

Political Pressure and Statistical Quality: An American Perspective on Producing Relevant National Data

Stephen E. Fienberg¹

Abstract: Federal statistical agencies in the United States have long prided themselves on their independence and their ability to produce data in a neutral fashion. Yet even when their data are intended to serve national needs, the mandate to collect such data for most agencies comes from the political arena, i.e., from Congress or from politically appointed federal administrators. This paper focusses on a series of problem areas where substantial political pressures have imperiled the collection and dissemination of quality statistical information of material importance: quality control and

the welfare system, underenumeration and the decennial census, the extent and consequences of the AIDS epidemic, and employment discrimination litigation. Statisticians need to be aware that this pressure exists and to develop professional mechanisms to help our governmental colleagues recognize unreasonable pressure and resist it.

Key words: Committee on National Statistics; decennial census; federal statistics system; statistics and the law; undercount adjustment.

1. Introduction

This paper focusses on a mixture of good and possibly disturbing news regarding national statistics in the United States and their quality. The good news is that the national statistical establishment (which for

me includes the federal statistical agencies as well as the statistical infrastructures found in university and business and industrial settings) has survived almost eight years of budgetary pressure and occasional political hostility. Many statistical activities have been substantially strengthened during this period while others have suffered. But overall the national statistical enterprise was in a reasonable position to provide quality statistical data for policy purposes and public information, as the new administration of President Bush assumed control of the executive branch in January 1989. While the Committee on National Statistics (CNSTAT) at the National Academy of Sciences/National Research Council and other pro-

¹ Stephen E. Fienberg is Maurice Falk Professor of Statistics and Social Science and Dean of the College of Humanities and Social Sciences at Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.

Acknowledgements: Preparation of this paper was supported in part by the National Science Foundation under Grant No. SES-8701606 to Carnegie Mellon University. An earlier version was presented at a Sesquicentennial Session on National Statistics at the American Statistical Association Meetings, August 1988. Robyn Dawes, William Kruskal, Janet Norwood and two referees provided extremely helpful comments and criticisms on earlier versions of this paper.

fessional groups have labored long and hard to ensure the health of national statistics, most of the credit for this good news should be given to the professionals in the statistical agencies and to those of our colleagues who have worked with them on specific projects.

The possibly disturbing news is that national statistical data are permeated with extra-technical issues and values. For many readers this message is certainly not new. As Alonso and Starr (1987) recently noted:

Official statistics do not merely hold a mirror to reality. They reflect presuppositions and theories about the nature of society Lest there be any confusion, we should emphasize that to say official statistics are entangled with politics and social life is not to say that they are "politicized" in the sense of being corrupt. In some circumstances they may be corrupt, but that is not our point Our point, rather, is that political judgments are implicit in the choice of what to measure, how to measure it, and how to present and interpret the results.

Kruskal (1989) makes a strong case for *attempting* to separate technical statistical issues and citizenly or political values, even though I believe he agrees with me that the task is in the end an impossible one. Today, as in the past, American society faces a large number of basically social-political issues that can clearly be illuminated by statistical information. Yet there are multiple perspectives on any of these issues and most individual statistical data sets will reflect, to a greater or lesser extent, political and social perspectives whether or not we openly admit it and no matter how much care was exerted by the statisticians involved.

Federal statistical agencies in the United States have long prided themselves on their independence and their ability to produce data in a neutral fashion. I am reminded of a meeting several years ago where one of the speakers commented that "government

statistical data should be valueless." Actually he meant value-free, and when pressed he was unwilling to admit that the wording of questions in a survey or the definition of a variable inevitably reflects a perspective that is almost certain to have a political or societal component. For example, the national unemployment rate, produced by the U.S. Bureau of Labor Statistics based on data from the Current Population Survey, is constructed from a battery of questions designed to exclude *all* of those not considered to be in the labor force from both the numerator and denominator of the rate. While this approach makes considerable sense, it clearly reflects a societal attitude towards unemployment and the need to search for work, i.e., the basic measure reflects social values. Starting from a different social perspective we might easily be led to use a different battery of questions and come up with what is in effect a different definition of the unemployment rate.

Despite the fact that national data cannot be value-free, I believe that most federal statistical agencies do an excellent job in insulating their data from unnecessary political aspects, including the political views of those working for the agencies. Yet, even when they succeed in their political "neutrality," most federal agencies are typically collecting data in the context of a mandate that comes from the political arena, i.e., from Congress or from politically-appointed federal administrators. I would argue that statisticians who collect data mandated by others can, at best, *attempt* to remain "disinterested" and impartial, but they certainly cannot uncouple their efforts from societal values and perspectives. Moreover, once produced, statistical data "enter the political fray on behalf of social interests" (Prewitt (1987, p. 262)) and thus the statistician's job does not end with the produc-

tion of impartial reports or data summaries. A good example here is the Office of Management and Budget (OMB) definition of poverty linked to a specific Bureau of the Census statistical series, and the annual production of data on families in poverty. Who is naive enough to believe that such data will not be used for partisan political purposes? Are we well served by the lack of alternatives to this admittedly flawed definition and the absence of empirical estimates to go with them?

The remainder of this paper focusses on a series of problem areas where substantial political pressures have imperiled the collection and dissemination of quality statistical information of material importance: quality control and the welfare system, under-enumeration and the decennial census, the extent and consequences of the AIDS epidemic, and Title VII employment discrimination litigation. These and other issues have been before CNSTAT and its panels over the past decade, and my discussion of them draws upon my work for the committee since 1978, first as member, then as Chair, and finally as a member of its Panel on Decennial Census Methodology. My personal experiences suggest to me that statisticians need to be aware that this pressure exists and to develop professional mechanisms to help our colleagues recognize unreasonable pressure and resist it. We might even try to analyze statistically the types and contexts in which this pressure is encountered as well as the ways with which it is dealt.

2. Examples of Political Pressure in Statistical Settings

Having stated my basic thesis, I would like to illustrate it by means of a series of anecdotes. I argue by example for at least two reasons. First, no one to date has done

any systematic thinking on these issues and thus it is not clear how to carry out a careful empirically based investigation (c.f. Dawes (1989)). Second, the examples are vivid and I hope that they will stimulate readers to think about the issues of politics and statistics. My examples have been clearly chosen to illustrate instances where I think that political pressure and intervention has been harmful to the quality of national statistical data.

The anecdotes that follow should be viewed in the context of a federal statistical system that has a number of mechanisms to protect it from political interference. One of the best known of these is the OMB *Circular on the Compilation, Release, and Evaluation of Principal Economic Indicators*. There are also instances where such intervention has taken place but has, in retrospect, produced reasonable or even positive results. For example, in his memoir, *An American Life*, Jeb Stuart Magruder (1976 (p. 102)) relays the following story:

One battle was with the Bureau of Labor Statistics (BLS), which puts out the monthly figures on employment. These figures had traditionally been put in terms of the unemployment rate, which in 1970 was holding steady at about six percent. At the same time, while the unemployment rate was constant, the number of people holding jobs was at an all-time high, and getting higher each month We saw no reason why BLS couldn't stress the positive fact – a record number of jobs – at its monthly briefings, but BLS did not agree. I spoke with its director, Geoffrey Moore, several times. I tried persuasion, and when that didn't work I finally told him, "Look, Mr. Haldeman says this is what the President wants. If you want to argue with Mr. Haldeman, fine. But if not, change your style."

The BLS changed its style It was a small victory, one achieved after a great deal of pushing, it seemed to us outrageous that a bureau of the Labor Department

should defy a reasonable request by the President.

Janet Norwood, the current Commissioner of BLS, has suggested that Magruder's recollection is not quite correct but it may well be that, as long as he and others perceive it to be correct, the perception of the intrusion of politics on statistics is reinforced. It is also interesting to note that Janet Norwood's monthly appearances before the congressional Joint Economic Committee to announce the latest data on employment and unemployment continue to feature both the number of jobs and the unemployment rate.

The following subsections describe four problem areas where substantial political pressures have imperiled the collection and dissemination of quality statistical information of material importance: quality control and the welfare system, underenumeration and the decennial census, the extent and consequences of the AIDS epidemic, and employment discrimination litigation. The discussion of underenumeration and the decennial census is the most extensive both because of my familiarity with the issues and because of its potential implications for national statistics.

2.1. Quality control in family assistance programs

During the 1960s the federal government initiated a series of family assistance programs to be administered by the states with federal financial support and oversight. The three principal programs (Aid to Families with Dependent Children, Food Stamps, and Medicaid) serve related purposes and overlapping populations. For each a special quality control (QC) system was established by the federal government to address concerns about ineligibility, fraud, and abuse.

The following remarks draw heavily on a review of these QC programs carried out by a panel of the Committee on National Statistics (CNSTAT) chaired by John Neter (see Affholter and Kramer (1987) and Kramer (1988)).

While there is a shared conception of a quality control system within the statistical community, this conception is not always well understood outside. The QC structure for family assistance programs has, from the outset, focussed on only a small part of what most statisticians would include in a complete system. Specifically they deal with the accuracy of eligibility and benefit determinations by the states. Each month, state officials take samples from two universes of cases: active cases (units receiving benefits), and negative cases (units denied benefits and terminations). These samples are used to calculate various error rates in the program administration and are then reevaluated by federal agencies to establish substantial monetary sanctions against states with poor QC performance.

These systems as described would not in and of themselves be problematic, except for the fact that they use statistical concepts and methods to achieve political as well as quality control aims. The political and punitive aspects of the QC systems have prevented the implementation of broad-based quality improvement by overriding technical judgment in several areas. For example, the current system basically ignores the differential precision of the estimated error rates produced by different states. Moreover, QC programs at the state level are placed in the position of trying to achieve two conflicting set of goals: the first is a statistical goal focussing on system improvement; the second is a regulatory goal of ensuring accountability for payment accuracy. Using one data collection system to achieve both

goals creates considerable problems; the political pressure to achieve the regulatory goal compromises the statistical goal, often severely. In addition, the sanction system is based only on overpayment errors rather than on overpayment and underpayment errors as well as improper denials. This choice of sanction system serves to reinforce the focus on a specific form of regulation and not on overall quality improvement and was made despite statistical advice on the need for a more comprehensive system. It will be interesting to see the administrative response to the panel's strong recommendation to revamp the financial accountability of the state QC programs by including all sources of payment inaccuracy.

An interesting sidelight that arose in the CNSTAT panel review of these programs was the controversy over the use of two-phase regression estimators for the state error rates in the three programs. Whether one supports the use of this particular estimator (that had been originally recommended for use to the federal agencies by Westat, a non-government survey and statistical consulting company) depends heavily on how one chooses to view the accuracy of the error estimates in the state and federal reviews. Different perspectives are rooted in fundamentally different assumptions that are not immediately verifiable from the statistical design of the sampling procedures or from the data themselves. The panel proposed a set of changes which would do away with the two-stage evaluation system and thereby sidestepped the resolution of a basically unresolvable statistical dispute.

In many ways the impingement of political pressure on the QC programs for family assistance programs is subtle and in other ways it is quite overt. But one need only look at the financial stakes to understand why statistical integrity might easily get lost

in the shuffle. The CNSTAT panel's two reports on these programs may lend support to improved statistical efforts in this area.

2.2. *Underenumeration and the Decennial Census*

Concerns about the accuracy of the census counts in the United States have existed almost as long as the census itself. In Vol. 2 of the *Journal of the American Statistical Association* General Francis A. Walker (1890), Superintendent of the U.S. Censuses of 1870 and 1880, wrote about the undercount of Blacks in the 1870 census. He elicited one of the earliest statistical proposals for adjustment for the undercount from H. A. Newton and H. S. Pritchett, both of whom used the method of least squares to fit a third degree polynomial to census data for 1790 to 1880 and then measured the undercount for 1870 as a residual from the fitted curve. See Stigler (1988) for further details.

There are basically two quantitative techniques that have been used to estimate the undercount at a national level: demographic analysis and the dual-system or capture-recapture technique. Demographic analysis combines birth, death, immigration, and emigration records with other administrative records to carry forward the population from one census to the next, deriving an estimate of the overall population size, and thus the undercount. The methodology can be used to provide population and undercount figures by age, race, and sex, but only at the national level. Demographic analysis cannot be used to provide reliable state, regional, and local estimates, principally because of the absence of accurate data on migration. In the dual-system estimation approach, those included in the census are matched with a second source (e.g., a random sample of the popula-

tion or a list based on administrative records) and this information is used to produce an estimate of those missed in both sources and thus an estimate of the undercount in the original census. This technique can be used directly at the national level as well as at state and sub-state levels. For further details see Fay, Passel, Robinson, and Cowan (1988) or Feinberg (1989a).

Beginning with the 1940 census, the Bureau of the Census estimated the size of the undercount by race, using demographic analysis. The estimated differential undercount between Blacks and Whites has remained between 5% and 6% up through the 1980 census (see Fay, et al. (1988)). While the explanations for the undercount have changed over the decades, as techniques for taking the census have changed, the differences in undercount among population groups has continued to be a major concern for demographers and statisticians. The dual-system estimation approach was used in conjunction with the 1980 census to evaluate population coverage as part of what was called the post-enumeration survey (PES) program. In this PES program, a sample of 110,000 households from the census, selected in clusters of approximately 10 housing units per enumeration district, was matched with data from households in the April and August Current Population Survey, each containing approximately 84,000 households, and estimates of the undercount were produced for the U.S. as a whole as well as for all 50 states and several large local areas. Substantial controversy surrounded the subnational undercount estimates that emanated from the 1980 PES program (for example, see Ericksen and Kadane (1985) and Freedman and Navidi (1986)).

Prior to the 1980 census there was extensive discussion in the statistical community

regarding the advisability of adjusting the census counts to correct for the undercount, and a decision was made shortly before the reporting deadline, in December 1980, not to adjust the results. A lawsuit was filed on census day by the city of Detroit requesting that the 1980 census be adjusted for the undercount, and this action was followed by 52 others, 36 of which requested adjustment. One of these cases, brought by the state and city of New York, gained considerable attention, with a large number of statisticians testifying for and against adjustment. The New York lawsuit, known as *Cuomo v. Baldrige*, ultimately went to trial in January 1984, but the judicial opinion was not issued until December 1987. The judge ruled that no adjustment need be made. He argued that, because statisticians and demographers can and do disagree on the reliability of an adjustment of the 1980 census, it would be inappropriate for the court to substitute its judgment for that of the experts at the Census Bureau.

Simultaneously with these activities the Census Bureau launched a major research program to improve the methodology used for census adjustment and it commissioned CNSTAT to establish a Panel on Decennial Census Methodology, whose charge included the review of the census undercount research program. The panel's 1985 report (Citro and Cohen (1985)) outlined the basic issues that needed to be addressed in the adjustment research program. Subsequently, the panel reviewed the proposed methodology developed by the Census Bureau staff for adjustment in 1990, and its implementation in two separate pretests. This methodology was based on a newly designed post-enumeration survey and the use of dual-systems estimation, and was designed to overcome problems with the PES/dual-systems approach used in the 1980 census (see Chil-

ders, Diffendal, Hogan, Schenker, and Wolter (1987) and Hogan and Wolter (1988) for a description of this methodology and its technical justification). In two separate letter reports, in 1986 and 1987, the CNSTAT panel made a positive assessment of the technical feasibility of the planned adjustment methodology. In addition, at two separate 1987 congressional hearings, bureau officials and independent statistical experts testified on issues relating to the technical feasibility of adjustment (see Wallman (1988a)).

In her presidential address to the American Statistical Association, Barbara Bailer (1988) noted:

A sizable group of eminent statisticians now believe that adjustment of the 1990 census is feasible, that it has been successfully demonstrated in test censuses and that it would substantially improve the accuracy of the 1990 census. Those who press for adjustment say that an undercount in 1990 is inevitable, a view that the Census Bureau largely shares. They argue that even an imperfect adjustment will be a move in the right direction and will increase the accuracy of census data for its many uses.

The key point here is that, unlike in 1980 when there was a serious split in opinion among statisticians knowledgeable about census-taking and adjustment methodology, the vast majority of statistical experts, inside and outside the Bureau of the Census, concluded that adjustment methodology proposed for use in 1990 was statistically sound.

Then, in October 1987, the Department of Commerce, in which the Bureau of the Census is located, announced that the 1990 census would not be adjusted for the differential undercount. This decision has been widely criticized by many in the statistical community (including myself). A special congressional hearing was held in March 1988, focussing on the technical feasibility

of the adjustment methodology and reviewing the decision not to plan for adjustment (see the report in Wallman (1988b)).

The issue I wish to focus on here is the effect of political pressures on the Census Bureau in the decision on whether or not to plan for adjustment in 1990. I begin by noting that the decision not to proceed with plans to adjust was announced not by the Census Bureau but by a political official, the Under Secretary of the Department of Commerce, the department in which the bureau is situated. A March 1988 congressional hearing provided documentation on the deliberations at the bureau prior to the announcement and substantiates the charges of political interference.

- There was virtually complete agreement among the statisticians in the Undercount Research Staff and others in the Statistical Standards and Methodology Division that the adjustment methodology had been successfully implemented in the Los Angeles Census Pretest and "that adjustment was technically sound and feasible."

- In May 1987, the bureau's Undercount Steering Committee recommended proceeding with plans for adjustment, and Census Director John Keane made the decision to proceed with appropriate plans.

- The bureau's plans were overruled by political officials in the Commerce Department.

It is beyond doubt that those most likely to benefit from the decision not to plan for adjustment are the Republicans, whose administration made the decision in opposition to professional statistical advice. Even though Census Director John Keane was a political appointee of the Republican administration he appears to have resisted pressure to take a politically expedient posi-

tion on adjustment. It was therefore especially disconcerting to many of those who had closely followed the planning activities for possible adjustment that the Republican administration attempted to put a scientific gloss over what had become an intensely political decision. To make matters worse, several individuals at the Census Bureau attempted to "rewrite history" by belatedly setting to paper a rationale, which they claim to have been true back in the spring of 1987, for not having proceeded with plans for adjustment.

In October 1988, the Commerce Department decision was challenged in a new lawsuit brought by the City of New York and other state, county, and local governments. They alleged that the decision not to adjust was arbitrary and not based on technical grounds. The trial was scheduled to take place in July 1989 but a last minute settlement was reached which included the following plans:

- the October 1987 decision was withdrawn and there will be a new decision on whether or not to carry out a statistical adjustment of the 1990 census,
- standards for adjustment would be publicly announced prior to the census,
- all census data released or published prior to an adjustment decision will carry a disclaimer, saying that the results are subject to possible correction,
- the Secretary of Commerce will appoint an independent panel of experts to advise on adjustment-related activities.

For further details on the settlement agreement, see Fienberg (1989c).

Clearly the story about politics and the 1990 census is not over. First, there is a professional consensus that there will be a differential undercount in the 1990 census and many statisticians believe it will exceed that for 1980. Second, if left uncorrected this

differential undercount is likely to effect statistical programs all over the federal government for at least ten years. This is because census results are used not only for reapportioning Congress and state and local legislative bodies, but also as the frame for virtually every government-sponsored household survey and for the allocation of various forms of federal funding. Third, the settlement in the New York City lawsuit does not preclude further legal challenges for and against adjustment. Finally, we need to wait in order to learn whether the Bureau of the Census is able to recover from this encroachment on its activities in order to carry out a quality post-enumeration survey and implement it for correcting the census counts.

2.3. The extent and consequences of the AIDS epidemic

Acquired Immune Deficiency Syndrome (AIDS) is the medical description for the final stages of a series of diseases caused by a human retrovirus known as HIV. This virus attacks the immune system, damaging its ability to fight other diseases. In 1981, the Centers for Disease Control (CDC), which is the lead federal agency for monitoring the outbreak and spread of infectious diseases, received reports of a number of cases of Kaposi's sarcoma, involving previously healthy young male homosexuals with severely compromised immune systems. This previously unidentified disease was subsequently labeled AIDS and the CDC has become the lead agency tracking the spread of the AIDS epidemic.

The term AIDS is typically used to denote various forms of the HIV infection even though progression from one state to the next is not automatic. A relatively large group of individuals possess the HIV virus (the first stage), and a much smaller group of those infected have gone on to develop

AIDS Related Complex while even fewer have actually suffered from AIDS itself. The HIV virus is known to be transmitted through sexual contact, parenteral exposure to blood and blood products, and from mother to child during the perinatal period, but information on rates of transmission is poor at best (Curran et al. (1988)). The population groups exhibiting the greatest incidence of AIDS are homosexual men, intravenous drug abusers, and hemophiliacs.

While the CDC has the principal responsibility for the collection of data on AIDS, it has a rather narrow focus. CDC has been joined by an array of federal government agencies interested in various aspects of the AIDS epidemic and its effect on individuals and families, the health care system, the economy, and society more broadly. Except for some of the efforts on assessing the medical cost associated with AIDS, most of these data collection efforts are only now getting underway.

Many people, for example, are surprised to learn that we still have only crude knowledge about how many people are infected with the HIV virus. At the moment, we know so little about the transmission and development of the disease that the uncertainty associated with any forecasts is often as large as the forecasts themselves. For example, the July 22, 1988 issue of the *New York Times* ("Halving of Estimate on AIDS Is Raising Doubts in New York") reports that the New York City Health Department revised its estimate of the number of people in New York City who have the HIV infection from 400,000 to 200,000, triggering renewed debate over the accuracy of the figures being used to set local and national policies with respect to AIDS. The revised figures are based heavily on data from the Kinsey survey of sexual behavior (conducted about 40 years ago) whose statistical quality

is questionable at best (e.g. see Cochran, Mosteller, and Tukey (1953)), and data from San Francisco on infection among homosexual and bisexual men. The AIDS figures for New York City were revised again less than a month later (see the article, "New York Again Revises Its Aids Virus Estimate," in the August 11 issue of the *New York Times*) to somewhere between "149,000 and 226,000" on the basis of a change in the estimate of the number of infected homosexual and bisexual men!

To address the basic issue of population-based infection rates, the National Center for Health Statistics (NCHS) has developed plans for a national seroprevalence survey of HIV infection (Weeks, Horvitz, Hurley, and Wright (1989)). The pilot test of this survey was originally scheduled for Washington, DC but, when local health officials publicly raised concerns about the survey on the grounds related to the confidentiality of the respondents, politically-appointed officials at CDC, which is NCHS's parent organization, canceled the test. In part they feared that the public debate would effect data collection in another CDC nonprobability AIDS related project (see the description of this other survey in Stoto (1989)). The pilot was later implemented in Allegheny County, Pennsylvania and the results were quite encouraging. The screening response rate for occupied households was 95.1%, and, for those containing an eligible sample person, 85.4% gave a blood sample and completed the questionnaire. Thus the overall response rate was 81.2% (Research Triangle Institute (1989)). One of the problems facing a possible NCHS survey is a Public Health Service Guideline on mandatory notification, i.e., individuals who participate in the study must be told if they are HIV-positive (the only recourse is not to participate). Another problem is that the

possible concentration of HIV infection in that part of the population with severe census undercount problems may severely bias estimates from such a survey. Finally, issues of confidentiality and anonymity continue to prevent the use of reverse record checks to test the accuracy of responses to behavioral questions that are part of the survey questionnaire.

There are a number of other planned data collection efforts about AIDS that are worthy of note, including studies of sexual attitudes and behavior. A recently released report from the National Research Council (Turner, Miller, and Moses (1989)) provides an excellent description of what is needed and why. Experts have been trying to launch these efforts for some time claiming that most of what we “know” in the behavioral area is “fake-lore” on sexuality that is based on bad data and bad data collection practices. The funding of the proposed survey has been approved by the National Institutes of Health but the Office of Management and Budget is currently withholding approval of the survey questionnaire because of the complaints from several congressmen and senators who have branded the survey as “pornographic.” For example, in a recent letter in *Science*, Congressman Wm. Dannemeyer (1989) has publicly stated:

What I have said is that the survey seems more apropos for the pages of a pornographic magazine, with explicit sexual questions for a very limited audience, than as something to be passed off as a scientific study.

Many government officials do not yet understand that the social and moral dimensions of problems surrounding the AIDS epidemic are critical to an understanding of what statistical data to collect and how to interpret them. And, as Mary Grace Kovar has pointed out, there is also the epidemic of

fear – one which is less visible than the AIDS epidemic itself, but whose social impact may be greater. The ability of federal agencies and university-based researchers to gather such social, behavioral, and attitudinal information is heavily influenced by the social and political attitudes of administration officials. Unfortunately to many of them, AIDS remains a social threat concentrated among some of the most undesirable groups in society.

Given the seeming extent of the AIDS epidemic, it is rather surprising that such data were not collected several years ago. Why are we only now getting around to measuring the prevalence of HIV infection? The answer that many people give to this question is politics and political pressure within the executive branch to conform the original administration position that the spread of AIDS was not a serious matter for most Americans. Because of the linkage of AIDS with intravenous drug usage aspects of statistical measurement are also bound up in the administration’s “war on drugs.” A recent article in *Science* describes the political pressure being brought to bear on the National Institute on Drug Abuse (NIDA) in connection with its role in the war on drugs (Booth (1988)). As a result of recommendations from the President’s own Commission on AIDS and those coming out of several related committees and groups, this negative political perspective has been somewhat curbed and funding for AIDS-related data collection is now growing rapidly. For example, NIDA will spend approximately 38% of its budget in the last fiscal year on AIDS-related activities.

2.4. Title VII employment discrimination litigation

The example is basically different from the preceding ones in that it is linked to national

statistics in quite a different way from the decennial census or quality control for family assistance programs. The focus of my concern here is the involvement of statisticians as expert witnesses for parties in employment discrimination litigation. The two links with the other examples are the extra-technical considerations for statisticians testifying in such adversarial proceedings and yet another CNSTAT study, by its Panel on Statistical Assessments as Evidence in the Courts (see Fienberg (1989b)).

As the CNSTAT panel reports, there was a remarkable growth in the use of statisticians as expert witnesses during the late 1970s and early 1980s, especially in connection with litigation brought under Title VII of the Civil Rights Act of 1964, a federal statute which states that it is an unlawful employment practice for an employer to "fail or refuse to hire or discharge any individual, or otherwise to discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual's race, color, religion, sex, or natural origin." Much of the evidence used in Title VII cases, especially those involving class actions, is statistical in nature and it has become typical for both plaintiffs and defendants to hire expert witnesses to present statistical evidence in support of or in opposition to a claim of employment discrimination. The statistical data in these cases tended to consist of employer records on applications and former and present employees, as well as labor market "availability" data often drawn from national sources, such as the Bureau of Labor Statistics, and state and local labor statistics offices.

In the typical Title VII case, the plaintiff's expert uses a data base and presents an

analysis supportive of the plaintiff, followed by a rebuttal by the defendant's expert who often presents a somewhat different data base and often a dramatically different statistical analysis. For example, the plaintiff's expert may use multiple regression analysis and five predictors to argue that men and women similarly situated received different levels of compensation while the defendant's expert uses reverse regression and seven variables to reach an opposite conclusion (for a discussion of the rationale for these two different analytical stances see Conway and Roberts (1986) and Dempster (1988)). The ensuing battle of the experts often turns the statisticians into advocates and invokes forms of political and social pressure on them in a fashion that is analogous to that which I have discussed in a government context above.

What becomes clear from an examination of actual Title VII cases (see Fienberg (1989b)) is that the very nature of the adversarial system draws the expert witness away from neutrality and objectivity. The process begins with the briefing of the expert by counsel who invariably presents the facts of the case from the perspective of the client. Moreover, access to various types of data, e.g., company employment files, is often a function of the party with whom the expert is working. As an expert's involvement with a case grows, so too may friendship with counsel. Fisher (1986) reminds us that "Particularly because lawyers play by rules that go beyond those of academic fair play, it becomes insidiously easy to see only the apparent unfairness of the other side while overlooking that of one's own side. Continuing to regard oneself as objective, one can slip little by little from true objectivity."

Addressing the broad range of litigation areas involving statistical testimony, Paul Meier (1986) describes these extra-technical

dimensions of the statistician as expert witness quite well:

As we have just seen, the professional integrity of the expert witness and, through him, of the profession that he represents is not well protected by the courts and hardly at all by counsel. But before we assume too readily that simple morality and personal ethics will be an adequate substitute, we should reflect for a bit on . . . *corrupting influences* First, there is the fact that the expert witness is playing someone else's game and, inevitably, has to accept the rules as he finds them. His instructor in these matters is, of course, his client's counsel, and the witness is ill-equipped to resist the role of adversary when his lawyer thrusts it upon him Among the most difficult of the corrupting influences to deal with is what I call *aggrandizement*. In Title VII cases . . . the Supreme Court has placed the statistician in the key role. Long ignored and treated with contempt in literature and in the courts, the statistician has been elevated to Olympian levels . . . He will be tempted to ignore or minimize those qualifications that he might emphasize in an academic setting, he may fail to emphasize schools of thought other than his own, and he may lay claim to overly broad scope for the inferences he draws.

Meier goes on to describe a host of other additional influences added by the adversarial system including bribery, flattery, co-option, and personal views. Meier also advocates the use of personal and professional codes as a way to defend the integrity of statistical testimony.

A reviewer of an earlier draft of this paper correctly pointed out that we should not simply view the statistician as an innocent noble soul corrupted by immoral or at least amoral lawyers. Rather we must recognize that some statisticians are fully aware that they are being hired as expert witnesses because their testimony will help win the case and they exploit this situation by securing high fees for their services. While we

should not condemn such professional colleagues, we should also not regard them as totally innocent of co-option.

That these issues extend far beyond the area of Title VII employment discrimination litigation is quite clear to anyone who has been involved as a consultant or an expert witness in the legal arena. For an illustration of how statisticians take markedly different and polarized perspectives on statistical proof of cancer causation and environmental issues, see the recent paper by Freedman and Zeisel (1988), which is followed by extensive comments from others. (This exchange grew out of a law suit involving DDT contamination that was ultimately settled out of court.) The CNSTAT panel report reviews two other major areas involving statistical testimony, antitrust litigation and environmental issues, as well as touching upon such areas as taxation and identification evidence in criminal cases.

In addition to the use of personal and professional codes of conduct as methods to insulate statisticians from the distortion of professional standards and practice in the legal setting, other mechanisms can be invoked such as the use of court-appointed experts. The report of the Panel on Statistical Assessments as Evidence in the Courts devotes considerable attention to this issue. Nonetheless the ethical issues surrounding the role of statisticians as expert witnesses are worthy of much greater attention by the statistical profession.

3. Some Lessons: Recognizing Political and Social Perspectives and Resisting Political Pressure

The basic themes of this paper have been three-fold. First, virtually all of what we might label as national statistics, whether produced by federal government agencies or

by those in universities or the private sector, will reflect political and social perspectives and values. Second, despite the fact that national data cannot be value-free, I believe that most federal statistical agencies do an excellent job in insulating their data from unnecessary political aspects. Third, there are a series of problem areas where substantial statistical pressures have imperiled the collection and dissemination of quality statistical information of material importance. The four illustrations in this paper have been quality control and the welfare system, underenumeration and the decennial census, the extent and consequences of the AIDS epidemic, and Title VII employment discrimination litigation. It is my view that statisticians need to be aware that this pressure exists and to develop professional mechanisms to help our colleagues recognize unreasonable pressure and resist it.

The theme that relevant national statistical data cannot be value-free is not a new one, as I noted at the outset, but the recognition of this perspective from within the statistical community is not wide-spread. A shared understanding of the issues I have tried to raise in this paper should help to strengthen rather than undercut the quality of national statistics programs. Writing in a related context, the sociologist Stanley Lieberson recently noted (1988): "It is perfectly appropriate at any and all times to ask about the social underpinnings of knowledge, but that is not the same thing as asking whether the knowledge is valid – at least valid under the broad criteria of what the society is able to define as 'true' in its current state of affairs." He goes on to note that "we are all too willing to allow socially and politically relevant subjects to be studied in the form in which the society states the questions, rather than by the way that our knowledge tells us to approach the problem."

What remains is for us as statisticians is to learn how to recognize the overt forms of political and social pressure that may subvert the production of quality statistical data of relevance to the problems such as in the examples discussed in Section 2. In the context of government statistical agencies we need both strong professional leadership that is capable of protecting the agencies from unreasonable political interference as well as congressional mandates that make clear the responsibilities for statistical quality and aspects of independence that are vested in the agencies. Professional statistical organizations have a special role to play in support of our colleagues in the government sector as they develop and enhance codes of ethics for professional statisticians. They can also monitor activities in the government statistics domain and thereby provide support for statisticians and agencies that appear to be the targets of unreasonable political pressure and interference. They might even try to analyze statistically the types and contexts in which this pressure is encountered and the ways with which it is dealt.

Finally, I note that the issues of political pressure and the quality of statistical data extend far beyond what is usually taken to be the domain of national statistics in the United States. The example of the battle of experts in the context of Title VII discrimination litigation was chosen to be illustrative and to provide an indirect link to the American federal statistics scene. The issues here are, in a real sense, universal; they are not simply American and not restricted to activities of national statistical agencies. The purity of the mathematical formulation of statistical ideas vanishes once we actually measure phenomena in the real world and attempt to set our analyses into an interpretive context.

4. References

- Affholter, D.P. and Kramer, F.D. (eds.) (1987): *Rethinking Quality Control: A New System for the Food Stamp Program*. (Committee on National Statistics, National Research Council), National Academy Press, Washington.
- Alonso, Wm. and Starr, P. (eds.) (1987): *The Politics of Numbers*. Russell Sage Foundation, New York.
- Bailar, B.A. (1988): Statistical Practice and Research: The Essential Interactions. *Journal of the American Statistical Association*, 83, pp. 1-8.
- Booth, Wm. (1988): War Breaks Out Over Drug Research Agency. *Science*, 241, pp. 648-650.
- Childers, D., Diffendal, G., Hogan, H., Schenker, N., and Wolter, K. (1987): The Technical Feasibility of Correcting the 1990 Census. Paper delivered at the Annual Meetings of the American Statistical Association.
- Citro, C.F. and Cohen, M.L. (eds.) (1985): *The Bicentennial Census. New Directions for Methodology in 1990*. (Committee on National Statistics, National Research Council) National Academy Press, Washington D.C.
- Cochran, Wm.G., Mosteller, F., and Tukey, J.W. (1953): Statistical Problems of the Kinsey Report. *Journal of the American Statistical Association*, 48, pp. 673-716.
- Conway, D.A. and Robert, H.V. (1986): Regression Analysis in Employment Discrimination Cases (with discussion). In *Statistics and the Law* edited by M.H. DeGroot, S.E. Fienberg, and J.B. Kadane, Wiley, New York, pp. 107-195.
- Curran, J.W., Jaffe, H.W., Hardy, A.M., Morgan, W.M., Selik, R.M., and Donero, T.J. (1988): Epidemiology of HIV Infection and AIDS in the United States. *Science*, 239, pp. 610-616.
- Dawes, R. (1989): Probabilistic Versus Causal Thinking. In *Thinking Clearly About Psychology: Essays in Honor of Paul E. Meehl*, edited by D. Chicchetti and W. Groves (in press).
- Dempster, A.P. (1988): Employment Discrimination and Statistical Science (with discussion). *Statistical Science*, 3, pp. 149-195.
- Dannemeyer, W.E. (1989): Proposed "Sex Survey." *Science*, 244, p. 1530.
- Ericksen, E.P. and Kadane, J.B. (1985): Estimating the Population in a Census Year: 1980 and Beyond (with discussion). *Journal of the American Statistical Association*, 80, pp 98-131.
- Fay, R.E., Passel, J.S., Robinson, J.G., and Cowan, C.D. (1988): *The Coverage of Population in the 1980 Census*. Bureau of the Census, U.S. Department of Commerce, Washington D.C.
- Fienberg, S.E. (1989a): Undercount in the U.S. Decennial Census. In S. Kotz and N.L. Johnson, eds., *Encyclopedia of Statistical Sciences, Supplemental Volume*, Wiley, New York, pp. 181-185.
- Fienberg, S.E. (ed.) (1989b): *The Evolving Role of Statistical Assessments as Evidence in the Courts*. (Committee on National Statistics and the Committee on Research on Law Enforcement and the Administration of Justice, National Research Council) Springer-Verlag, New York.
- Fienberg, S.E. (1989c): An Adjusted Census in 1990? *Chance*, 2 (3), pp. 23-25.
- Fisher, F.M. (1986): Statisticians, Econometricians, and Adversary Proceedings. *Journal of the American Statistical Association*, 81, pp. 277-286.
- Freedman, D. and Navidi, W.C. (1986): Regression Models and Adjusting the 1980 Census (with discussion). *Statistical Science*, 1, pp. 3-39.

- Freedman, D. and Zeisel, H. (1988): From Mouse to Man: The Quantitative Assessments of Cancer Risks (with discussion). *Statistical Science*, 3, pp. 3–56.
- Hogan, H. and Wolter, K. (1988): Measuring Accuracy in a Post-enumeration Survey. *Survey Methodology*, 14, pp. 99–116.
- Kramer, F.D., ed. (1987): From Quality Control to Quality Improvement in AFDC and Medicaid. (Committee on National Statistics, National Research Council) National Academy Press, Washington.
- Kruskal, Wm.H. (1989): The Statistician as Citizen. Paper presented at ASA Sesquicentennial Session, Joint Statistical Meetings, New Orleans, August, 1988.
- Lieberson, S. (1988): Asking too Much, Expecting too Little. *Sociological Perspectives*, 31, pp. 379–397.
- Magruder, J.S. (1976): *An American Life*. Atheneum, New York.
- Meier, P. (1986): Damned Liars and Expert Witnesses. *Journal of the American Statistical Association*, 81, pp. 269–276.
- Prewitt, K. (1987): Public Statistics and Democratic Politics. In *The Politics of Numbers*, edited by Wm. Alonso and P. Starr, Russell Sage Foundation, New York, pp. 261–274.
- Research Triangle Institute (1989): *National Household Seroprevalence Survey: Pilot Study Report*. Report submitted to the National Center for Health Statistics.
- Stigler, S.M. (1988): The Centenary of JASA. *Journal of the American Statistical Association*, 83, pp. 583–587.
- Stoto, M.A. (1989): Statistics for the AIDS Epidemic. *Chance*, 2 (2), p. 9 and p. 52.
- Turner, C.F., Miller, H.G., and Moses, L.E. (eds.) (1989): *AIDS. Sexual Behavior and Intravenous Drug Use*. National Academy Press, Washington, D.C.
- Walker, F.A. (1890): Statistics of the Colored Race in the United States. Publications [later Journal] of the American Statistical Association, 2, pp. 91–106.
- Wallman, K.K. (1988a): Adjusting the Census: A Tale of Two Cities. *Chance*, 1 (2), pp. 48–52.
- Wallman, K.K. (1988b): A Tale in Two Cities: Act III, Washington, D.C. *Chance*, 1 (3), pp. 55–57.
- Weeks, M.F., Horvitz, D.G., Hurley, P.L., and Wright, R.A. (1989): Designing a Household Survey to Estimate HIV Prevalence: An Interim Report on the Feasibility Study of the National Seroprevalence Survey. Paper presented at the Fifth Conference on Health Research Methods, Keystone, Colorado, May 4, 1989.

Received November 1988
Revised July 1989