

Session 2

WHAT THE PUBLIC NEEDS TO KNOW ABOUT FEDERAL STATISTICS

# Organizational Perspectives and the Agenda of Federal Statistical Agencies: How What We Know Reflects Who We Are

By Daniel Melnick<sup>1</sup>  
National Science Foundation

Here is the dilemma federal statisticians face every day in their work.

When we use numbers to enlighten public [and private] policy, particular perspectives are key to the meaningfulness of assertions. Agency mandates shape perspectives and in turn data collection, aggregation and analysis. The models of data analysis in common use assume these perspectives are or should be uniform. In fact, they are diverse. How can we create general purpose statistics if the goals we serve are particular to the questions put by our constituents? Yet, without this infrastructure, individual studies would be impossible.

## Some Examples

Consider three examples of apparently flawed procedures that might be criticized from a statistical perspective but have considerable validity in the context of the data collectors' expectations and their relationships with respondents and constituents.

- A Member of Congress sends out a questionnaire to all of the people living in the district. It is a simple flyer addressed to occupant and it contains questions which express the Member's views as much as ask for opinions. 15,000 constituents or about five percent of the adults in the district return the form. The Member touts the results and seems to pay a great deal of attention to them.
- A group of mental health researchers designs a survey to measure the prevalence of mental disorders in the general population. Because they are concerned about the need to accurately measure the respondent's condition, they recruit psychiatry students and staff from five leading tertiary care teaching hospitals with strong departments of psychiatry. In each catchment area near these institutions, they draw a probability

---

<sup>1</sup> Daniel Melnick, Ph.D. is a Senior Advisor, Research Methods at the National Science Foundation. The views expressed in this paper are his own and do not represent an opinion of the National Science Foundation.

sample of the population and administer a detailed diagnostic protocol. Subsequently, they use demographic data from the decennial census to extrapolate national estimates. When the next census is published, they update the extrapolation and report on expected changes in the prevalence of mental and addictive disorders. These figures become part of the calculations for a major national health care initiative in an attempt to estimate the number of people who will demand mental health services under the new plan.

- A group of substance abuse researchers wants to estimate the number of people needing drug abuse treatment, the number receiving it and the treatment deficit. They are particularly concerned with measuring the changes that followed increased funding for treatment. Drawing on an annual survey of a probability sample of US households, they create a composite indicator of treatment need which the psychiatrists in the second example would probably not trust. They calibrate these results against administrative data to try to adjust for reporting biases. They compare the treatment deficits of population groups. Yearly change can be reported to satisfy OMB and the White House Drug Policy office.

It is easy to show why each of these procedures is wrong in the abstract:

- The Member of Congress obviously does not have a representative sample of the district and the questionnaire probably did not provide a reading of the "true" opinions of the respondents—i.e. the views they held before they received the Members mailing.
- The psychiatrists don't really have a nationally representative sample either. They may also be missing variation between places with and without tertiary care institutions. When the data are updated with new census findings, the researchers assume that there is no change in the relationship between the demographic characteristics of study participants and mental health conditions.
- The substance abuse researchers make the grand assumption that the limited data they have from their survey really measures the need for treatment. They also assume that respondents have been frank about very sensitive personal matters. They justify this by the need to measure the amount of change from one year to the next and between groups.

While its easy to use an abstract standard to judge these studies, its wrong to do it. Each of these procedures must be understood in terms of the underlying assumptions and expectations of the data

collectors and their mandate. As long as they remain within the limits of their own framework, the results have meaning. When they depart from it, they experience difficulties.

- The Member of Congress uses the questionnaire to provide constituents with a way to express their views. While information about a representative sample would be useful, from the Member's perspective the mailing provides a way to measure the number of people who are deeply concerned about issues. The questionnaire is a crude kind of behavioral test—those who return it tend to be more exercised about the issues included in the questionnaire than those who do not. For example, knowing that there are at least 300 people who are deeply troubled about the siting of a government office helps the Member fulfill his duty to represent his constituents.
- The psychiatrists believe mental disorders result from long lasting personality or biological factors which are not likely to vary much from place to place. Because they assume that the key to understanding the prevalence of these conditions lies in expert diagnosis, they believe it's better to compromise on sampling and timeliness to insure accurate measurements. When they have to respond to demands for updated information, it seems reasonable to adjust decade old results with the new census data. A subsequent National survey using probability sampling techniques uses procedures that are calibrated to the original 5 site study.
- The substance abuse researchers focus on tracking change and geographic variation. Their constituents believe that drug abuse is the product of overt infusion of pernicious activity. Therefore, they demand indicators that can track its progress yearly [if not semiannually, quarterly or monthly] and report results for cities and states. The demands to justify expenditures in annually submitted budgets force them to estimate the effect of treatment initiatives. Criminal justice authorities also use this data to plan their strategies. They compare the reported prevalence to estimates of the supply of drugs which is equally if not more uncertain.

The psychiatrists and the substance abuse researchers both report on drug abuse, but their different perspectives lead them to different conclusions. Policy makers sometimes ask which of these is "correct", but in fact they both appropriately answer different questions and careful analysis shows they can be reconciled.

We may think that federal statistical agencies are shielded from these kinds of concerns. After all, many of them are organizationally distinct from mission agencies. But, each data collection and statistical reporting organization carries its own perspectives to the field, in the office and ultimately in its data tapes and reports. Each constituent evaluates data in terms of their particular inquiry--but statistical agencies cannot adopt a narrow view or they risk eroding their main role as providing the infrastructure for many particular analyses. This dilemma is at the heart of many disputes. For example, consider how BLS's wage statistics program must meet the needs of federal employee wage adjustments while still providing general wage statistics.

Organizations shape the content, quality and meaning of federal data series. Their expectations, relationships and mandates shape the view provided by seemingly unbiased factual presentations. To account for these factors, we need to go beyond the definition of uncertainty used by statisticians in which total error equals sampling error plus non-sampling error.

We can draw upon a large literature about organizations which has recently been enhanced by studies of the decision and cognitive sciences. These perspectives have been applied to the interactions between survey interviewers and respondents.<sup>2</sup> But, we need to assess their implications for the entire system of collecting and reporting statistics. These factors are not exclusively found in federal statistical agencies, they are common to every organization that collects and reports information. But, today we are focused on the federal effort.

### **Expectations Shape the Outcome**

What we expect shapes what we report, because it limits what we look for and thereby what we find. Different approaches to compiling information grow out of the different questions policy makers and the public ask. It helps to think about the stages at which information is used, and what it is used for. Whatever else they do, statistical agencies provide a factual foundation for the "authoritative allocation" of resources.<sup>3</sup> Relationships with users and constituents provide the context for this work and shape its impact on how public policy is grounded in facts.

Decision makers and the public use facts to:

- **monitor conditions**, by reporting on trends, tracking changes, or comparing groups,
- **plan action or propose responses**, in legislation, government rules or corporate policies,
- **implement action strategies**, [For example, companies use them to decide how to shape their marketing messages, candidates use them to decide where to put scarce campaign resources and legislators use them to fashion allocation formulas for block grants.]

---

<sup>2</sup> For example see Norbert Schwarz and Seymour Sudman, editors, *Answering questions : methodology for determining cognitive and communicative processes in survey research* / 1st ed. San Francisco : Jossey-Bass Publishers, c1996. xi, 469 p. : ill. : 24 cm. Includes bibliographical references (p. 403-441) and index.

<sup>3</sup> David Easton. *A systems analysis of political life*. New York, Wiley [1965] xvi, 507 p has an elaborate presentation of the implications of this phrase.

- **evaluate the results** of those strategies, by trying to judge what actually happened compared to what would have occurred if they had not acted, *and*
- **weigh evidence** during the adjudication of disputes arising from these actions. [For example, when experts report about the results of psychological tests or the number of children from different groups admitted to a school.]

There seems to be a temporal logic to these activities, but it is a constructed logic. In fact, frequently these activities occur out of order. Thus, their linear appearance is just that, a neat way to describe them, but far from what actually happens.

### Modes of Data Collection

The relationship between statistical agencies and their sources might best be summarized by borrowing from a classical categorization that divides data into those coming from administrative records, censuses, periodic surveys, focused or single project surveys, and experiments.

- **Administrative records** are the result of corporate activity either in private companies, community organizations [like universities or schools] or government agencies [like the IRS];<sup>4</sup> For example, when goods sold in a store, loans paid to the bank, votes cast, passports issued, taxes paid or unpaid, tickets issued for a performance, books lent in a library, grant applicants processed, and payments made to mothers with children, transaction records are a part of the administrative process.
- **Censuses**, although specially organized, are generally so comprehensive that they use many of the same procedures and have many of the same characteristics as Administrative records except that censuses count the stock of people, organizations or property while administrative records track the flow of actions on these elements;<sup>55</sup>
- **Periodic** sample surveys can be much more carefully controlled. They generally employ a permanent and professionally trained survey crew, led by skilled statisticians and analysts to collect the same or very similar information each day, week, month or year.
- **Focused surveys** are specially organized to collect information about a specific topic, population or issue. They use many of the same methods as periodic sample surveys but have greater flexibility to target information needed to answer specific questions. And
- **Experiments** purposefully vary the lives of people or organizations to test the effect that this has on attainment of stated goals.

The match between the questions policy makers and other users want to answer and the particular type of data available is a key factor in the validity of reported results.

---

<sup>4</sup>These have been the stock in trade of researchers for hundreds of years. What differs now is the increasing mining of them for quantitative estimates that are integrated with data from more controlled studies.

<sup>55</sup>Censuses also have a long history, going back to Biblical times. See Anderson, Margo J., 1945-The American census: a social history / Margo J. Anderson. New Haven: Yale University Press. c1988. xiii, 257 p. : ill. ; 25 cm



Disputes about statistical results were focused by the impact that statistical work has on policy choices, limiting leaders options and setting the context for constituent support. Nothing gets the public's attention more than learning that a new report will shift resources or power, when new census figures shift representation from one state to another, new crop figures effect prices, or the latest Consumer Price Index shifts billions of dollars from employers to employees, and the government to its beneficiaries.<sup>6</sup>

This effort is important because:

- Federal data are the foundation for what we know about our country, its people, economy and society. A large part of the information generated in the private sector and social science research conducted by Universities relies on federal data for benchmarks, sampling frames or controls. And
- Federal data must suit multiple purposes: providing both the descriptive and analytical information and mechanisms to directly allocate resources. When policies are drafted, statistical results provide key information to calibrate impacts. The data is used to determine if proposals are feasible. Just consider the way the Congressional Budget Office uses federal statistics to cost out legislative proposals which are then scored against goals to see if they comply with the budget resolution. And when the statistical basis for allocations is ambiguous, it is difficult to get the provisions enacted.

### Facts, Figures, and War

Statistical indicators are central to our public policy debates. 55 years ago, in October of 1941, President Roosevelt appointed the Librarian of Congress as the head of a Bureau of Facts and Figures.<sup>7</sup> This office which was ultimately absorbed into the Office of War Information, marshaled accurate information--the Facts and Figures-- to counter war rumors. The largest part of its staff was assigned to a statistical operation- the Bureau of Intelligence which monitored US public reactions. Its results enabled Archibald McLeash to have extraordinary influence in the early months of the war-- because he had charts and graphs to back up his assertions. The OFF also monitored and tried to gain control over the reports of US statistical agencies so as to provide

---

<sup>6</sup>Other disputes are summarized in *The Decennial census : an analysis and review* / prepared for the Subcommittee on Energy, Nuclear Proliferation and Federal Services of the Committee on Governmental Affairs, United States Senate : by the Congressional Research Service. Library of Congress. Washington : U.S. G.P.O. : For sale by the Supt. of Docs., U.S. G.P.O., 1980. ix, 465 p. : ill. : 24 cm. OTHER NAMES: Melnick, Daniel. United States. Congress. Senate. Committee on Governmental Affairs. Subcommittee on Energy, Nuclear Proliferation and Federal Services. Library of Congress. Congressional Research Service. NOTES: At head of title: 96th Congress, 2d session. Committee print. Authors: Daniel Melnick and others. "November 1980." And, Anderson, Margo J., 1945-The American census : a social history / Margo J. Anderson. New Haven : Yale University Press, c1988. xiii, 257 p. : ill. : 25 cm. Also, Numerous issues have been reviewed in reports issued by the National Research Council's Committee on National Statistics.

<sup>7</sup>This account is based on papers found in boxes 52 and 53 of the Archibald McLeash Papers at the Library of Congress. See, History of the Office of Facts and Figures, typescript HFG 9/23/1943.

needed information to the public while denying intelligence to the enemy. In support of this effort, Library staff went on a 24 hour schedule, so that the information would always be available on short notice. The speed with which this was done and its importance illustrates the role that statistical results play in policy analysis, strategic planning and implementation.<sup>8</sup> By the summer of 1942, these activities were subsumed in the work of the Office of War Information. Analysts applied the results to pressing decisions. The importance of statistical data and the attention paid to it marks a real shift in our view of government action.

Of course, this library mobilization followed several decades during which a revolution in federal statistics was begun.<sup>9</sup> The period after the First World War saw the institutionalization of federal statistical efforts and the introduction [into the 1930's and 1940's] of modern sampling and statistical controls. At the same time, the changed role of the American government and the altered economic and social system brought on by the depression and World War II intensified demands for statistical information. For example, the rise of the radio and television created a whole new kind of demand for numbers to show the impact [and monetary worth] of totally new information products. At its base, each new private statistical service [like the Nielsen ratings] rested on federal statistics which provided its foundation.

### **Rhetoric and Facts**

Most of us recognize that there are different versions of the facts depending on our points of view. Sometimes, we easily see that an apparently factual presentation is rhetoric subtly or not so subtly expressing the policy preferences of analysts or the decision makers who use their work. Yet, these purposeful attempts to present facts so as to make an argument are not the most difficult or troublesome instances of distortion. As long as the purposes are clearly set forth, these arguments are a legitimate part of the policy dialog. For given the constructed nature of facts, there is no escaping the impact of opinion on fact. [But of course we can try to present different views of the same circumstances in an attempt to balance the argument.-- the sort of pro/con analysis developed into a high art at the Congressional Research Service.]

In fact, a conscious attempt to mold the public's view of the world by shaping the statistics we report is easier to deal with than the more subtle impact of organizational expectations. In our open society, the biases are often apparent or can be easily identified. Opposing analysts are quick to articulate them. Even without a planned attempt to shape statistical reports there are numerous disputes about the fairness and accuracy of reported results. If there were none, we would wonder if the reports were important. In fact, the very controversies generated by statistical reports show their importance for public and private decisions.

---

<sup>8</sup>Annual reports of the Librarian of Congress detail this work. It is described in Jean M. Converse, *Survey Research in the United States. Roots and Emergence 1890-1960*. University of California Press, Berkeley, 1987 pp. 171-172.

<sup>9</sup>Duncan, Joseph W. *Revolution in United States Government statistics, 1926-1976* / written by Joseph W. Duncan and William C. Shelton. [Washington] : U.S. Dept. of Commerce, Office of Federal Statistical Policy and Standards : for sale by the Supt. of Docs., U.S. Govt. Print. Off., 1978. ix, 257 p. Shelton, William Chastain, 1916- joint author. See also Converse



How we view statistical inference is one of the most important expectations shaping our assessment of these facts.

### **Organizational Structure and Inference**

It is easy to think about inference as if the projects generating statistics were investigations conducted by a single researcher or a group focused on a defined set of issues. They formulate an hypothesis and design a procedure to test it including the data collection procedure [sample, questionnaire, interview, etc.] and a specific inferential plan. The statistical tests we use work best when part of this kind of process.

Statistical agencies, on the other hand, collect data for general purposes. Even statistical offices in mission oriented agencies frequently have mandates to report on conditions rather than test hypotheses. Analysts using this data impose post hoc hypotheses and try to test them. The agencies operate with a distinct mandate that shapes their expectations and defines the conceptual maps that guide their work. These cultures give purpose to the efforts of the agency by showing how its work fits into a broader pattern of action.

Before the growth of the counting industry, it was sufficient to address the problem of analysis by reflecting on the dual nature of facts: that they are defined by our ideas and refract our environment. What we know arises from our own expectations about what we might know. At the same time, what we know reflects learned definitions of what there is to know. Thus, facts are a product of individual innovation and learned behaviors. Language sets up the definitions and names the feelings that allow individuals to see and not to see aspects of the world. Without such filtering, the mind cannot comprehend the unlimited impressions thrown at it. Facts are therefore created and arise from the fictions inherent in the learned environment.

As anthropologists have recently shown<sup>10</sup>, intelligence--the skill of manipulating information and relating it to activity--may have grown from the increasing dependence and communication between primates and ultimately men and women. Thus, the apparent innate skill of knowing arises from cooperation and is learned. It lies exposed as yet another cultural artifact-- perhaps the master artifact of society.

As we try to use modern procedures to collect, compile, analyze and report facts, the organizations we work in [such as the Census Bureau, the Income tax department, the social entitlement office, the polling firm, the market research department, the academic survey research center, or the statistical office of a large company] become dominant instruments of fact finding. It is as if our fictions [the meaning we gave to sense impressions that make them facts] have taken on a corporeal existence and are manifest in these Bureaus and Institutes. We should recognize that this continues a tradition of institutions which came before our enchantment with counting. Yet, the increasing quantitative expression of facts has radically altered the way we think.

---

<sup>10</sup>For example see Esther N. Goody ed. Social Intelligence and Interaction. Expressions and Implications of the Social Bias in Human Intelligence. Cambridge University Press. 1995.

To be sure, the distinction is a matter of degree. The family, guild, state, church, library and legislature surely also embody such domains of meaning. Yet, with the rise of the professional fact finders, we see a newly ordained fact-space in which dogma and values are not supposed to dominate. These new bureaucracies certify the veracity of stated realities while they also specify the limits to our confidence in them. Meanwhile, they legitimate what we know by giving it a quantitative face that adds an appearance of definiteness even as it allows analysts to indicate the probable degree of associated uncertainty. The fact finders treat events as mundane and emotionless, but yet, their product provides bold new banners--sentinels of modern life. And as such, they become the vortex for heated debates.<sup>11</sup>

Consider the question of accuracy: how do we know that our reports accurately and appropriately reflect the truth? For many of the most interesting questions, no gold standard of veracity can be found. We replace it with more relative yardsticks testing the adequacy of what we do know. Thus, truth is estimated by comparing the answers with the questions--and consistent responses often stand in for valid ones. Validation is predicated on the expectation that analysts pose problems, they observe phenomena focused on these issues, and the results approximate an answer. Statisticians step in to help us measure the limits of the outcome.

Yet, as we periodically collect large scale surveys and censuses or compile data from the records of bureaucratic files, frequently the hypotheses are only implicit-- and often the only way to collect data is to allow different participants to adhere to their own view of the purposes.

All of what we think we know about society, politics and its environment -- the most basic facts -- are refracted through these structures. This means that, we cannot know about the uncertainty of our results -- the variances from some absolute truth-- without understanding the effect of these organizational mechanisms on what we think we know.

To do this, we must look beyond the conventional statistical tests of variance and the impact of measurement as well, to the systematic effects of these social structures. The lead indicators that are the artifacts of our age drift off into this uncharted space. Analysts struggle to extract meaning from this constructed environment where bits of data are transformation into information which becomes fact.

Each bit of data ultimately relates to our perception of the subjective relationship. And each in turn is both observation and activity. By focusing attention on a particular aspect of social relations, we take a stand on the issues. Yet, we agree to suspend our acknowledgment of this ultimate subjectivity--attempting to cast what we want to say in objective fact.

---

<sup>11</sup>Herbert Simon, *Administrative Behavior*, The Free Press, New York, 1945. See especially Chapter III Fact and Value in Decision-making.

**Implications**

Although it may seem convenient to treat the organizational perspectives discussed above as biases to be identified or controlled, they are as much a part of statistical work as questionnaires, forms, samples or formulaic tests. They provide the purpose and direction for much of the work. Inferences are best judged in light of their contribution to these purposes. There is no escaping that who we are greatly influences what we know.

# IMMIGRATION STATISTICS: NO LONGER NEGLECTED, BUT STILL INADEQUATE\*

by  
Jeffrey S. Passel†  
Urban Institute  
2100 M St., NW  
Washington, D.C. 20037

## Introduction

Immigration remains a very contentious area of public policy, even after a total overhaul of U.S. immigration policy between 1980 and 1990. The 104<sup>th</sup> Congress passed three significant pieces of legislation relating to immigration and narrowly failed to pass a fourth. The Antiterrorism and Effective Death Penalty Act of 1996 contained a number of provisions relating to immigration, focussing mainly on easing removal of aliens from the United States. The law limited appeal and judicial review for persons seeking asylum and expanded criteria for deportation to include relatively minor crimes. Welfare reform legislation — the Personal Responsibility and Work Opportunity Reconciliation Act of 1996 — included major modifications to immigrant and immigration policy. Two notable legal changes were severe limits on legal non-citizens' access to a range of public assistance programs and making the sponsor's affidavit of support legally enforceable by the government and the sponsored immigrant.

The final law passed, the Illegal Immigration Reform and Immigrant Responsibility Act of 1996, represents a significant alteration of immigration policy. The Act increased border enforcement by doubling the size of the Border Patrol and strengthening the physical barriers at the Mexican border. It also reinforced the limits to review, appeal, and access to public benefits in the previous two laws. The law also criminalizes many activities related to illegal immigration

---

\* Presented at the Seminar on Statistical Methodology in the Public Service: A Conference Dedicated to the Memory of Maria Gonzalez, sponsored by the Council of Professional Associations on Federal Statistics, Bethesda, Maryland, November 12–13, 1996.

† Jeffrey S. Passel is Director of the Urban Institute's Program for Research on Immigration Policy, a program of public policy research supported principally by The Ford Foundation. The views expressed are attributable to the author and do not necessarily reflect those of other staff members, officers, trustees of the Urban Institute, The Ford Foundation, or any organizations that provide financial support to the Program.

(e.g., smuggling and document fraud) and sets up pilot programs to test employment verification systems.

In an environment such as this, with numerous policy changes and legislative considerations, one would assume that a great deal is known about immigration, immigrants, impacts, and behaviors. Unfortunately, many basic facts about immigration and immigrants remain unknown or, at best, based on estimates of unknown or questionable quality, often because of serious data deficiencies. Detailed studies of immigration impacts and immigrant characteristics face similar limitations. The result is often “dueling studies” in which even reputable researchers reach diametrically opposite conclusions. Partly as a consequence of data deficiencies, development of new immigration and immigrant policies has only occasionally been the result of measured analysis of immigration phenomena and considered deliberation of empirical research.

At the heart of these difficulties over alternative interpretations lies the inadequacy of data on immigration and immigrants, principally from federal sources. Two studies by the National Academy of Sciences (Levine et al. 1985; Edmonston 1996) have described various barriers to producing needed data and prescribed potential solutions. Both of these studies led to some improvements in immigration statistics. However, notwithstanding the significant progress of the last decade, immigration data are still woefully inadequate for addressing a number of key areas of both immigration and immigrant policy.

Organizational and conceptual obstacles continue to plague analysts and policymakers attempting to deal with immigration issues. No agency or office is in charge, overseeing the planning, collection, and production of data on immigration and immigrants. Statistics and statistical analysis at INS still do not have a central role in the agency's mission; they remain a stepchild of administrative processes. The same is true of many other agencies as well, including the Bureau of the Census. Nonetheless, a number of new data sources are now available, with yet more coming on line soon. However, we still cannot answer some very basic questions, such as: “How many people come to the United States each year?”, “How many illegal aliens live in the country?” or, even more basic, “How many legal aliens and naturalized citizens live in the country (and each state)?” Finally, there are so many areas where experts' opinions are at odds in large

part because we have not developed a common framework for analysis and discussion, nor the appropriate data to address the pressing questions.

This paper covers four main areas of concern. First, what data do we need on immigration and immigrants? What would we like to know? Second, what data do we have? What is being planned and is likely to be available? What needs remain? Throughout these discussions the focus is on some basic measurement issues, largely dealing with the demography of immigration. Notably, how well are the concepts measured? What deficiencies and gaps are there? Finally, the paper closes with some proposals for reform and change.

### **Data Needs and Analytic Issues**

In assessing our data needs, we must examine linkages among data, data users, policy formulation, and policy implementation. Specifically, what data and analyses are needed to *formulate* policy. We also need to examine how the policies are *implemented* and to *evaluate* both the implementation and the policy outcomes. For all of these purposes, we need to keep in mind that we must not only investigate *immigration policy* (i.e., the numbers and types of persons admitted to or excluded from the country), but also and *immigrant policy* (i.e., those policies and programs for dealing with immigrants after they arrive in the country — integration issues — and the impacts of the immigrants).

Impacts of immigration across a broad spectrum of realms are either poorly understood or remain subjects of contention. The bases for admission of immigrants into the country are principally: family unification, employment needs, and humanitarian concerns. The labor market impacts of not only the employment-based immigrants, but also other legal immigrants and illegal immigrants represent the major arena of competition between immigrants and natives; the purported labor market impacts are frequently cited in policy debates. The research in this area has been generally inconclusive. Case studies and anecdotal evidence indicate the possibility of significant displacement of low-skilled natives. However, macro studies do not find support for such displacement effects, nor do they find substantial wage effects (Fix and Passel 1994). Some researchers (Borjas 1996) have argued that such effects must exist, but that the data and



measurement techniques available are simply not sufficient to detect them. Clearly more and better data are needed.

While the issue of economic competition between immigrants and natives may eventually be settled, there are more subtle impacts of immigration that continue to be even more difficult to discern. The availability of immigrant labor may affect the mechanization and transformation of entire sectors, such as agriculture. Interactions between trade, capital flows, and the migration of labor also require better data and methods to be measured accurately.

The impact of immigrants on the public fisc is another area where more and better data are sorely needed. While some researchers have found an overall fiscal surplus from immigrants (Passel and Clark 1994), others found significantly more costs than revenues (Huddle 1993). There does, however, seem to be general agreement that negative fiscal impacts are felt at state and local areas, driven principally by education costs, while revenue streams flow to the federal level. Not only are data generally inadequate for careful and complete assessments of immigrant (and native) impacts on the balance of public costs and revenues, but the proper analytic framework has yet to be developed and agreed upon. A National Academy of Sciences panel is currently addressing these issues.

One of the difficulties in fully assessing the fiscal impacts of immigrants is the inadequacy of methods for measuring indirect economic benefits derived from immigration. Again, anecdotal evidence abounds across the country for the positive impacts of immigrants in revitalizing once deteriorating neighborhoods. Similarly, immigrant businesses are known to be the principal driving factor behind economic activity in some neighborhoods and to account for significant amounts of tax revenues. However, none of the available fiscal impact studies takes these factors into account.

While analytic and policy issues abound in the immigration field, even more fundamental measurement issues need to be addressed. We simply do not have accurate measures of either the stock or flow of immigrant populations. Such basic questions need better answers and, in many cases, the answers to the larger policy questions raised above must be predicated on having basic demographic information.

## Immigrant Numbers

The passage of welfare reform in August 1996 has brought into focus some of the basic deficiencies of the nation's immigration data systems. This law limits access to a range of benefits for many categories of non-citizens, some of whom had formerly been eligible and receiving benefits. For example, legal permanent residents who have not worked 40 quarters in the United States and refugees who have been in the U.S. for more than 5 years are no longer eligible; naturalized citizens remain eligible. In addition, whereas eligibility rules governing aliens (legal and undocumented), had been the exclusive province of the federal government, the welfare reform law devolved responsibility for setting many of these eligibility rules, such as those covering undocumented aliens, to the states. Since the financial responsibility for providing services to the groups no longer eligible for federally-provided coverage was also devolved to states and localities, a number of new parties became interested in the numbers of naturalized citizens and aliens of various types, and in rates of naturalization; moreover, such data are needed for states and even smaller governmental units. Many of the population numbers needed are only available as rough estimates or simply do not exist.

There are five main immigrant populations of interest to most observers. In roughly decreasing order of size, they are:

- Legal permanent residents;
- Naturalized citizens;
- Undocumented aliens;
- Refugees, asylees, and parolees; and
- Nonimmigrant residents.

In addition, for some purposes, many want data for subgroups of these major groups — family-sponsored immigrants, employment immigrants, foreign students (a category of non-immigrants) to name just a few. Data are needed on the numbers in each category, the inflows and outflows each year (or at least the annual net change), and characteristics of the particular individuals. Interestingly enough, the only one of the five groups for which there are “official” estimates of size and annual net change is the one most would characterize as the hardest to measure — undocumented aliens.

There are three major impediments to obtaining data (or estimates) on the sizes of the various immigrant populations. First, the data we actually do get from INS relate to the numbers of people attaining the legal status each year, not the numbers actually entering the country. Thus, many persons becoming legal permanent resident aliens each year are currently living in the United States, for many years in some cases. The difference between new entrants and the total number can be quite substantial in some years, with new entrants generally running about 60 percent the total. However, in some years the difference can be so large as to be totally misleading. For example, in fiscal year 1991, official INS figures indicated about 1.8 million persons attained legal permanent resident status. This figure was touted, in the press and by anti-immigration advocates, as being the largest annual immigration in the nation's history. About 1.2 million of these "new" immigrants, however, were formerly undocumented aliens, most of whom had been living in the country for more than 10 years. The number of non-legalization legal permanent residents that year was approximately 600,000 or the same number as the year before; the number of true new entrants was even smaller.

A second major deficiency is the lack of any hard data on departures from the United States. The most recent estimates available cover emigration of the foreign-born population during the 1980s (Ahmed and Robinson 1994). These data suffer from a number of known problems. There is little detail available on the emigrants — only age, sex, and country of birth. Information of the emigrants' legal status and socioeconomic characteristics is completely absent. The third significant impediment to obtaining good demographic data is that there is virtually no information on movement of immigrants between immigration categories or within the United States.

## **Data Sources**

There have been major improvements in the amount of data on immigration and immigrants available in recent years. There are more sources, more topics covered, and more types of data. New data sources will be coming on-line soon. Nonetheless, there remain significant shortfalls and deficiencies. The major data gaps will continue to be a need for better, more up-to-date information on legal status and more longitudinal data. Government data sources remain the most extensive, but nongovernment surveys often provide richer data.

**Immigration and Naturalization Service.** The Immigration and Naturalization Service naturally remains the principal data source on in-flows of various types of immigrants. However, the INS is not organized as a statistical agency, even in its data-gathering divisions. One consequence of this organizational perspective is that administratively-based categories dominate data collection efforts. Counts of immigrants are based solely on official categories. Thus, it is very difficult to derive a figure for the demographically-based concept of *the number of people moving into the country each year* (Levine et al. 1985). The focus on administrative needs means that only limited information is available on the demographic and socioeconomic characteristics of the immigrants. Strangely, however, there is little linkage across data sets even though many administrative needs could be served by linked data sets. For example, little information is available on sponsors of immigrants.

The INS' focus on administrative data means that little analytic work related to immigration or integration policy is done in the agency. Even the analyses done in the agency tend to be idiosyncratic rather than institutionalized. The INS estimates of undocumented immigration serve as an example of both analytic work done at the agency and some of the institutional issues raised by these efforts. Generally, the analytic estimates of undocumented aliens released by the agency are of high quality (Warren 1994; INS 1997), use advanced methodological techniques, and respond to criticism and comments (e.g., Woodrow-Lafield no date; GAO 1995). However, a detailed description of the methodology employed, such as generally supplied by other statistical agencies, is lacking. The review and release procedures tend to be somewhat *ad hoc*. INS did convene an expert panel to review the latest estimates, but the release provides virtually no information on the statistical properties of the estimates. Furthermore, a lack of institutional commitment is quite apparent. The estimates are the product of a single researcher working largely without support staff and the principal data system required to measure visa overstayers has been allowed to fall into a state of disrepair and neglect.<sup>1</sup> In fact,

---

<sup>1</sup> Warren's estimates make significant use of matched arrival and departure forms from the Non-Immigrant Information System. These data may no longer be available on a regular basis. A similar event occurred in the early 1980s. Warren and Passel (1987) made extensive use of the Alien Registration system to develop estimates of legal residents and of undocumented aliens included in the 1980 Census. The collection of Alien Registration data was halted in 1982, shortly before Warren and Passel first released their estimates in 1983.

Warren's estimates became the "official" INS estimates largely because the new Commissioner (Meissner) knew of his work and started citing them, rather than through a concerted agency effort to measure this politically relevant, clandestine migration.

There are some other examples of analytic, policy-related research done at the INS, but the list is not long. Analyses of naturalizations, including cohort-based measures, are presented in recent *Statistical Yearbooks* (INS various dates). However, even this work does not go much beyond simply reporting rates of naturalization; there is little analysis of determinants of naturalization or attempts to correct the rates for emigration and mortality. Another recent example is an INS analysis of records of immigrant sponsors.

**Current Population Survey.** Within the last few years, there has been a substantial expansion of data on the foreign-born population available from national surveys. There is now more data available on socioeconomic characteristics of the foreign-born population for intercensal dates than ever before. Efforts of INS, NICHD, and a few other agencies have led to the creation of some new data sets. Substantial gaps remain, but the first steps have been made.

Beginning in 1994, the Current Population Survey (CPS), the monthly labor force survey conducted by the Bureau of the Census for the Bureau of Labor Statistics, has included questions on country of birth, year of immigration, citizenship, and country of birth of parents. These data are available monthly, but the annual March demographic supplement to the CPS provides a wealth of information on social and economic characteristics, income sources, and program usage. Although the sample is fairly large (approximately 50,000 households), its size does limit to some extent the amount of detail available on immigrants. The CPS is, however, the first significant source of data on second-generation Americans (i.e., the native-born children of immigrants) since the 1970 Census.

The new CPS data are extremely important and their continued collection needs to be guaranteed and institutionalized. The data, however, are not without problems. Some of these can be attributed directly to the lack of institutional support for immigration statistics. The data on the foreign-born population from the 1994 CPS were available only for a very limited list of about 22 countries of birth. Immigrants from other countries were grouped together into an "all

other" category. Thus, it was not possible to tabulate information on immigrants from Asia, Europe, or any aggregation other than the specific list of countries. Although this problem was fixed by 1995, it would not have occurred if the Census Bureau had a staff or branch devoted to immigration statistics. In 1995, another glitch occurred. The list of countries on the data collection instrument was expanded and all responses on country of birth were coded directly, permitting aggregations at any geographic level desired. However, data for 6 countries were lost: all new responses of South Korea, Taiwan, Thailand, Trinidad and Tobago, Vietnam, and Yugoslavia were treated as "unknown country of birth" and imputed to other specific countries. The result was serious shortfalls in the estimated numbers from these countries and substantial overstatements for other Asian countries including India and the Philippines. The Census Bureau has recently released corrected data some 15 months after the initial release, but again the lack of staff dedicated to immigration statistics undoubtedly delayed finding the error and resulted in release of erroneous information.

A yet more serious problem affects the CPS data for 1994 and 1995. A change in editing procedures for the race question led to inconsistencies between the CPS data for persons who were not white or black and the demographic estimates used as control totals for the survey. (See Passel 1996 for more details.) As a result, CPS estimates of Asian/Pacific Islanders and American Indians/Alaska Natives are understated by approximately 20 percent in 1994 and 30 percent in 1995. In March 1995, the shortfall amounts to more than 1.5 million Asians. Since about two-thirds of this group is foreign-born, the resulting data on the foreign-born population, its origins, and its geographic distribution within the United States are seriously distorted. By 1996, this weighting problem and the coding problems described above have all been fixed.

Two other issues illustrate some of the problems deriving from the lack of institutional support for immigration statistics. The Census Bureau has consistently downplayed the significance of the weighting problem and has no plans to issue reweighted data.<sup>2</sup> Thus, many researchers remain unaware of a deficiency that could have a substantial impact on research and

---

<sup>2</sup> The Urban Institute is planning to release new weights for the March 1994 and 1995 socioeconomic supplements which should permit immigration researchers to better approximate the size and characteristics of the foreign-born population. See Passel 1997.



have no official mechanism for correcting the data even if they are aware of it error. The citizenship data also fail to meet fully the needs of researchers. These data, like other information on immigration, are collected only at the initial interview of CPS respondents. Because of the rotation patterns for the CPS sample, some of the information released on naturalization may be 16 months old. This lag makes very little difference for invariant statistics such as country of birth, country of birth of parents, and year of immigration. However, when more than 1 million immigrants have naturalized in each of the last two fiscal years, data on naturalized citizens could be significantly affected. More fundamental, however, is the lack of information on immigration status (e.g., refugee, legal permanent resident, undocumented, student visa).

**Other Government Surveys.** Data on immigrants from other government surveys is more limited, but new information will be available in the foreseeable future. The Survey on Income and Program Participation (SIPP) is a semi-longitudinal survey, tracking respondents for two-and-one-half years, which provides information on immigrants. The sample size is smaller than the CPS so information on immigrants is more limited. However, the welfare reform legislation provides funds for a substantially expanded SIPP (called the Survey of Program Dynamics) that will follow respondents for longer periods and have a larger sample.<sup>3</sup>

The American Community Survey (ACS) has been proposed as an alternative to detailed collection of sample data in the decennial census. As currently envisioned, the ACS would consist of large, independent monthly samples amounting to perhaps 3 million households every year. As an alternative to decennial census data, the ACS would fill most needs of immigration researchers who currently use census data. Moreover, the ACS data would be more timely since they will be collected on an on-going basis rather than once every 10 years. The ACS also represents an opportunity to expand data related to immigration. More information on the second generation and more detailed information on citizenship and legal status, if collected, would make the ACS an exceedingly useful source of immigration data. At this time, plans and funding for the ACS remain uncertain at this time.

---

<sup>3</sup> The PSID, a longitudinal survey that has been tracking respondents for almost 30 years, does not have many immigrants since the sample selection predated the current wave of mass immigration and new immigrants have not been added in representative numbers.

Another new survey, if fully operationalized, will provide sorely needed longitudinal information on immigrants. A pilot test for the so-called "Green Card" survey is currently in the field, funded by NICHD. This survey will be a sample of aliens admitted for permanent residence (i.e., legal immigrants) during a year, augmented by a sample of certain nonimmigrants. The immigrants will be followed for a number of years at specific intervals. This survey will have good data on legal status and will provide information on adaptation of immigrants the U.S. society. The survey will be limited in that it is only a single cohort and limited only to legal immigrants. Nonetheless, it will be an unprecedented source of data when fully funded and operationalized for a number of years.

**Administrative Data.** There are a number of potentially useful administrative data sets available for the study of immigration, including data from Social Security, Supplemental Security Income, and Food Stamps. These data have been underutilized for immigration research, in part, because their scope is limited. (For example, a study of Food Stamp recipients needs a comparison group of non-recipients that requires another source of data.) Nativity and immigration status are often not routinely or systematically collected. New data from welfare reform may be a gold mine for immigration researchers, as certain data collection activities are mandated in the law. To be fully utilized, researchers will need to ensure that needed data, especially on nativity and citizenship are collected, and that there are some parallel data on no-welfare users.

## **Data Deficiencies**

The preceding discussion has covered a number of issues and problems affecting immigration data. While there are a number of new sources available now and in the near future, the major deficiencies can be summarized by four key points:

**INS Administrative Data.** The main purpose of data collected by INS is administrative; this limits severely its utility for analytic work. Only data for administrative actions are routinely collected, even when other information could be collected from immigrants at minimal cost, in terms of either money or respondent burden. The concepts and definitions used can be quite confusing and are often irrelevant for analysis or policy implementation. In addition, significant

changes in collection methods, coding, and data processing have occurred, making comparisons over time difficult.

**Legal Status Information.** The legal status of immigrants is a key characteristic for many new programs and policies as well as for many existing ones, with the most important statuses being legal permanent resident, naturalized citizen, refugee (on admission), undocumented alien, and nonimmigrant. These characteristics are not routinely collected in censuses and surveys, nor even in some administrative data sets. A number of technical issues must be addressed before such data can be systematically collected. The impact of legal status questions on response rates is unknown, but one would expect response rates to be reduced overall by such questions; undocumented aliens, in particular, may be dissuaded from participating in government surveys. In addition, many respondents may not be able to provide accurate information on their own legal status, nor for other household members. Children who derive U.S. citizenship when their parents naturalize may be an especially problematic group.<sup>4</sup> INS record systems may, in fact, need to be upgraded to provide accurate and timely information on the legal status of individuals.

**Demographic Flows.** As noted, we do not have even basic information on the numbers of immigrants living in the United States in various legal statuses; the only official estimates are for undocumented immigrants. Estimates of resident legal permanent residents and naturalized citizens are critical for assessing welfare reform and its impacts. Yet, available estimates are little more than "back of the envelope" approximations. Nor do we have information on the flows of immigrants into and out of the country each year. The in-flows can be closely approximated by manipulation of various INS data bases. However, estimates of out-flows are almost completely the product of informed guesses.

Even more serious for the future of research and impact assessment is that no one in either INS or the Census Bureau is tasked with providing such estimates. To the extent that estimates are made, they are often by-products of other work.

---

<sup>4</sup> Even INS does not have complete information on this group.

**Longitudinal Data.** Many of the significant issues concerning immigrants, their impacts, and integration require data on changes over time in the number of immigrants and their characteristics. There is virtually no such data covering significant numbers of immigrants. Synthetic cohorts, put together from successive data collections (i.e., decennial censuses, monthly CPSs) can fill some of the gaps, but selective outmigration hinders most serious analyses. Eventually, if systems currently in testing come to full fruition, we will have the required longitudinal data. Unfortunately, it takes time to collect such data.

## **Conclusions and Recommendations**

Policies get made and implemented even if no data, or only deficient data, are available. Furthermore, some policies are instituted even in the face of data and analyses. For example, a number of studies showed that discrimination against foreign-looking and sounding persons legally in the country resulted from IRCA's employer sanctions. The law called for sunseting of these provisions in such circumstances, but the sanctions remain in place today, largely because of the politics surrounding the issue. Nonetheless, it is generally thought that data and research aid the policy formulation process. Immigration and immigrant policies should be no different. More plentiful, more targeted, and more accurate research is clearly a desideratum. However, better and more plentiful data are a necessity. What can be done to improve upon the current situation? The following four main points, if implemented, will clearly lead to a better understanding of immigration and its impacts on the United States.

***Centralize Federal Control of Immigration Statistics.*** No agency or office has official responsibility for immigration data collection (and analysis) activities. Consequently, there are many important activities which are either not being done or are being done in a very *ad hoc* manner. In some cases, there is no institutional support for essential data systems.

Some active oversight is needed for immigration statistics. There must be someone or some group defining data needs, setting data standards, and allocating resources so that essential tasks do not get overlooked. The lack of such oversight and planning is very apparent — in a recent book assessing the entire Federal

statistical system (Norwood 1996), immigration statistics are not even mentioned. Such an oversight or policy office could be located in the Office of Management and Budget or at INS, but in any case it needs cross-agency powers for allocating resources and defining data needs.

***Strengthen INS Data Production & Analysis Functions.*** INS data production activities are clearly not an agency priority. There are some resource allocation problems, although the agency is flush with dollars and actually faces more problems of finding sufficient staff than budget problems. The administrative and policing mindset of the agency is more of an impediment to collection of useful, high quality data. These priorities need to be changed.

Analytic capacity in the area of immigration and immigrant policy needs to be strengthened within the federal government generally. INS' policy office has begun to expand in this area. It should be given higher priority and expanded. However, the ideal locus is not immediately apparent — INS may not be the best location. This issue is the type that could best be addressed by a centralized immigration statistics office. Some important projects could also be done outside the government. Wherever the research is done (either inside or outside the INS or federal government), analysts need to have access to essential data systems.

***Improve Data Collection Systems.*** Progress has been made in this area, but a great deal more could be and needs to be done. One deficiency affecting virtually all current data systems is the lack of data on legal status. With the new welfare regimes, particular attention needs to be given to acquiring quality data on legal status, especially naturalized citizenship. Integration of data from disparate sources could be a particularly valuable method of filling gaps in the data.

***Develop a Consensus on an Analytic Framework.*** A common language and reference system facilitates communication and policy formulation. In the area of immigration, such a common analytic framework may be extremely difficult, if not impossible, to achieve, given the variety of views on the subject. Nonetheless, a general, if not

complete, consensus will still be valuable. Several National Academy of Sciences' panels are beginning a significant effort in this area, as they assess existing studies of fiscal impacts and conduct their own. These efforts need to be encouraged and strengthened. "Dueling" experts are, in part, a consequence of this shortfall, as many studies simply "talk past" one another. However, even a consensus framework will not be helpful if we only have bad data.



## REFERENCES

- Ahmed, Bashir and J. Gregory Robinson. 1994. "Estimates of Emigration of the Foreign-born Population: 1980-1990." Population Division Technical Working Paper No. 9. Washington, DC: U.S. Bureau of the Census. December.
- Edmonston, Barry (editor). 1996. *Statistics on U.S. Immigration: An Assessment of Data Needs for Future Research*. Washington, DC: National Academy Press.
- Fix, Michael and Jeffrey S. Passel. 1994. *Immigration and Immigrants: Setting the Record Straight*. Washington, DC: Urban Institute.
- General Accounting Office. 1995. *Illegal Immigration: INS Overstay Estimation Methods Need Improvement*. Letter Report, GAO/PEMD-95-20. September 26.
- Huddle, Donald. 1993. "The Costs of Immigration." Washington, DC: Carrying Capacity Network Research Series, Updated July 20.
- Immigration and Naturalization Services. 1997. Press release on estimates of undocumented immigration. February 5.
- Immigration and Naturalization Services. Various dates. *Statistical Yearbook of the INS*. Washington, DC: U.S. Government Printing Office.
- Levine, Daniel B., Kenneth Hill, and Robert Warren (editors). 1985. *Immigration Statistics: A Story of Neglect*. Washington, DC: National Academy Press.
- Norwood, Janet L. 1995. *Organizing to Count: Change in the Federal Statistical System*. Washington, DC: Urban Institute Press.
- Passel, Jeffrey S. 1997. "New Weights for the March 1994 and March 1995 Current Population Surveys." Washington, DC: Urban Institute. Memorandum, forthcoming.
- Passel, Jeffrey S. 1996. "Problem with March 1994 and 1995 CPS Weighting." Washington, DC: Urban Institute. Memorandum, November 12.
- Passel, Jeffrey S. and Rebecca L. Clark. 1994. "How Much Do Immigrants Really Cost? A Reappraisal of Huddle's 'The Costs of Immigrants.'" Washington, DC: Program for Research on Immigration Policy, Urban Institute. Photocopy, February.
- Warren, Robert. 1994. "Estimates of the Unauthorized Immigrant Population Residing in the United States, by Country of Origin and State of Residence: October 1992." Washington, DC: Statistics Division, Immigration and Naturalization Service.
- Warren, Robert and Jeffrey S. Passel. 1987. "A Count of the Uncountable: Estimates of Undocumented Aliens Counted in the 1980 Census." *Demography* 24 (3, August): 375-393.
- Woodrow-Lafield, Karen A. No date. "A Sociology of Official Statistics on Undocumented Immigrants." Photocopy.

**Seminar on Statistical Methodology in the Public Service**

**November 12-13, 1996**

**Session 2: What the Public Needs to Know About Federal Statistics**

**Discussion**

**Constance F. Citro**

**Committee on National Statistics**

[Note: The views expressed are those of the author and not of CNSTAT.]

When I was asked to be a discussant at this session, my first question was: What do papers on immigration statistics and organizational perspectives have to do with the topic of the session? My second question was: What is the topic of the session anyway? After all, the title of the session, "What the Public Needs to Know About Federal Statistics," can be interpreted in several ways simply by changing a word or two:

(1) One interpretation is: "What the Public Needs to Know FROM Federal Statistics."

That is, what data does the public need to have from the statistical system to inform policy debates and serve other important purposes? I interpret "the public" in broad terms as the data-using community of policy makers, analysts, planners, researchers, the media, and the general public; and I stress the need for the statistical system to be relevant to policy concerns. Jeff Passel's paper on immigration data relates directly to this interpretation of the topic; Dan Melnick's paper on organizational perspectives also bears on it, although indirectly.

(2) A second interpretation is: "What the Public Needs to Know TO FIND OUT About Federal Statistics." This has to do with one-stop shopping, which is the topic of another session, but it is relevant to the first interpretation of this session's topic because data of which users are ignorant cannot be used to inform policy or serve other public purposes and most likely represent a waste of resources.

(3) A third interpretation is: "What the Public Needs to Know About THE QUALITY AND USEFULNESS OF Federal Statistics." Both papers relate to this very important topic: bad or irrelevant data will almost always be worse than no data at all.

(4) A fourth interpretation is: "What the Public Needs to Know About Federal STATISTICS-PRODUCING AGENCIES." I had not thought of this one before seeing Dan Melnick's very interesting paper. His thesis is that how different agencies in the system go about providing what the public needs to know is shaped by their organizational perspectives, which, in turn, are shaped by their mission, history, experience, and expertise. The public needs to understand the different organizational perspectives in order to be able to assess the accuracy and appropriateness of the different outputs from the system.

**Interpretation (1): What the Public Needs to Know FROM Federal Statistics**

Users, if asked, will say they need to know everything about everything, but this expectation is not reasonable. Given resource constraints (and resources are always constrained, more so

now than ever), priorities must be set and choices must be made. What the data-using community *can* rightly expect is that the federal statistical system will continually monitor current and emerging policy issues and other important user needs and seek to determine the data collection vehicles that can best serve those needs-- rather than that the system will look first to the care and feeding of existing data collection vehicles and only secondarily consider whether they continue to be relevant to user needs.

Yet, from Jeff Passel's paper, it is clear that currently available immigration statistics only partly address today's policy needs in what is an increasingly contentious area of public concern. Agencies have made improvements in specific immigration databases, but no agency has a mandate to improve immigration statistics overall. The Immigration and Naturalization Service, with the perspective of an administrative agency, does not want this mandate. The Census Bureau, with the perspective of a statistical agency, could have but has not sought this mandate either, most likely because it already has a lot of other things on its plate. But the result for users is that the system of immigration statistics does not provide even basic numbers, such as how many people leave the United States each year, let alone more detailed information for policy purposes.

What Jeff Passel says for immigration statistics could be said for many other areas of current and emerging policy debate. Here are some examples from work at the Committee on National Statistics.

(1) **Data on Children** Historically, it has been difficult to study trends in the well-being of children--a topic of increasing public concern. One reason is that, until recently, longitudinal surveys have almost always followed adult members of households and not children, even

though many children these days move in and out of different family situations. Also, statistical agencies that collect data about children have tended to look only at the domain of direct relevance to their mission--school, health, work, and so on.

**(2) Data on Transportation** Historically, transportation data have been collected by "mode," that is, in terms of highways, rail, airlines, etc. Consequently, while there is a great deal of information about each mode, basic data on the total movement of freight or people in the country are lacking, and there are not comparable data across modes with which to address important public concerns about the safety, cost, and efficiency of the transportation system.

**(3) Data on Retirement Income Security** A growing policy debate that may be all-consuming in future years is how the United States will provide for the baby boomers' retirement. Employer-provided pension and health benefits are a key component of retirement income security, and employer behavior is one of the critical things to know about when considering alternative policies. There are many employer surveys conducted by such agencies as the Census Bureau and Bureau of Labor Statistics (BLS); the Department of Labor has an administrative database on pensions; and the Department of Health and Human Services is starting up a large new employer health insurance survey. But all of these data, while serving many useful purposes, do not in fact provide what the public needs to know to address important policy questions about employer-provided benefits and how employers might respond to different policy incentives.

In each of these areas, the statistical system recognizes the need to inject a cross-cutting, system-wide perspective, and is taking action. There is a recently-established federal interagency forum that is working hard to improve data on children (CNSTAT had a workshop on the topic, from which came a 1995 report on *Integrating Federal Statistics for Children*). In transportation, the 1991 Intermodal Surface Transportation Efficiency Act (ISTEA) established the newest federal statistical agency, the Bureau of Transportation Statistics (BTS). In its short life, BTS has made significant efforts to fill gaps in intermodal data and make accessible the data that have been buried away within USDOT. In immigration, there have been some efforts to bring agencies together, but Jeff Passel suggests that more needs to be done--specifically, that a single agency should be made the center of responsibility for improving immigration data. In the area of retirement income security policy, a CNSTAT study panel has just issued a report, *Assessing Policies for Retirement Income: Needs for Data, Research and Models*, that recommends an interagency working group, led by BLS and the Pension and Welfare Benefits Administration, to address how to provide the employer data that are needed to inform the policy debate.

These developments are all promising--or potentially promising. However, the U.S.'s highly decentralized statistical system--the most decentralized in the developed world--can make it hard to keep cooperative efforts going and, indeed, can make it hard to identify areas in which cooperative efforts are needed in the first place. Speaking in the context of Dan Melnick's paper, the individual agencies in the system bring valuable perspectives that ensure that it does not get locked into one approach or viewpoint, but without cross-cutting perspectives, the system is in danger of not providing what the public needs to know in many areas of policy concern. All of the participants in the system--the agencies, the OMB



Statistical Policy Division, of which Maria Gonzalez was a leading light for so many years, and such outside review groups as CNSTAT, need to work continually to find creative--and low-cost--ways to inject the necessary cross-cutting perspectives.

Here are some ideas for consideration by OMB and the system as a whole:

(1) Follow up Jeff Passel's recommendation and designate a central or lead agency for immigration statistics and other cross-cutting areas. A lead agency would not try to take over all of the data collection in an area, but it would have responsibility to identify data gaps and work with other agencies to develop strategies to fill them.

(2) Mandate that interagency forums and working groups establish specific action goals and timetables so that they do not lose momentum, which can all too easily happen as staff are pulled back to the perspectives of their own agency.

(3) Periodically charge an interagency task force to look around at what is happening in the U.S. society, economy, physical environment, and polity to identify emerging areas of policy concern that may require new or improved data. (Emerging issues will almost always require a cross-cutting perspective and consideration of how to revamp existing data systems.)

(4) Establish outside advisory groups in cross-cutting areas--that is, in addition to committees of outside experts that advise particular agencies or data programs, establish committees that advise in such areas as data for children or retirement income security policy needs.

(5) Interpret the Government Performance and Results Act (GPRA) to require statistical agencies not only to establish performance goals for their own programs and

services, but also to set goals for contributing to the effectiveness of the federal statistical system as a whole in providing what the public needs to know.

## **(2) What the Public Needs to Know TO FIND OUT About Federal Statistics**

This topic is addressed in the session on one-stop shopping. I would simply make a point about the potential of the Internet and World Wide Web to help the user community. Statistical agencies and OMB are already making innovative use of the Internet to provide large amounts of data to users; the agencies and OMB could use the flexibility of the Internet to even greater advantage to serve users' cross-cutting data needs. For instance, the White House Economic and Social Statistics Briefing Rooms, which provide key statistics and links to the agencies, are likely to be a popular route of access for users. The briefing rooms could be made even more useful by such simple means as adding cross-cutting categories--for example, statistics for children, the elderly, immigration. Such expansion in the number of categories could help users find needed data and also help the federal statistical system identify cross-cutting areas in which more or different data are needed to serve user needs.

In the move toward electronic dissemination of data, which the agencies are absolutely right to pursue, I would raise one caution. The agencies hope not only to serve users more effectively via the Internet, but also to save costs. However, such cost savings should not target agencies' in-house analysis capabilities. In-house statistical and subject-matter specialists are needed to work with their agency's data in order to provide benchmarks for users and valuable insights for their agency about needed improvements in data quality

and relevance.

**(3) What the Public Needs to Know About THE QUALITY AND USEFULNESS OF  
Federal Statistics**

This is a very important topic on which there could be an entire session. I would simply make a point again about the role of the Internet. Data dissemination via the Internet is exciting in that it is expanding the user community, but it is also scary in that it is expanding the opportunities for users to misapply data because they do not understand their quality and limitations. The federal statistical system has an obligation to provide information to Internet users about data quality and to make the issue of quality standards a priority concern.

Perhaps the Federal Committee on Statistical Methodology could work to develop standards for documentation of data on the Internet. Also, OMB could enforce such standards by including statistics on the White House Economic and Social Briefing Rooms only for those agencies that follow the standards. It is true that many users will not read the footnotes, but they cannot heed them or have the opportunity to be educated if information on quality does not accompany the data.

**(4) What the Public Needs to Know about Federal STATISTICS-PRODUCING  
AGENCIES**

It is unrealistic to expect that the public, or data-using community, will seriously study the sociology of the federal statistical system. However, it is important that federal statistical agencies and OMB periodically remind themselves of the insights from Dan Melnick's paper about the role of organizational perspectives in shaping how the system responds to data needs for policy and other public purposes. Each agency needs to be cognizant of and periodically reassess its own perspective. Each agency and OMB also needs to consider how best to inject the cross-cutting perspectives that are so necessary to enable the system to provide what the public needs to know.

**Council of Professional Associations on Federal Statistics**

**Seminar on Statistical Methodology in the Public Service  
November 12 - 13, 1996**

**WHAT THE PUBLIC NEEDS TO KNOW  
ABOUT FEDERAL STATISTICS**

**Summary of Remarks by  
TerriAnn Lowenthal, Discussant**

I am pleased to serve as a discussant this morning on the topic of, "What the Public Needs to Know about Federal Statistics." I will focus my remarks on two specific groups of users of Federal statistics: the general public, and Congress as the elected representatives of the general public. It is important to bear in mind that Congress both **directs** Federal statistical policy through agency authorizations, funding decisions, and data requirements, and **uses** (or misuses, as the case may be) Federal statistics to support policy and program development.

I want to address two basic issues that I think will help us get to the heart of the topic at hand. First, how do the public and Congress form their opinions about Federal statistics? And second, what is your responsibility, as experts in the field, to help these two large, generally non-expert, groups of users understand the value of the work you do?

Let's start by looking at what the public and Congress know about Federal statistics. Actually, it is much easier to look at what they **don't** know. Generally, you can assume that the level of knowledge about Federal statistics among Members of Congress and the people they represent isn't all that much different: very low. There appears to be a vast 'disconnect' between the prevalence and use of statistics and any understanding of how and why they are produced and what they represent.

It seems to me that, intuitively, people ought to have some level of faith in the numbers that assault them daily through the media and public officials. After all, basic math (to which most of the population has had some exposure) is taught as an exact science. There is usually a right answer and a wrong answer, a right number and a wrong number. So why are people so skeptical or distrustful of Federal statistics?

It's important to understand that Federal statistics probably suffer from the company they keep (although usually not by choice). It is unlikely that the public distinguishes between Federal statistics and statistics produced by a wide range of private sources. Jeff Passel aptly referred to the phenomenon of "dueling" figures, when reporters of official statistics have a need to find a competing statistic in order to present what they believe is a balanced view of an issue. Ironically, I think it's likely that the

public would tend to trust statistics produced by a Federal agency above all others. But it is difficult (and, I suspect, probably not worth the effort to non-statisticians) to sort through the barrage of statistics they may be confronted with every day through the news, advertising, politicians, and other sources.

The fact is that non-Federal (or otherwise unofficial) statistics are often produced for a narrow purpose, to support a specific policy outcome or cause. Some may be useful and produced through reliable means, and some may not be. But the reporting of these numbers is unlikely to explore such technical issues. And if the public, as the recipient of this information, is skeptical of the cause that the numbers support or the organization that produced the data, then it is likely to be distrustful of the statistics themselves, and any other numbers that happen to be reported at the same time.

Another, similar cause of this environment of mistrust may be found in the persona of the bearer of the news. Federal statistics often are released publicly not by the statistical agency itself but by a senior spokesperson of the parent department, often in conjunction with a policy initiative or political (small "p") statement. Therefore, the perception of objectivity that might otherwise attach to the statistics may be undermined if the underlying policy context is politically (capital "P") charged.

I also think that the media, a primary source of statistics for the general public and Congress, also contributes to the confusion surrounding the meaning of statistics in several ways. First, as I mentioned earlier, in the name of balance or objectivity, the media often finds a statistic to counter every statistic it reports, thereby diminishing the meaning of any given statistic and contributing to the perception that numbers only mean what the people producing them want them to mean.

Second, the media often doesn't take the time (or doesn't have the time) to understand the full meaning of statistics, how and why they were collected, and what they tell us and don't tell us.

And third, members of the media also can be skeptics, further undermining public confidence in the numbers to a significant degree. I have the perfect example, which I've been waiting to share with an audience like this for quite awhile.

As 5,000 statisticians were meeting in the Windy City this past August for the annual meetings of the American Statistical Association, an Op-Ed on welfare reform (which had just been passed by Congress) by columnist Mike Royko appeared in the Chicago Tribune. Mr. Royko started his column by conjuring up an image of the often-quoted statistic that 1.2 million American children would be thrown into poverty if the welfare reform bill being considered by Congress were enacted. This statistic, said Royko, was produced by "alleged experts." (This, by the way, is probably better than being called a "know nothing.") Heartless, if true, Royko said. But then he asks: How can anyone be certain about something like this? "How can these experts be sure it is



true? How do they know it won't be 1.4 million. Or 900,000. Or 2,365. Or 317," Royko asked.

Well, for one thing, Mr. Royko, even a non-statistician couldn't live with that large of a margin of error.

Mr. Royko then goes on to say, with unveiled sarcasm, that he doesn't doubt the findings because they came out of "some kind of think tank, where bright people ponder stacks of stats, charts, and other data, then make significant pronouncements." That sounds like statisticians aren't too far above lawyers in the public perception department! You've got tough skins, I know. The trouble is, if memory serves me correctly, that the study in question actually came out of the Department of Health and Human Services. But let's not let the facts get in the way of a good argument.

I suppose we should be comforted by Mr. Royko's statement two-thirds of the way through his column: "Since I'm not an expert and statistics give me a migraine, my guess is that nobody knows what the effect of the new welfare law will be." Mr. Royko, if you don't know what you're talking about, how come you get to write an influential column in a major newspaper?

Right after informing us (rather proudly) that he's not an expert, though, Mr. Royko proffers the following statement: "The biggest of all welfare problems is unmarried, uneducated, unemployable young women having illegitimate kids." Hmmmm. I wonder if Mr. Royko has any data to back up that statement? Or maybe it's just another guess?

Now, aside from the fact that this column really made me angry, I have a larger point to make. It seems to me that because the media is responsible for a great deal of the public's exposure to statistics, both Federal and non-Federal, it ought to be a little more responsible in how it portrays and reports numbers. And Federal statisticians have a responsibility here, too. Don't underestimate the value of the advice about dealing with the press that Victor Cohn gave you in this morning's keynote address. It may be tempting to dismiss reporters as hopeless when it comes to reporting fairly or fully on scientific matters, but please bear in mind that the media is the primary avenue of communicating the work you do to most of the nation, so you might as well strive to be friends, if not partners and allies.

Now that we've established that the public and Congress have a weak understanding of Federal statistics, we probably need to determine what they know about the Federal statistical system. The answer is "not much," I'm afraid. In fact, not many more Members of Congress than the people they represent even know that there is a Federal statistical system. There are several reasons for this.

First, oversight and funding responsibility among congressional committees is dispersed too widely for all but a handful of members to see the big picture (or even

know that there is a big picture). There are many different committees and subcommittees in both the House and Senate overseeing the work of the many Federal statistical agencies, and it is rare that any one committee or subcommittee is more than mildly aware of the activities of other panels with respect to oversight or funding of statistical agencies within an entirely different Federal department.

Second, the activities of individual Federal statistical agencies often are overlooked or lost within debates about larger agency or department programs. The decennial census is probably the one notable exception, but few Members of Congress can name any other activity of the Census Bureau. Even the relatively noteworthy monthly release of the employment and unemployment figures is generally viewed in a vacuum by most legislators. The statistics are eagerly anticipated and much analyzed both within and outside the halls of Congress, but rarely is there an effort to understand how the statistics were compiled and produced.

Speaking of the decennial census, don't look for any deep understanding among Members of Congress of that process, either. For example, while the Committees on Appropriations have been busy directing a reduction in number of questions on the census form, the rest of Congress was **adding** a question -- on children who are being raised by their grandparents -- to the 2000 census through the welfare reform bill. And to demonstrate the level of sophistication in Congress on matters statistical, the welfare reform provision requiring this new data directs the Census Bureau to collect "statistically significant data." And that's in connection with the decennial **and the mid-decade census**. You remember that much anticipated data-collection vehicle. If you blinked in 1995, you missed it!

The bottom line is that Congress doesn't even have a system for organizing its own activities affecting Federal statistics, never mind understanding the organization of the Federal statistical system!

And if you think Members of Congress have a minimal understanding of the Federal statistical system, generally, just think about their ability to grasp methodological issues. Many of us are familiar with the ongoing debate about the use of sampling and statistical techniques in the decennial census. It's often painful to watch, as Members of Congress make well-intentioned but usually uninformed pronouncements about the accuracy of sampling methods.

But perhaps the most egregious example of the failure to understand the meaning of Federal statistics, how they are produced, the constraints in producing them, and how Congress itself may have contributed to the limitations of the data it now rails against, is the debate (if one can call it that) over the accuracy of the Consumer Price Index. I never thought I would see the day when Congress would actually consider setting a statistic **legislatively**. I think this is outrageous; manipulation of data at its worst!

To me, this effort to change the CPI symbolizes nothing more than how palpable the fear over our most significant policy questions has become. Some of our most respected, thoughtful legislators see a legislative change in one of our most important statistics as the key to resolving some of the most fundamental, difficult, and titanic problems of social and economic policy -- namely how to control the burgeoning costs of entitlement programs, such as Social Security, that threaten our ability to balance the Federal budget. **You ought not to let that happen! The damage to the Federal statistical system may be irreparable if it does.** The non-Federal statistical community must speak up soon, loudly, and often, before this misguided effort goes too far.

I will close my remarks this morning by asking: What do you need to do to raise the level of public understanding of Federal statistics, as well as the level of public confidence in the ability of the Federal statistical system to provide the information that the country needs to develop and implement sound policy?

First, you need to establish useful working relationships with legislators and their staff, apart from the parent agency or department within which you operate. I know this may be difficult from an organizational or protocol standpoint, but you need to find a way to do it. Confidence in the reliability and objectivity of your products will increase in direct proportion to your ability to establish an independent identity as a statistical agency.

Second, rather than hiding from policy debates, you should strive to understand fully the policy context within which your statistical products will be used, now and in the future. You don't have to take sides in a debate, but you must be able to understand how your statistics will be used by people with differing goals and agendas, so that you can present and explain your product in a way that is most meaningful to the user.

Third, it is important to build meaningful and enduring relationships with the media. Try to understand the constraints under which reporters operate (such as the 5:00 p.m. deadline) and then develop innovative ways to increase the media's ability to use your statistics properly within those constraints.

And, finally, reach beyond your traditional audiences and professional relationships to establish meaningful and mutually beneficial relationships with public constituencies that are the end users and beneficiaries of the numbers you produce. Talk to the NAACP and the National Council of La Raza, to the U.S. Conference of Mayors and the National Association of Community Action Agencies, to senior corporate executives (as well as corporation statisticians or demographers) and editorial writers (as well as beat reporters). Reach out not just to professors of statistics, but to professors of history and political science and government.

When you hold conferences and seminars, such as today's session, include individuals and organizations outside of your traditional professional circles. Try to understand their concerns and perspectives, the worlds in which they operate, the people

they represent, and the objectives they are trying to achieve. These individuals and organizations are your window on the broader audience which you hope will understand what you do, and that understanding must be mutual in order to be effective.

Federal statistical agencies should go beyond their official advisory committees and communicate directly to the widest range of users and stakeholders possible, on their turf, in their environment. You will find, I think, that this kind of communication will help you move beyond the walls that we naturally build around our sciences, walls that often serve as barriers to a full understanding of what you do and why.

It is these kind of relationships that will help you build a strong and lasting foundation for the acceptance of, appreciation of, understanding of, and confidence in Federal statistics, and instill true meaning in the work you do for your government and country.