoring agencies made a series of presentations to familiarize panel members with the NHIS and CEQ methodologies and response rate histories. At the end of the day, a Census Bureau interviewer and a survey supervisor made presentations to describe the challenges of gaining respondent cooperation in the field.

On the second day, the panel members engaged in open discussion to brainstorm ways to increase the surveys' response rates. They looked at the system used to administer the survey, the organization of the survey itself, and the workload of the field representatives. Using open discussion, reviews of current literature, and field experience, the group thought of a variety of different ways in which response rates can be increased for the NHIS and the CEQ.

The discussions revolved around two main questions: How can we get a higher response rate and how can we lower noncontact rates? During the course of the day, panel members offered various solutions and techniques for reducing nonresponse. Some topics of discussion included front loading cases for more efficient field assignments, better use of call record histories during contact attempts, use of respondent and interviewer incentives, refusal aversion training for interviewers, changes to the advance contact materials, and changes to the life-cycle of the survey itself. At the end of the day, the Summit facilitator presented a list of twenty-two recommendations and suggestions. Each panel member was asked to prioritize the most important ideas of the day and rank them one through three. The top five recommendations in order of rank are presented briefly below.

#1 - Call Records

In the context of the panel conversations, the term "call records" was used loosely to refer to any means of gathering histories on contact and contact attempts. It was noted by some that the practice of collecting good contact histories was lost (or at least diminished) in some organizations during the transition from paper and pencil interviewing (PAPI) to computer assisted personal interviewing (CAPI).

There was unanimous agreement that the collection of call record information would benefit interviewers, regional offices, and headquarters. Some panel members were concerned about when interviewers should open their laptops and if they would bother to open them if the respondent was not at home. Every member of the panel seemed to agree that more specific outcome codes for noninterviews needed to be developed. For example, was initial contact never made or was there initial contact but no follow-up? Would you call a follow-up contact that was never reached a "not-at-home," a "broken appointment," or a "temporarily unavailable"? Data collection on refusals was another recommendation. Classifying refusals by demographic and attitudinal variables of those who refused and by physical and social environment variables about the living unit and the neighborhood could provide valuable information for reducing refusals.

By and large, the group found call records to be very valuable tools to better discern refusals and noncontact, develop an interview history of contacts, design targeted materials, and understand motivations for noncooperation. Consequently, the number one recommendation from the

Summit was to collect and analyze detailed contact/call record data. Such data have a twofold purpose: (1) for analytic purposes, and (2) as a feedback mechanism for regional offices and interviewers. It was suggested that agencies devise an automated system to capture case histories—including number of contacts, interim outcomes (outcome of each contact/attempt), reasons for refusals, day of week and time of contact/attempt, and demographics of refusal/noncontact households. The data could also be used in real time by the supervisory staff and senior interviewers to quickly identify and aid interviewers having problems and to determine optimal callback patterns.

#2 - Refusal Aversion Training

The second-ranked idea was that of refusal aversion training. This training promotes a transfer of the skills from the very best interviewers to less experienced interviewers providing a strong knowledge of the survey and explanations for specific concerns that reluctant respondents have. It teaches interviewers how to listen for these concerns and immediately counter them (see Groves and McGonagle, 2001). There was a consensus reached that interviewers would be better off and more prepared if they were trained in how to respond effectively to refusals from respondents in an effort to persuade them to participate. The panel agreed that training interviewers how to react was better than sending them into the field with a pre-scripted list of responses to possible refusals. Most of the group liked this idea; however, some speculated that it would only help the lower-tier interviewers and waste the time of the more experienced interviewers.

The panel recommended expanding the use of this training method beyond an experiment conducted in 2001 in the NHIS[34]. The panel suggested building a model that allows senior interviewers in the regions to administer the actual training. Additionally, members recommended continual tracking of interview-level response rates to further refine the training and test whether the training has a diminishing effect over time.

#3 - Incentives and Motivators

The topic of incentives ranked third among the top five recommendations. Many panel members were of the mind that incentives would generate increased response rates, however there was considerable debate as to how incentives should be applied. Another general observation made was the most motivated interviewers also had a propensity to be the best interviewers. Thus incentives and motivators must work in the direction of both the respondent and the interviewer. While thinking about incentives and motivators, members of the group were able to narrow their focus to the perspective of the respondent and ask, "What's in it for me?" A general outline of the dialogue follows.

34 (For more information on this experiment, see O'Brien, Mayer, Groves and O'Neill, 2002)

Here is what we know about incentives:

1. There are two types: conditional and unconditional.
2. There are two subtypes: monetary and nonmonetary.
3. Incentives are more complicated to give in government surveys (because of OMB approval).
4. Incentives are applicable for both respondents and interviewers.

Suggestions as to what conditional and unconditional incentives should be used for respondents included:

1. Give a little up front, the rest upon completion (e.g., split incentives).
2. Only give incentives to refusals to get them to convert.
3. Give everyone an incentive in the advance letter.
4. Only give to those who complete the survey.

Suggestions as to what incentives/motivators could be used for interviewers:

1. Goal sharing.
2. Performance statistics based on response rates, consent rates, refusals, etc.
3. Exponential bonuses (especially for refusal conversions).
4. Training for refusal conversion.

Suggestions made as to what the monetary and nonmonetary incentives could include:

1. Social and personal benefits derived from uses of the data collected.
2. Civic duty.
3. Material goods such as patriotic paraphernalia.
4. Fixed dollar amounts.
5. Variable compensation schemes.
6. Debit card in the advance letter and the four-digit code at completion.

One argument raised against refusal conversion incentives was the frustration it causes interviewers. They consider it paradoxical to reward someone for being a problem case, thus lowering morale. The panel recommended conducting experiments with the CEQ and the NHIS to explore unconditional incentives for respondents. The incentives do not necessarily have to be monetary but should be survey-specific. For example, provide respondent bonus for full compliance (complete entire diary, successful completion of all survey sections) or provide up-front monetary incentive with promise of more to come if they cooperate. At the same time, consideration should be given to developing an incentive program to motivate interviewers.

#4 - Changes in the Survey Design

The fourth-ranked recommendation involved specific changes to the CEQ and NHIS survey design. Discussion began with an analysis of the life cycle of NHIS. There were two general groups: those who thought that NHIS should maintain its weekly sampling procedures for comparability and those who thought that it would decrease burden to do a monthly or biweekly sample. Some NCHS staff argued the sampling process should remain as it is for comparability across years.

The next major life cycle discussion was about the length of the field period and its impact on response rates and noncontact rates for NHIS. Everyone agreed that the field period was short, which led to decreased response rates and increased noncontact rates. To lengthen the field period, a suggestion to add an extra weekend to the field period was made.

Someone suggested that both the CEQ and NHIS should reduce the complexity and length of their questionnaires. A panel member remarked that it had been his experience that too many surveys were designed without taking into account the concerns of the survey takers. Introducing matrix sampling was one idea mentioned to decrease the length and complexity of the surveys. Evaluating the usefulness of each question was another suggestion to reach the same goals.

In summary, the panel suggested the NHIS should experiment with a monthly or biweekly sample design. It could also expand the interview period to include a third weekend and/or spread out the caseload. For the CEQ, consider using a clustered sample design, then estimate impact of clustering on response rates. Additionally, consider a redesigned coordination of sampling and workload.

#5 - Improve Advance Contacts and Other Materials

The fifth ranked suggestion was to make improvements to the advance materials. The discussion began with a thorough examination of the current methods used for advanced contacts. Most of the group agreed that an advance letter is helpful but the panel strongly recommended that instead of a "face lift" to the materials, we revolutionize the advance contact protocol.

Citing low readership of advance letters, most agreed that if there are required statements (like collection authority statutes), these should be kept in the letter. Cosmetic changes suggested to increase readership included adding jazzy graphics and changing the font.
It was agreed that the more detailed information should be displayed in a brochure or Frequently Asked Question (FAQ) format rather than the advance letter. Content changes suggested including more personalization—such as the business card of the field representative, making the language of the letter less bureaucratic, and tailoring the letter to dispel possible objections.

Other issues about advance letters that were raised included who should sign the letter, the sponsor or the Census Bureau? Should multiple advance letters be sent? Should letters be customized for each region? Should we use FedEx to deliver advance letters? No decisive consensus was reached on the aforementioned questions. Everyone agreed that experiments should be conducted before reaching a conclusion.

The importance of the advance letter to the response rate and the interviewers' comfort level was discussed as well. It was unanimous an advance letter is important for interviewer comfort because it makes the house calls seem less like "cold calls." In the words of one interviewer, "It gives us a reason to be there."

Additionally, panel members deliberated over whether having phone numbers of respondents would help get a leg up on advance contacts. Those against advance contacts argued that the advance phone call or letter could give the respondent more time to think of reasons not to participate. Most agreed that if someone is going to refuse, they do so regardless of the advance contact. The panel suggested conducting a controlled experiment to best understand the impact that advance phone calls might have on unit nonresponse.

Moving Forward on Call History Recommendation

Following the Summit, the co-chairs of the IHSNG presented the recommendations to their subcommittee and at the Census Bureau survey sponsors quarterly meeting. Several initiatives are underway to follow-through on the Summit suggestions generated. For example, efforts are underway to get a program of refusal aversion training implemented nationwide in the NHIS. Additionally, the Census Bureau is currently developing a stand-alone prototype instrument designed to collect detailed call record data in the NHIS. This system, deemed the Contact History Instrument or CHI, hopes to be in production for the NHIS by 2004 with additional personal-visit demographic surveys to follow. A CHI steering committee has been formed to oversee the development of CHI for longitudinal/panel surveys and to ensure that the IHSNG and agency sponsors provide adequate input during development (see Oneto, 2002 for more information on both initiatives).

SIPP Methods Panel Contact History Logs

Shortly after the Summit, the Survey of Income and Program Participation Methods Panel (SIPP MP) decided to sponsor a short-term research project to explore the use of contact history logs. The SIPP MP is an experimental survey sponsored by the Census Bureau and carried out in six regional offices. It is designed to test improvements and alternative measurement approaches for the core SIPP instrument. For each experiment, a random sample of approximately 1,500 addresses receive the experimental MP instrument and another 1,500 receive the control SIPP instrument (the production SIPP). The majority of MP interviews are conducted in-person by computer assisted personal interview (approximately 80%) and the remainder are completed by telephone.

Based largely on discussions from the Summit, we decided to introduce a paper and pencil contact history log into the MP field procedures. This allowed a quick means of collecting detailed information on contact and contact attempts during a Census Bureau personal-visit survey. The contact logs were printed front and back on heavy weight colored paper. One side contained a grid for interviewers to record the date, time and mode of contacts along with the interim and final contact outcomes and comments from each contact or contact attempt. The

other side contained instructions for completing the logs and a list of final outcome codes (see attachment 1).

Interviewers were instructed to complete a record of the personal visit or telephone call each time they attempted to contact a household. Interim outcome codes were divided into Contact and Non-Contact categories and included subcategories such as: eligible household member not home, language problem, respondent too busy, respondent refused, household did not answer door, unable to reach - gated community, telephoned - busy signal, telephoned - answering machine, etc. When the case was deemed 'complete' the final outcome code was also recorded on the contact history logs (e.g., completed interview, no one home, household refused) along with the date and time.

Interviewers used the logs for the first time during Replicate 3 Wave 1 of the 2002 SIPP MP which took place in July and August, 2002. At the time of writing, most of the July logs were available for analysis, but the August cases had not yet been keyed. Additional analyses of both months are planned once all data become available. The July SIPP MP sample consisted of a total of 1,587 cases and at the time of writing, we had contact log data from approximately 91% of these cases. A series of random 20% sample checks of the data entry suggest an error rate of approximately 0.8% for the July logs.

In the discussion that follows, we set about answering a few basic questions previously unknown in the absence of interim contact histories. Namely, we want to know the average number of contacts/contact attempts for completed interviews and non-interviews, the workload distribution and status after each successive contact/contact attempt, and the interim status distributions leading up to cases resulting in an interview, a 'no one home', and a refusal. We also sought to answer basic questions about the time of day and day of week interviewers are making contacts resulting in successful interviews versus 'no one home'. These represent only a fraction of the potential research questions the contact logs will eventually yield.

To set the context for the contact log analysis, it is important to first report the overall response rates for the 2002 SIPP MP. In the month of July, the survey achieved a response rate of 87.2%[35]. Nonresponse cases were comprised of 7.5% refusals, 2.1% 'no one home' and 3.2% residual nonresponse comprised of unable to locate, language problems, and an 'other' noninterview category.

Chart 1 illustrates the mean number of contact/contact attempts by region for cases ultimately classified as a completed interview, no one home, or refusal. This average reflects both personal visit and telephone contacts. The bar chart clearly shows a trend whereby completed interviews required the lowest number of contacts. There is some variation across regions but for the most part, the average number of contacts is around 3.5 for cases resulting in an interview.

Cases that ultimately got classified as a refusal required more effort – here the overall average

35 The response rate is calculated using the American Association for Public Opinion Research Response Rate 2 (RR2) definition (AAPOR, 2000).

number of contacts was closer to 6. In the Charlotte region, the average number of contacts for refusals climbed above 7 (7.4 contacts). However, the number of refusals is rather small when broken out by region (approximately 17 cases per region), so it will be important to re-visit these means once the August data are available. Likewise, while the trend appears that 'no one home' cases required by far the greatest amount of contacts (8.6 overall), these cases are relatively rare and, in fact, means are not shown for the Atlanta and Seattle regions as they each had only one such case. But, regardless of the limitations, the chart begins to tell the story of how much time and effort is being expended for different classes of outcomes.

Chart 2 illustrates the case status by each successive contact/contact attempt. The base for the chart is restricted to cases determined to be in-scope and eligible for the survey (i.e. placed in the denominator of the response rate formula). According to Chart 2, around 22% of the eligible workload became completed interviews after just one contact (noted by the blue band on the first bar). The majority of cases remain active after one visit and a small fraction are deemed non-interviews after just one visit (lightest portion of the bars). At twenty visits, all of the active cases finally became classified as an interview or non-interview.

Chart 3 graphs the number of completed interviews by contact number. Of all interviews eventually conducted, the greatest number are performed during the first and only visit, the number drops steadily with each successive contact/contact attempt until the line starts to flatten out around the eighth contact. Chart 4 illustrates the percentage of completed interviews relative to the cases attempted at each contact. The horizontal red line illustrates the average percent of completed interviews across all contact attempts (21%). The percent of completed interviews are slightly above average during contacts 1-3 and slightly below during the fourth and fifth contact. A simple 'eyeball' of the chart suggests the average starts to drop around the sixth or eighth visit – roughly the same points we saw the number of interviews start to flatten out on Chart 3. Additional analysis should be performed to explore the cost/benefits of extending the maximum number of contacts beyond the 6-8 contact range. In their study of call record data from a face-to-face survey in the UK, Purdon, Campanelli, and Sturgis (1999) report that contact becomes less likely as the number of calls increases. The trick is finding the point at which additional calls begin to result in diminished returns, that is, the likelihood of getting a completed interview drops significantly.

Charts 5 and 6 explore what is occurring at the contact/contact attempts leading up to a final case code disposition. Chart 5 combines the 31 interim status codes into 6 major categories:

- Personal Visit Contact - no interview,
- Personal Visit - no one home,
- Personal Visit - fluttering curtain,
- Personal Visit - physical/environmental barrier,
- Telephone Attempt - no contact, and
- Other - noncontact.

The 'fluttering curtain' pertains to personal visit attempts where a respondent did not answer the door but interviewers reported evidence that someone was at home. Such situations represent a

grey area – should a no one home or a soft refusal be recorded? The 'barrier' category refers to those situations where a personal visit was attempted but contact was not made due to a physical barrier (gatekeeper, buzzer entry, gated community) or an unsafe environmental element (dogs, crime, drugs).

Far and away the most common outcome leading up to a final disposition is to make a personal visit but find no one at home (accounted for approximately 50% of all interim outcomes). Second behind the no one home category is the situation of finding someone at home, but not obtaining an interview due to a variety of reasons including no eligible member available, respondent too busy, or respondent initially refuses. Both the 'fluttering curtain' and barrier situations occurred far less frequently.

Chart 6 presents the distribution of selected interim situations by final disposition of the case. For example, the blue line graphs the frequency of situations leading up to cases that ultimately became successful interviews. The pink line charts cases that were classified as 'no one home' and the green line charts cases finally transmitted as refusals. The most common scenario leading up to all three cases was a personal visit, no one home, albeit it was more common in cases that received a final code of no one home. In fact, the top three interim outcomes for no one home cases are not very surprising with the most frequent being personal visit , no one home; the second being telephone, no contact; and the third a personal visit, environmental/physical barrier.

The most common scenario leading up to a completed interview was a personal visit, no one home followed by a personal visit, respondent too busy followed by telephone, no contact. Prior to classifying a case as a refusal, the most common contact attempt was a personal visit, no one home followed by telephone call, no contact followed by personal visit, respondent too busy. It is interesting to note that the interim category of 'respondent refused' was reported in less than 10% of the attempts leading up to cases ultimately determined to be refusals. It is also noteworthy to mention that environmental/physical barriers rarely appear to predict cases that ultimately became refusals. More likely, these impediments increase the number of contacts and lengthen the time needed to either secure an interview or close out a case as a 'no one home.'

Chart 7 graphs the distribution of contacts by day of week and time of day. Specifically, it illustrates the day/time distribution when interviews were successfully completed (the blue line) contrasted with the distribution of attempts that resulted in an interim 'no one home' status (pink line). Day of week and time of day are combined into 7 categories separating weekdays from weekends and dividing time of day into four major spans. The most successful day/time combination for interviews was weekdays between 5:00-9:00 pm (36%) followed by weekdays between noon and 5:00 pm (28%). Saturdays and Sundays between noon and 5:00 pm yielded another 10% of interviews.

Contact attempts resulting in 'no one home' were tried most often during a weekday between noon and 5:00 pm. Considering this is the second most successful time/day combination for interviews, the data suggest interviewers are being pretty efficient with their visits. However, additional efficiencies might be gained if field representatives shifted slightly away from this slot

and concentrated more visits in the 5-9:00 pm weekday period.

Conclusion

The Response Rate Summit proved to be a good catalyst for formulating research initiatives to deal with nonresponse. While many of the ideas were not new, the panel provided a forum for academia, survey managers, production staff, and field personnel to discuss a common problem and possible solutions. As a result of the Summit, the participating agencies are moving forward on at least two of the top five recommendations. These include the design of an automated call history instrument for face-to-face surveys and the wider implementation of a special refusal aversion training for interviewers. We are grateful to the panel members who were willing to attend the Summit and lend their expertise on the vital topic of increasing participation in government surveys.

Acknowledgments: I would like to thank Pat Doyle, Adele Alvey, Elaine Hock and Liz Griffin for technical assistance with the data and help in conducting this research. I would also like to acknowledge Nancy Cioffi for keying the contact history logs.

References

American Association for Public Opinion Research (2000). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Survey.* Ann Arbor, Michigan: AAPOR.

Groves, R. M. And McGonagle, K. (2001). A Theory-Guided Interview Training Protocol Regarding Survey Participation. *Journal of Official Statistics*, Vol. 17, No. 2, pp. 249-265.

O'Brien, E.M., Mayer, T., Groves, R.M., and O'Neill, G.E. (2002). Interviewer Training to Increase Survey Participation. A paper presented at the American Statistical Association annual meetings, New York, N.Y.

Oneto, A. (2003). National Health Interview Survey Response Rates: Influences and Interviews. A paper presented at the FCSM 2002 Statistical Policy Seminar, November 6-7, Bethesda, MD.

Purdon, S., Campanelli, P. And Sturgis, P. (1999). Interviewers' Calling Strategies on Face-to-Face Surveys. *Journal of Official Statistics*. Vol. 15, No. 2, pp. 199-216.

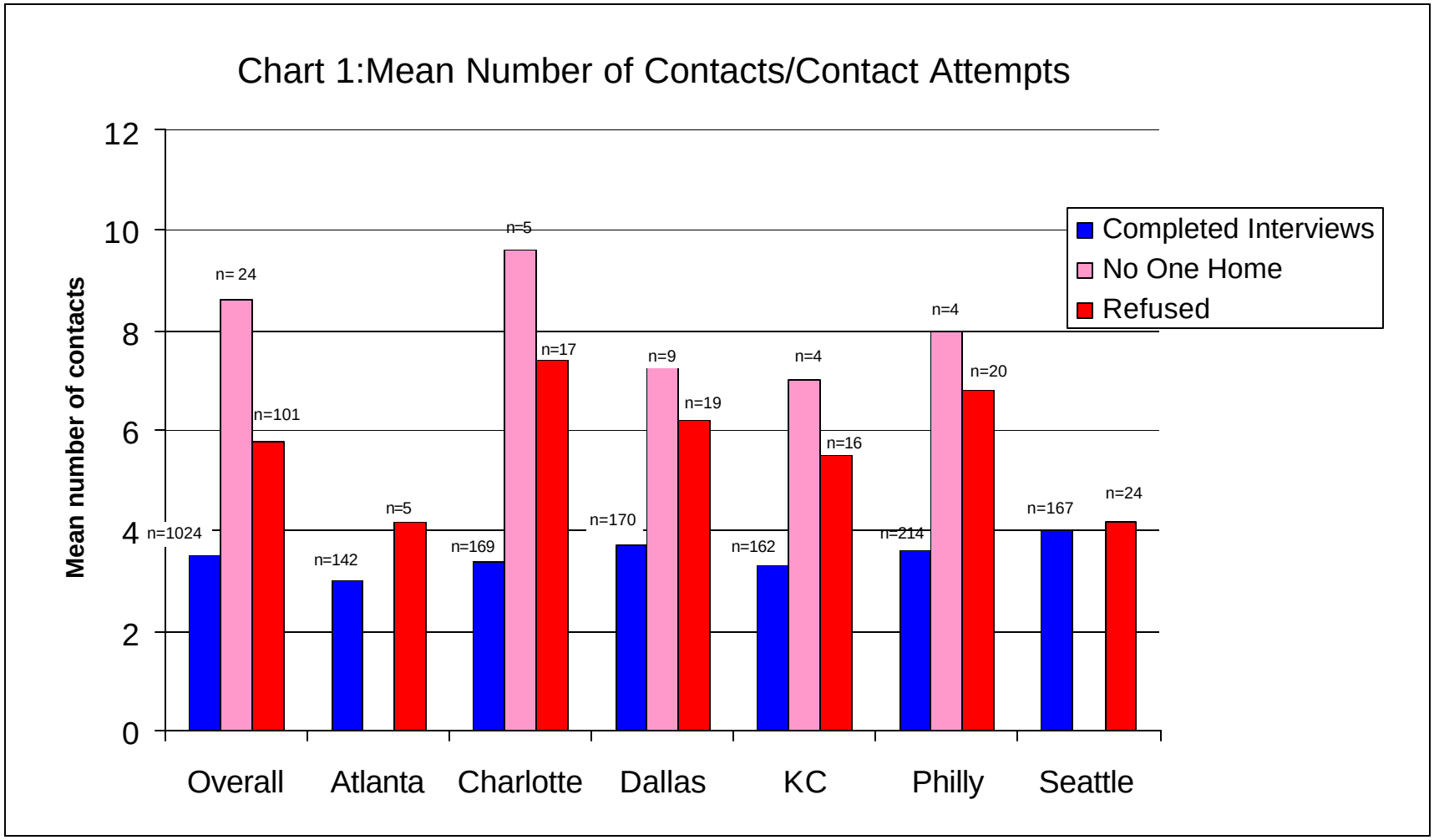


Chart 1:Mean Number of Contacts/Contact Attempts
Mean number of contacts
12
10
8
6
4
2
0
Completed Interviews
No One Home
Refused
n=1024
n= 24
n=101
n=142
n=5
n=169
n=5
n=17
n=170
n=9
n=19
n=162
n=4
n=16
n=214
n=4
n=20
n=167
n=24
Overall
Atlanta
Charlotte
Dallas
KC
Philly
Seattle

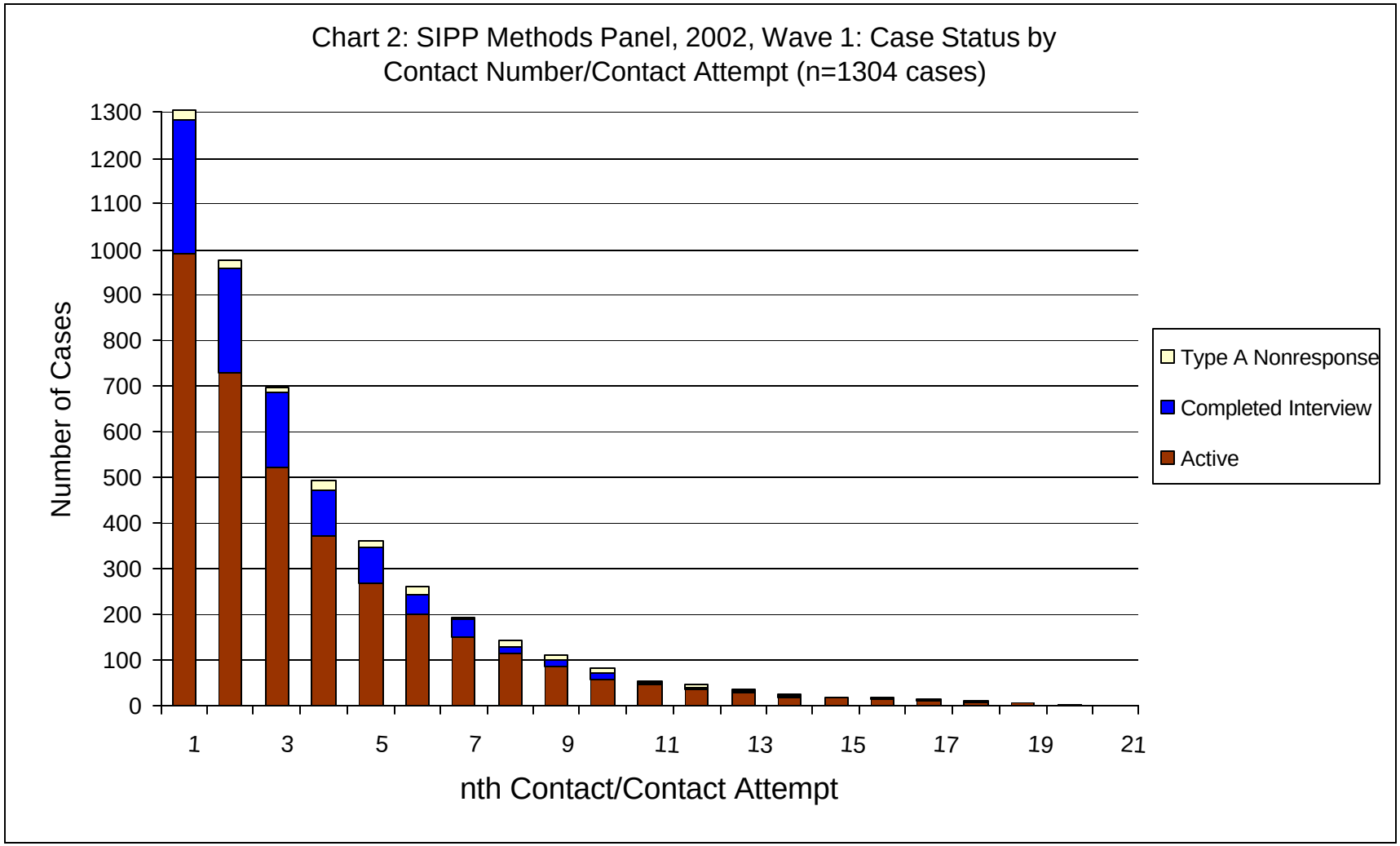


Chart 2: SIPP Methods Panel, 2002, Wave 1: Case Status by Contact Number/Contact Attempt (n=1304 cases)
Number of Cases
1300
1200
1100
1000
900
800
700
600
500
400
300
200
100
0
1
3
5
7
9
11
13
15
17
19
21
nth Contact/Contact Attempt
Type A Nonresponse
Completed Interview
Active

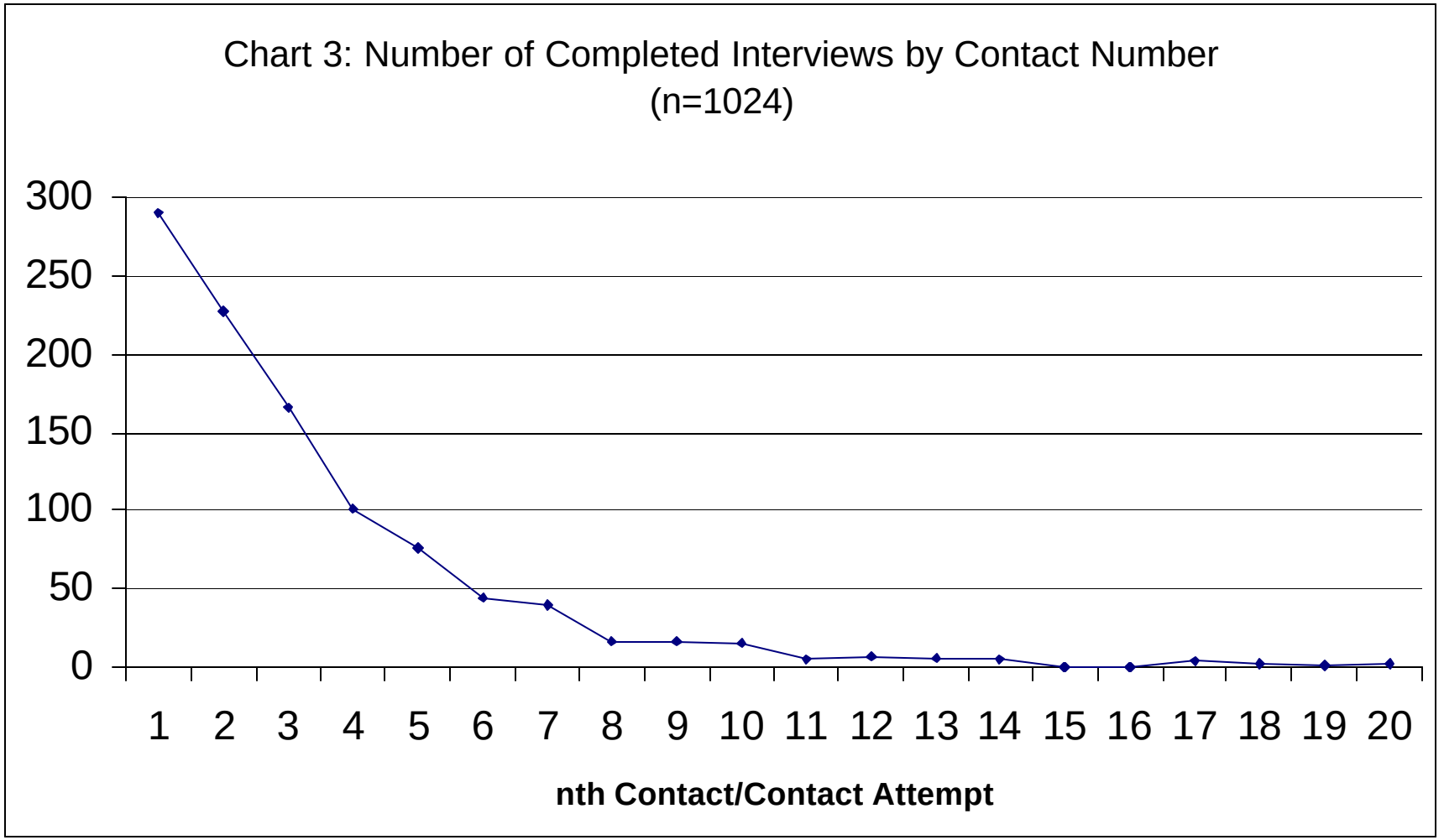


Chart 3: Number of Completed Interviews by Contact Number
(n=1024)
300
250
200
150
100
50
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
nth Contact/Contact Attempt

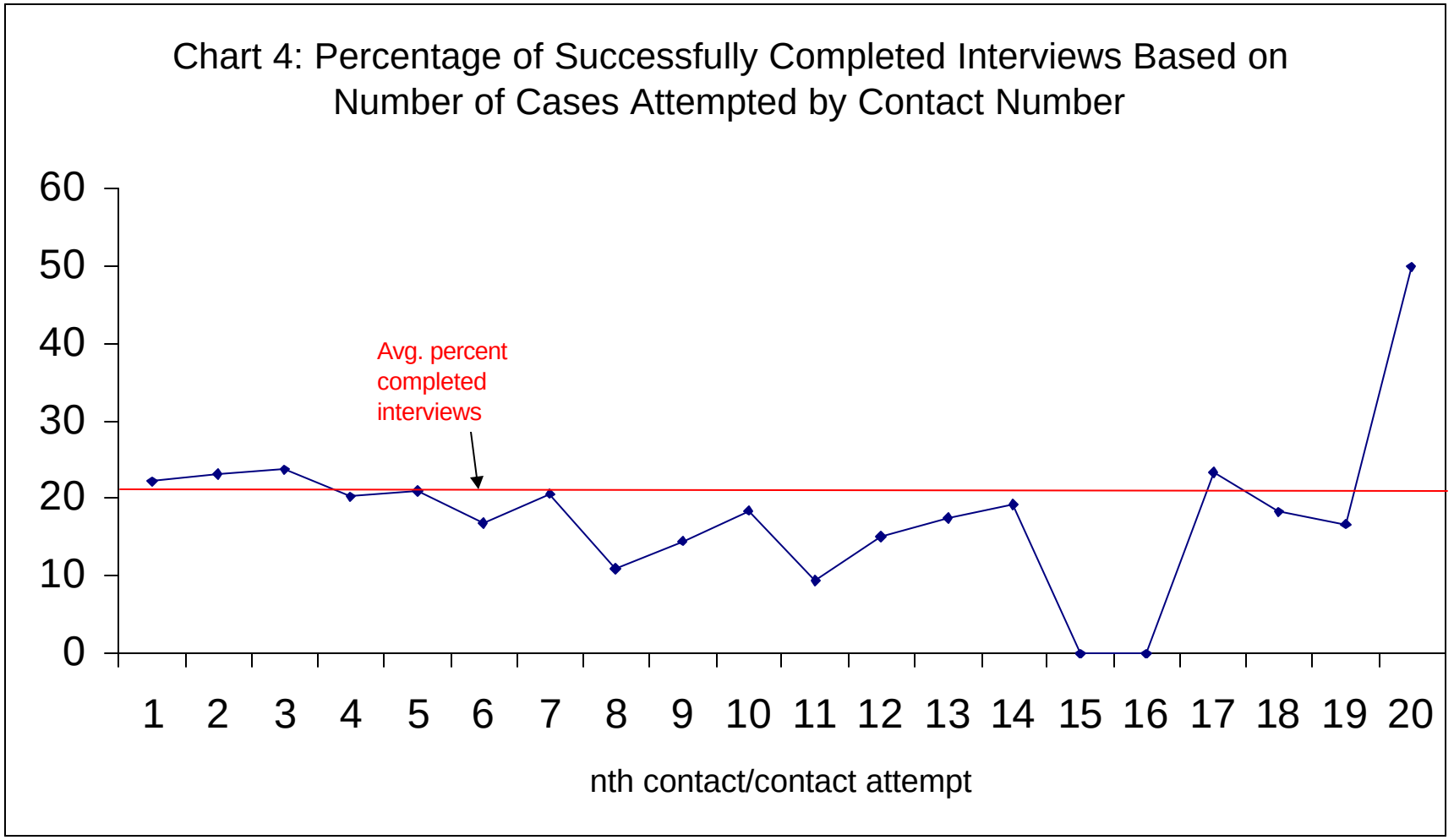
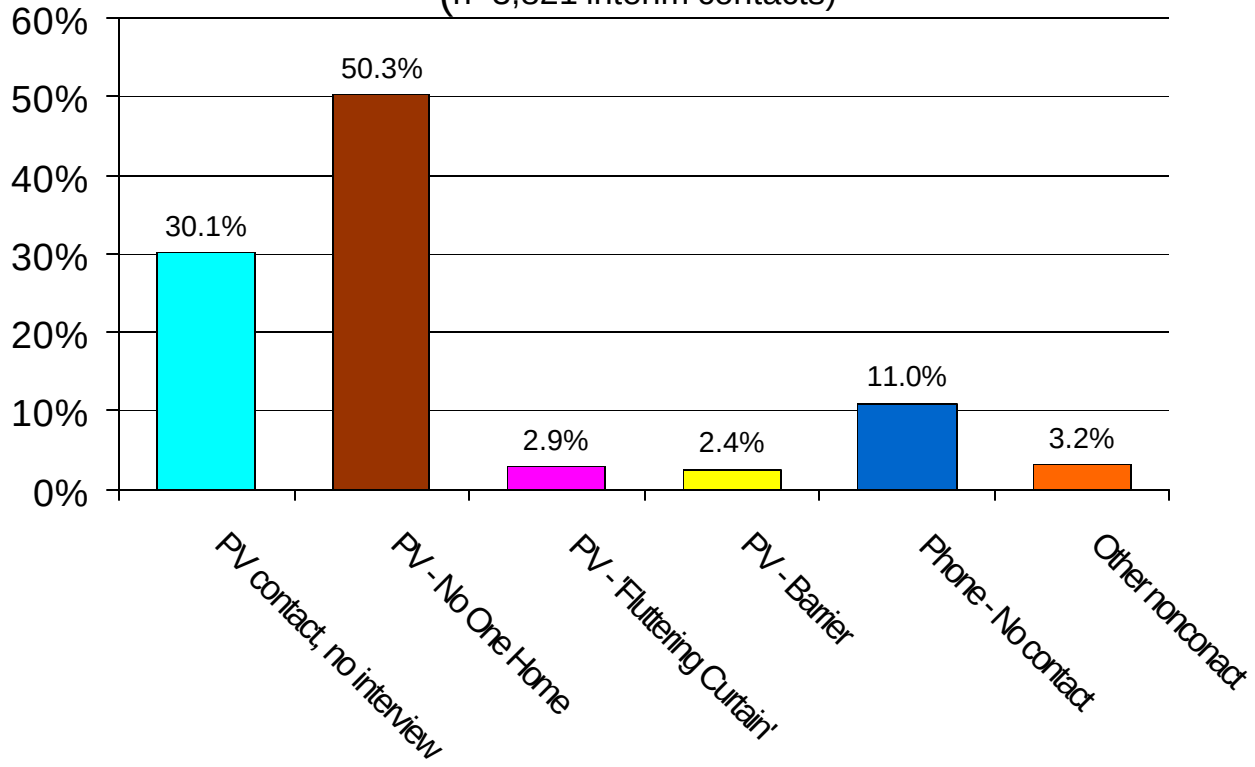


Chart 4: Percentage of Successfully Completed Interviews Based on Number of Cases Attempted by Contact Number
60
50
40
30
20
10
0
Avg. percent completed interviews
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
nth contact/contact attempt

Chart 5: SIPP Methods Panel, 2002 Wave 1:
Interim Status Distribution Prior to Final Disposition
(n=3,821 interim contacts)



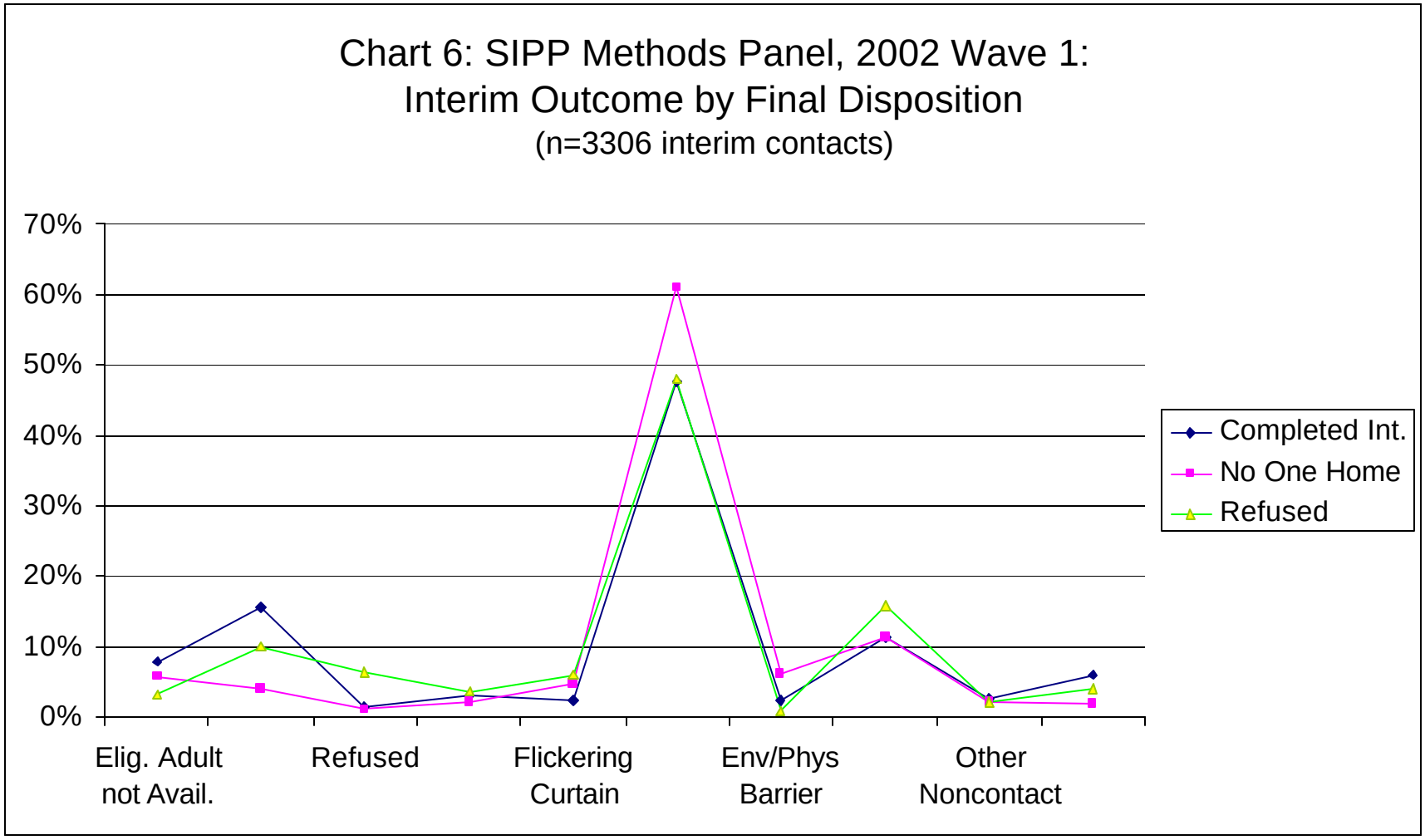


Chart 6: SIPP Methods Panel, 2002 Wave 1:
Interim Outcome by Final Disposition
(n=3306 interim contacts)
70%
60%
50%
40%
30%
20%
10%
0%
Elig. Adult not Avail.
Refused
Flickering Curtain
Env/Phys Barrier
Other Noncontact
Completed Int.
No One Home
Refused

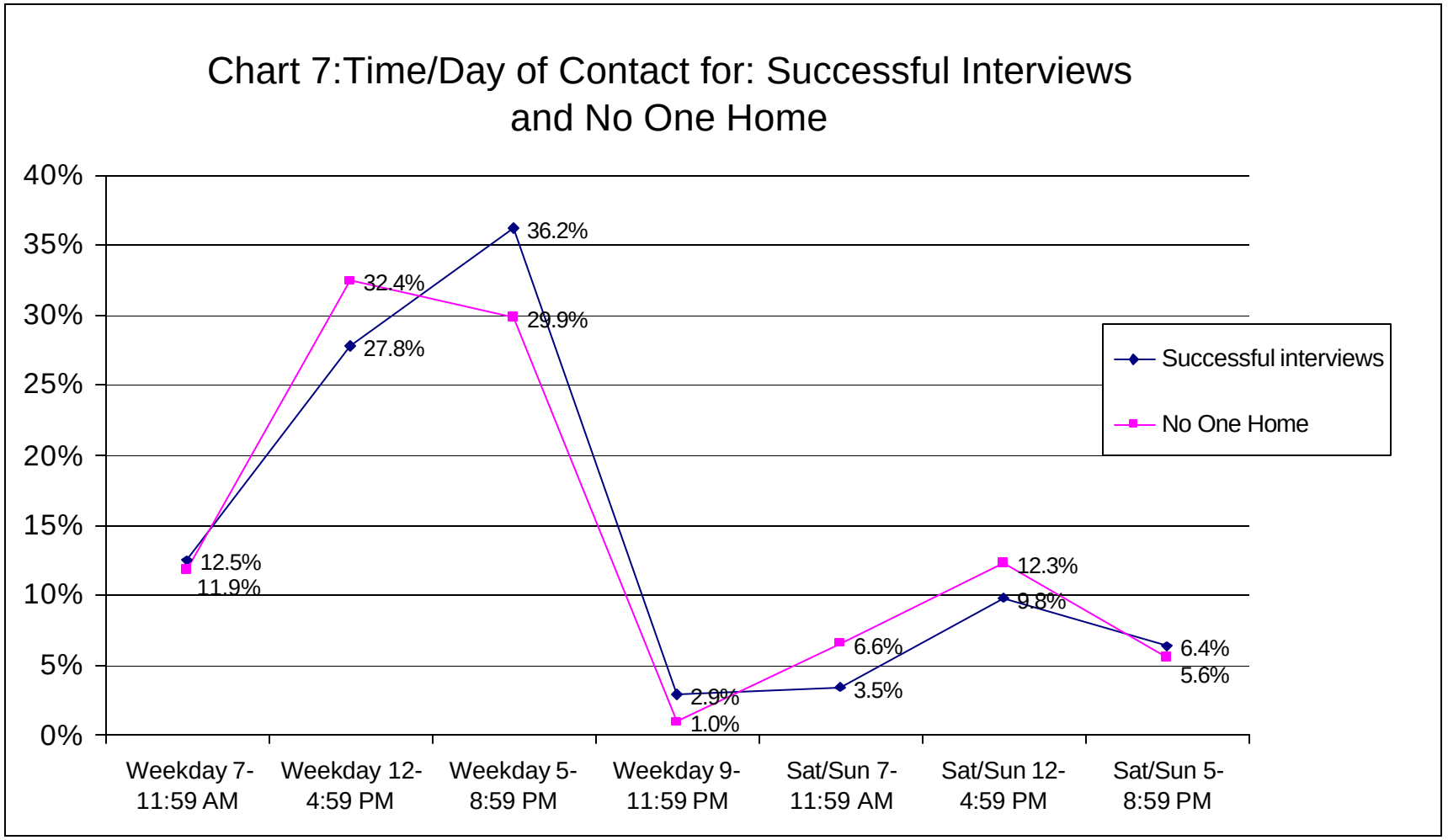


Chart 7:Time/Day of Contact for: Successful Interviews and No One Home
40%
35%
30%
25%
20%
15%
10%
5%
0%
36.2%
32.4%
29.9%
27.8%
12.5%
11.9%
12.3%
9.8%
6.6%
6.4%
5.6%
2.9%
3.5%
1.0%
Successful interviews
No One Home
Weekday 7-11:59 AM
Weekday 12-4:59 PM
Weekday 5-8:59 PM
Weekday 9-11:59 PM
Sat/Sun 7-11:59 AM
Sat/Sun 12-4:59 PM
Sat/Sun 5-8:59 PM

2002 MPSIPP WAVE 1 CONTACT LOG

1. (S)FR Code:_____
2. Control Number:_____
3. HU Address:_____

4. RO *(circle one)*: ATL CHA DAL KC PHI SEA
5. Sample (*circle one*): Test Control
6. Case ID Number:_____

7. (S)FR's Record of Personal Visits/Telephone Calls

Date **(a)**	**Time** ***(Write in hour and circle am or pm.)*** **(b)**	**Personal Visit/ Telephone Call (*circle one)*** **(c)**	**Outcome Code of Contact** **(d)**	**Comments** **(e)**
	am / pm	PV Tele		
	am / pm	PV Tele		
	am / pm	PV Tele		
	am / pm	PV Tele		
	am / pm	PV Tele		
	am / pm	PV Tele		
	am / pm	PV Tele		
	am / pm	PV Tele		
	am / pm	PV Tele		

MPSIPP WAVE 1 INTERIM OUTCOME CODES

Contact Code	Definition
01	Eligible HH Member Not Home
02	Language Problem - No Translator in HH
03	Eligible HH Member Hospitalized
04	Eligible HH Member has Mental Problems
05	Death in the HH
06	Respondent too Busy, Appointment Set
07	Respondent too Busy, Unable to Set Appointment
08	Appointment Broken - Rescheduled
09	Appointment Broken - Not Rescheduled
10	Respondent Refused - Requested Survey Info
11	Respondent Refused - Followup Required
12	Partial Interview - Followup Required
13	Instrument Problems
14	Other [Specify in Column (e)]

****

Non-Contact Code	Definition
15	HH did not answer door, even though there is evidence someone is at home
16	No One Home, Left a Note
17	No One Home, No Note left
18	Informed that HH is on Vacation
19	Informed that HH is on Business Travel
20	Informed that HH is Seldom Home
21	Unable to Reach - Unsafe Area (Drugs/Crime)
22	Unable to Reach - Unsafe Area (Dogs/Animals)
23	Unable to Reach - Gated Community
24	Unable to Reach - Locked Gate
25	Unable to Reach - Buzzer Entry
26	Telephoned - Got a Busy Signal
27	Telephoned - Left a Message on Machine
28	Telephoned - Disconnected, Wrong #, FAX
29	Telephoned - No Answer
30	Telephoned - Call Blocked
31	Other [Specify in Column (e)]

INSTRUCTIONS FOR COMPLETING THE CONTACT LOG

ITEM	SPECIFIC INSTRUCTIONS
1	Enter your (S)FR code.
2	Enter the control number from the "Control Number" column on the case list screen.
3	Enter the HU address from the "Address" column on the case list screen.
4	Circle the appropriate regional office. (*Only circle one*.)
5	Circle "Test" if the case has a "T" in the T/C column on the case list screen. Circle "Control" if the case has a "C" in the T/C column on the case list screen.
6	Enter the case ID number. Highlight the appropriate address on the case list screen and press the F3 function key. The case ID number is in the upper, right-hand corner of the F3 screen.
7a	Enter today's date.
7b	Enter the time that you visited **or** telephoned the case. Then circle the appropriate time of day: if before noon, circle "am" **OR** if after noon, circle "pm."
7c	Circle "PV" if you made a personal visit to the case. Circle "Tele" if you contacted the case by telephone.
7d	For each contact (other than the final contact), enter the two-digit interim outcome code from the list at the bottom of the Contact Log Form. On the final contact line, enter the three-digit outcome code from the list below.
7e	Further explain interim outcome codes 14 and 31, or any additional comments you feel are necessary.

Note that you must complete a line on the contact log every time you visit or telephone a case.

MPSIPP WAVE 1 FINAL OUTCOME CODES

Code	Definition
INCOMPLETE CASES	
200	New case - not started
202	Started - no interview or insufficient partial
204	Partial - Missing data (callback items)
206	Partial - 1 or more persons incomplete (not missing callback items)
COMPLETED CASES	
201	Completed interview
203	Complete partial - Missing data; No Type Zs; No followup
207	Complete partial - Type Zs; No further followup
TYPE A NONINTERVIEWS	
213	Language problem
214	Unable to locate
215	Insufficient partial
216	No one home (NOH)
217	Temporarily absent (TA)
218	HH Refused
219	Other occupied (specify)

Code	Definition
TYPE B NONINTERVIEW	
223	Entire HH Armed Forces
224	Entire HH age 14 or under
225	Temporarily occupied by persons with URE
226	Vacant - Regular
227	Vacant - storage of household furniture
228	Unfit, to be demolished
229	Under construction, not ready
230	Converted to temp. business or storage
231	Unoccupied tent or trailer site
232	Permit granted, construction not started
233	Other (specify)
TYPE C NONINTERVIEWS	
240	Demolished
241	House or trailer moved
242	Outside segment
243	Converted to perm. business or storage
244	Merged
245	Condemned (and unoccupied)
246	Built after April 1, 1990 (does not apply to Area frame - Non-Permit)
247	Unused serial number or listing sheet line
248	Other (specify)

National Health Interview Survey Response Rates: Influences and Interventions

Adrienne Oneto and Lindsey Dougherty
U.S. Census Bureau U.S. Census Bureau

Introduction

The National Health Interview Survey (NHIS) is a national, cross-sectional household interview of the United States noninstitutionalized civilian population. The NHIS, which has been conducted annually since 1957, obtains information about the health status, the amount and distribution of illness, its effects in terms of disability and chronic impairments, and the kind of health services people receive. Annually rotated supplements gather data on special, emerging topics such as: AIDS, cancer risk factors, immunization, disability, family resources, access to care, health care coverage, mental health, and alternative medicines. The NHIS is sponsored by the National Center for Health Statistics (NCHS).

The survey has extensive field activities, operating out of 12 regional offices (ROs) across the nation. A new data collection cycle begins every January. Refresher training for the NHIS Field Representatives (FRs) is conducted during the first two weeks of January every year. New hires are trained throughout the year to replace attriting FRs. The NHIS sample is distributed to the FRs weekly, with a nationwide average of 1460 cases per week, or approximately 122 cases per RO. Each FR works two assignments each month, and the average assignment size is 20 cases. The "interview week" begins on a Monday and spans 16 days.

In 1990, the NHIS household response rate was 95.6%, compared with the 2001 response rate of 89.3% (Table 1). To analyze this trend, we concentrate on the "Type A" rate, which measures nonresponse for households that are eligible to be interviewed. The "Type A" nonresponse rates over the same time period have inclined from 4.4% in 1990 to 10.7% in 2001 (Table 2). A larger portion of that increase is attributable to refusals rather than the "no one home" or "temporarily absent" rates, or other traditional "Type A" reasons (Table 3).

Most of our demographic survey response rates reflect the complexities of an increasingly difficult survey environment and a more challenging respondent landscape. Contributing to the problem is a nationwide disenchantment with surveys, a somewhat negative perception of the federal government, a greater number of gated communities, heightened concerns about security and privacy, and the economic challenges of personal visit interviewing. A number of different influences have contributed to NHIS nonresponse, and multiple interventions are being put in place to address response rate issues.

The National Health Interview Survey Response Rates: Influences

A series of events have influenced NHIS response rates throughout this period spanning 1990 through 2001 (Table 4). For example, in 1993 six alternating weeks of interviewing were cancelled in the third quarter, enabling FRs to concentrate more on lighter workloads. In 1994 and 1995, FRs dealt with two large supplements resulting in an interview that averaged an hour and a half; stacks of paper questionnaires proved daunting for FRs and respondents. Also in 1995, the survey switched to a 1990-based sample design, the government was shutdown for four weeks, and the NHIS implemented a screening operation in order to increase the reliability of certain minority statistics.

The NHIS saw major operational changes in 1996 when it began the switch from Paper Assisted Personal Interview (PAPI) to Computer Assisted Personal Interview (CAPI). The CAPI pretesting took place in 1996, and the NHIS was in dual-mode PAPI/CAPI interviewing for the entire year. Nationwide CAPI interviewing was introduced in 1997, replete with a revamped instrument, computerized random sample person selection, and navigational complexities. An Informed Consent Process was introduced in 1999 by the NCHS' Institutional Review Board. The procedure calls for all survey respondents to sign a consent form - indicating that they have been fully informed about the NHIS prior to the beginning of the interview - and agree to participate. Many FRs report that the Informed Consent Process detracts from techniques they are trained to use to be persuasive in capturing an interview during that tenuous initial contact episode. The 2000 Census provided added publicity and a heightened sense of civic duty. Most recently, the events of September 11, 2001, imbued Americans with a pronounced sense of patriotism but also an increased need for security.

The National Health Interview Survey Response Rates: Interventions

Several initiatives are being discussed and implemented to ensure that, given the current survey environment, positive response rate influences are operationalized for the NHIS.

A. The Inter-agency Household Survey Nonresponse Group

The Inter-agency Household Survey Nonresponse Group (IHSNG) was formed in 1997, under the directive of Chet Bowie of the U.S. Census Bureau and Clyde Tucker of the Bureau of Labor Statistics. The mandate of the IHSNG is to examine nonresponse in major U.S. Governmental Surveys, including: the NHIS, the Consumer Expenditures Surveys, the Current Population Survey, the Survey of Income and Program Participation, the National Crime Victimization Survey, and more recently, the American Community Survey. Specifically, the goals of the IHSNG are to document nonresponse trends, quantify the consequences of nonresponse for the quality of survey estimates, and identify nonresponse causes and their potential solutions. The IHSNG maintains multi-agency participation, and is an extremely effective conduit for keeping nonresponse issues center stage.

B. Regional Office Remedies

In 2001, headquarters staff consulted with the Regional Directors to discuss concerns about household and sample adult response rates, partial interview rates, and survey costs. A series of

suggestions and recommendations resulted in a summary of ideas for our sponsor, NCHS, to consider (Table 4). Some of the suggestions for increasing response require less lead time to implement, while others require significant lead time and would incur additional costs. In addition, some of the long-term proposals involve changing unique features of the survey design. A Response Rate Summit in 2002 provided expert discussion of these remedies and recommended strategies for prioritization and implementation.

C. Response Rate Summit

Under the sponsorship of the U.S. Census Bureau and the IHSNG, a Response Rate Summit was convened in 2002. The purpose of the summit was to provide a forum for discussion among experts in the field about how to address concerns related to the decreasing response rate trend in household surveys. The NHIS was one of the two surveys chosen, since it represents national, cross-sectional household surveys. The following discussion summarizes NHIS progress on the top five summit recommendations.

1. Collect and Analyze Detailed Contact Record Data. The Response Rate Summit panel of experts gave a strong recommendation for the Census Bureau to develop a system that collects a case history of contact. Although case history information is reported for CATI surveys and for many prior PAPI surveys, the data are scattered in various files or not captured at all, and analysis has been sporadic at best. A Contact History Instrument (CHI) is being designed that can be called from case management as an external program, or automatically launched from the data collection instrument. The CHI routinely and systematically captures detailed quantitative and qualitative information on the nature of each contact attempt for survey cases. The program will automatically update dates, day, time, and maintain a cumulative counter of contacts. There will be a composite CHI record available as a management tool for both survey supervisors and supervisory field representatives during real-time data collection, to identify and assist FRs experiencing problems. CHI data will also be available for post-data collection analysis, to help survey managers better discern the patterns and reasons for nonresponse, and to design materials and procedures that support response achievement. Eventually, CHI is expected to be enhanced through an evolutionary process as it is implemented for longitudinal and mixed-mode surveys.

2. Implement Refusal Aversion Training. In 2001, we piloted the Census Response Achievement Field Training (CRAFT) for the NHIS, out of the Dallas and New York ROs. First, a round of teleconferences was conducted to generate a list of environmental cues, respondent characteristics, visual nonverbal cues, and verbatim respondent concerns that are associated with respondent reluctance specific to the NHIS. A second round of focus groups was conducted in person, and successfully generated situation-specific, tailored rebuttals that are effective in addressing respondent reluctance. Training materials were developed in August of 2001. A handbook was created to describe theory, list themes of reluctance, and provide examples of verbatim respondent concerns along with appropriate verbatim rebuttals. Exercises were created that involved active listening, role playing, and more. Training in both ROs focused on five steps to encourage survey response: prepare for the visit; engage in active listening; diagnose the respondent's main concern; quickly identify a situation appropriate response; and then quickly deliver a clear and brief response. In November, a Dallas and New

York debriefing teleconference produced recommendations for several training modifications: less lecture, more time and flexible use of exercises, small groups, and follow-up.
Census staff are currently coordinating a nationwide implementation of CRAFT for NHIS. An hour-long Computer Based Training "home study" lecture component will precede the condensed version of the classroom training. CRAFT is scheduled for a pretest in June 2003, train-the-trainer sessions in November 2003, and nationwide implementation in January 2004.

In June 2000, a Blue Ribbon Task Force was convened to reengineer FR training to better equip our FRs to be successful in the current survey environment. The Interviewer Curriculum was developed as a series of generic modules using a short mock survey that all demographic surveys can use for initial training. This Blue Ribbon Interviewer Training will be piloted in 2003 for all Consumer Expenditure Survey new hires.

3. Promote Incentives for Respondents...Motivators for Interviewers. Monetary incentives were suggested as response motivators for NHIS interviews, particularly to address the partial response rate. Concerns about this recommendation are financial: the NHIS interview can have up to four respondents, and the control process for distributing and tracking incentives poses an additional cost. Still, incentives may be more cost effective than repeated callback attempts to reach sample adult or sample child respondents. Further research is needed to determine the optimum scenario for awarding incentives to respondents. Also, the NHIS budget is currently constrained by an ongoing reengineering project, and the upcoming 2000-based sample redesign.

An often repeated request from the Field staff to modify the way we currently handle screener cases in our production measures would certainly boost FR morale. Households targeted for screening and appropriately screened out could be included in a revised calculation as a screener/response measure. We need to further discuss whether this should represent an additional or a replacement calculation of completed cases, and the appropriate time to introduce this revised calculation.

The true impact of Informed Consent on the NHIS is difficult to measure, although our Field staff believes the process has negatively affected both response rates and cost. The sponsor's Institutional Review Board has agreed to discontinue the consent process in 2003 – a decision extremely well received by NHIS FRs.

4. Consider Changes in Sample Design...Survey Design. Recommendations for improving response rates involve sample design and survey design issues. Panelists agreed that the current weekly sample design creates a short field period, which can increase noncontact rates and lead to decreased response rates. We did a preliminary test in the Atlanta Regional Office for the fourth quarter of 2001 to add an additional weekend for NHIS interviewing. More specifically, we delayed the start of the NHIS interviewing week from a Monday to a Wednesday, which introduced a third weekend into the 2½-week interviewing period. Our FRs continue to report a heavy reliance on weekends to obtain a large portion of their interviews. While Atlanta FRs were very appreciative of the additional weekend, the desired increase in response rates was not achieved. Also, this reconfigured interviewing week with the additional weekend increases interviewing costs.

Another more significant sample design change recommended the substitution of a monthly sample design for the current weekly sample design. Although interviewing would occur all through the month, this design would enable our field staff to plan their itineraries more efficiently, particularly when overnight travel and per diem are involved. We would attempt to balance the work across weeks, but the flexibility to interview adjacent units that would normally be assigned in different weeks could result in significant economies of scale. A monthly sample design also mitigates the conflicts posed by multiple survey assignments. This is not an insignificant sample control change, however, and we would need adequate lead time to modify and test our systems. During initial discussions, the sponsor was not in favor of such a radical sample design change, citing data comparability for trend analysis, the periodic need to cut weeks of sample, and preservation of their ability to make weekly estimates.

Survey design recommendations addressed the length and complexity of the NHIS interview. We get feedback that respondents break off because the survey is too long, that many of the questions are too wordy and confusing, and that some question series are redundant. The longer and more complex the interview, the harder it is to deal with respondents' frustration and irritation and maintain their participation. Our FRs continue to urge us to move income and other sensitive questions to the end of the interview, to replace very detailed income questions with fewer and more straightforward questions, and to provide respondents the option of selecting from a broader income range up-front. Instrument design and interview length issues require more evaluations and proposed solutions from both agencies.

5. Redesign the Advance Contact Materials. The Response Summit panelists recommended that we revolutionize our initial contact materials. We believe that a proposed series of revisions to the Advance Letter—in style, format, and content—would result in a more favorable first contact with prospective respondents. Suggested changes include the following: a friendlier tone, a larger font, substituting median for average length of interview and moving it to the back of the letter, curtailing the emphasis given to the voluntary nature of the survey, moving the discussion of the Social Security and Medicare numbers to the back of the letter (with questions and answers), and directing all questions and concerns about the survey to the appropriate RO rather than to NCHS' Institutional Review Board's voice messaging service. We continue to work with the sponsor to improve the Advance Letter.

Panelists also suggested that engaging publications that discuss data uses would help boost respondent interest. The NHIS Promotional Packet is being revamped for 2003. The sponsor has also begun to issue early releases of their data, which we quickly get into the hands of our FRs.

D. Response Rate Summit Recommendations: Charting Our Progress

Clearly our progress on these Response Rate Summit recommendations is more advanced in some areas than others (Table 5). Both the CHI and the CRAFT endeavors are scheduled to be implemented in January 2004, coincidental with the introduction of the reengineered NHIS. We are confident with our progress on both of these initiatives. Respondent incentive initiatives are currently in conflict with budgetary concerns, particularly at this time, with both 2004 NHIS

reengineering and 2000 based sample redesign competing for funds. Clearly, the decision on the issue of incentives lies with our sponsor. The Response Rate Summit panelists made a number of excellent suggestions regarding the motivating of FRs. While we have taken some action here, we could do more. Recommendations regarding sample design and survey design require further discussion and research. Steps are being taken to improve NHIS advance contacts, but we realize that much more could be done to improve the format and contact of these advance contact materials. Under consideration are: a colorful brochure with frequently asked questions, FR business card inclusion in the advance mailed letter, issuance of thank you cards for all FRs to use, advance letters tailored to the needs of each region, and more frequent dissemination of early release data newsletters to our FRs. Response rate problems continue to challenge NHIS operations and we are appreciative of the direction and guidance provided by the Response Rate Summit panel of experts.

Obtaining Respondent Cooperation Discussion

Richard L. Bitzer
U.S. Census Bureau

Response Rate Achieved in Government Surveys: Results from an OMB Study (Ruey-Pyng Lu, Energy Information Administration)

This remarkable paper summarized the results of a survey of over 200 data collection activities submitted to the OMB for clearance in 1998. The author used an easily understood, descriptive, consistent, and straightforward response rate definition which resulted in meaningful comparisons among the selected programs. The author is congratulated for having a 100% response rate and collecting/organizing a wealth of valuable data.

These data collections were conducted, generally speaking, through 2001, but it must always be kept in mind that the interviewing landscape changes constantly as evidenced by the recent shootings in the greater District of Columbia metropolitan area and the resulting impact on respondents and interviewers. This event was concurrent with the monthly administration of the Current Population Survey and served as a reminder of the spontaneous complexities of survey administration.

The author provided a complete distribution of unweighted response rates that showed 65.8% of ALL data collections achieved a response rate of 80% or more, the recommended standard for minimal documentation to the OMB for approval. He then selected relevant factors to describe the differences in response rates according to the components of these factors, such as the survey/census design, type, and data collection medium. The author provided great insight by displaying the number of data collections within these components by the response rates achieved in 10% increments from 30% to 100%. He also provided the average response rate for the total of all data collections by factor component, but did not calculate the percentage that achieved a specific response rate as he did for ALL surveys. The percentage (by my calculation) of a selected data collections in the study that had the lowest percentage achieving at least an 80% response rate, reveals that 59% of the Individual or Households Component (Household versus Establishment versus Government Factor), 60% of the Mail Survey Component (Data Collection Mode Factor) and 45% of the Telephone Interview Component (Data Collection Mode Factor) meet the 80% standard. Others across the balance of factors/components ranged from 64% up to the high of 94% of those meeting the 80% standard by my calculation. The 94% was the Government Component (Household versus Establishments versus Government Factor). This distribution of response rates might add an interesting bit of additional information to the evaluation.

The interventions to improve response rates were comprehensive and well stated. However, with the exception of the suggestion for improved training, these suggestions were geared primarily toward program design and respondent motivation but not specifically toward interviewer motivation, morale, and support. Additional support for interviewer management strategies would complement an already impressive list of tips to improve response rates.

The 2002 Response Rate Summit: Recommendations from an Expert Panel (Nancy Bates, US Census)

The author did a detailed job summarizing the results of the summit and the relative priorities. One of the most impressive things about the summit is the involvement of the field staff that work with the survey instruments and respondents on a daily basis.
In the final analysis, it is the responsibility of managers to provide the best system for employees who do the actual work. There is no better way to get feedback to improve the system than to ask those doing the task to explain the difficulties they encounter and suggest methods to overcome them. This is exactly what done in the presence of the experts who could put this into context.

The Call Record Information was listed as the highest priority and this, certainly, deserves such a ranking. The three most important pieces of information that are needed for optimal use in managing surveys (other than actual case data during the interviewing period) are:

1. The average number of calls to complete a case by interviewer,
2. The outcomes by attempt by interviewer and,
3. The successful outcomes by interviewer by day and time.

Using these data, a survey manager can easily determine what interviewers are making too many attempts and determine how those are distributed by call attempt. Combining this with the day and time by successful outcome by interviewer, the survey manager can diagnose potential problems and suggest ways to modify interviewer behaviors to be more productive. Later in the paper, the results of call history data from the
MPSIPP Survey, show these data in the aggregate and provide a great starting place for developing a system to monitor interviewer performance.

The Refusal Aversion Training (RAT) listed second is another high priority area. Interviewers constantly ask for new ways to "convert" reluctant respondents. With the design based on feedback from experienced interviewers, this is another application of using feedback from those who are actually doing the work to promote continuous improvement.

The last three priorities, Survey Design, Improve Advance Contacts and Other Materials are good responses to common suggestions from interviewers. The very fact that interviewers suggestions are considered is a motivator by itself. In particular, interviewers always ask for simplified advance letters that are made possible by the suggestion to add a brochure with detailed information and a brief cover letter.

The Contact History data displays were invaluable. It was encouraging is see that the statistical results of the average number of contacts by final outcome (complete, refusal, and no one home) were confirmed by previous research (Purdon, S., Campanelli,P., and Sturgis,P. (1999)). The author suggested in her presentation that it would be valuable to try to determine the optimum average number of calls that should be made to a household before resulting in diminishing returns. The data seem to suggest 6-8 visits. However, I would suggest that 9-10 visits be given

some study to account for variation in different geographic areas and to preserve the highest possible response rate possible within resource constrains. The data also suggest that about 40% of completed interviews are done during the normal workday, another 40% during the weekday evenings and the balance on weekends. While the significant percentage of daytime interviews may be a function of how interviewers prefer to schedule their visits, one must remember that interviewers (US Government employees) who work after 6:00PM receive a 10% night differential salary premium. And, if an interviewer's productive time were limited to selected night/weekend hours, more staffing may be required to complete surveys by prescribed deadlines increasing training, recruiting, and administrative costs.

The author is commended for creating an excellent record of the expert proceedings and data analysis. The emphasis on input from field staff is the highlight.

National Health Interview Survey Response Rates: Influences And Interventions (Adrienne Oneto and Lindsey Dougherty, US Census)

The authors are to be commended for a excellent summary of the recent evolution of the National Health Interview Survey (NHIS). While the decline in survey response rates in general have decreased, the NHIS survey response rate decline from 95.6% in 1990 to 89.3% in 2001 may be related to other influences, both positive and negative. The authors pointed out the major influences:

1. The cancellation of six alternating weeks of interviewing in 1993,
2. The addition of two large (paper) supplements in 1994-1995,
3. The 1995 survey redesign (resulting from the previous Decennial Census),
4. The 1995 government wide shutdown,
5. The 1996 transition from paper to Computer Assisted Personal Interviewing(CAPI),
6. The completion of the CAPI transition in 1997,
7. The introduction of a written informed consent process in 1999,
8. The 2000 Decennial Census publicity, and
9. The historic 9/11tragedy.

Looking at these events along with the graph of response rates is a very compelling demonstration of how internal/external factors could easily be related to response outcomes. In all of these events, since interviewers and respondents were affected both must be considered in any treatment of response analysis.

Many of the interventions mentioned were already discussed in previous papers, particularly the primary importance of the Census Response Achievement Field Training (CRAFT referred to as RAT in the Bates paper) and the Contact History Instrument. On a specific note, caution must be used in the case of magnetic "trinkets" as respondent incentives, since they may cause problems with magnetic hard disks and other storage devices that interviewers rely on to perform their daily activities.

The author is congratulated for her research and presentation.

Reports Available in the Statistical Policy Working Paper Series

1. ***Report on Statistics for Allocation of Funds***, 1978 (NTIS PB86-211521/AS)
2. ***Report on Statistical Disclosure and Disclosure-Avoidance Techniques***, 1978 (NTIS PB86-211539/AS)
3. ***An Error Profile: Employment as Measured by the Current Population Survey***, 1978 (NTIS PB86-214269/AS)
4. ***Glossary of Nonsampling Error Terms: An Illustration of a Semantic Problem in Statistics***, 1978 (NTIS PB86-211547/AS)
5. ***Report on Exact and Statistical Matching Techniques***, 1980 (NTIS PB86-215829/AS)
6. ***Report on Statistical Uses of Administrative Records***, 1980 (NTIS PB86-214285/AS)
7. ***An Interagency Review of Time-Series Revision Policies***, 1982 (NTIS PB86-232451/AS)
8. ***Statistical Interagency Agreements***, 1982 (NTIS PB86-230570/AS)
9. ***Contracting for Surveys***, 1983 (NTIS PB83-233148)
10. ***Approaches to Developing Questionnaires***, 1983 (NTIS PB84-105055)
11. ***A Review of Industry Coding Systems***, 1984 (NTIS PB84-135276)
12. ***The Role of Telephone Data Collection in Federal Statistics***, 1984 (NTIS PB85-105971)
13. ***Federal Longitudinal Surveys***, 1986 (NTIS PB86-139730)
14. ***Workshop on Statistical Uses of Microcomputers in Federal Agencies***, 1987 (NTIS PB87-166393)
15. ***Quality in Establishment Surveys***, 1988 (NTIS PB88-232921)
16. ***A Comparative Study of Reporting Units in Selected Employer Data Systems***, 1990 (NTIS PB90-205238)
17. ***Survey Coverage***, 1990 (NTIS PB90-205246)
18. ***Data Editing in Federal Statistical Agencies***, 1990 (NTIS PB90-205253)
19. ***Computer Assisted Survey Information Collection***, 1990 (NTIS PB90-205261)
20. ***Seminar on Quality of Federal Data***, 1991 (NTIS PB91-142414)
21. ***Indirect Estimators in Federal Programs***, 1993 (NTIS PB93-209294)
22. ***Report on Statistical Disclosure Limitation Methodology***, 1994 (NTIS PB94-165305)
23. ***Seminar on New Directions in Statistical Methodology***, 1995 (NTIS PB95-182978)
24. ***Electronic Dissemination of Statistical Data***, 1995 (NTIS PB96-121629)
25. ***Data Editing Workshop and Exposition***, 1996 (NTIS PB97-104624)
26. ***Seminar on Statistical Methodology in the Public Service***, 1997 (NTIS PB97-162580)
27. ***Training for the Future: Addressing Tomorrow's Survey Tasks***, 1998 (NTIS PB99-102576)
28. ***Seminar on Interagency Coordination and Cooperation***, 1999 (NTIS PB99-132029)
29. ***Federal Committee on Statistical Methodology Research Conference (Conference Papers)***, 1999 (NTIS PB99-166795)
30. ***1999 Federal Committee on Statistical Methodology Research Conference: Complete Proceedings***, 2000 (NTIS PB2000-105886)
31. ***Measuring and Reporting Sources of Error in Surveys***, 2001 (NTIS PB2001-104329)
32. ***Seminar on Integrating Federal Statistical Information and Processes***, 2001 (NTIS PB2001-104626)
33. ***Seminar on the Funding Opportunity in Survey Research***, 2001 (NTIS PB2001-108851)
34. ***Federal Committee on Statistical Methodology Research Conference (Conference Papers)***, 2001 (NTIS PB)

Copies of these working papers may be ordered from NTIS Document Sales, 5285 Port Royal Road, Springfield, VA 22161; telephone: 1-800-553-6847. The Statistical Policy Working Paper series is also available electronically from FCSM's web site <*http://www.fcsm.gov*>.