

Statistics and Counterterrorism: The Role of Law, Policy, and Ethics

William Seltzer
Fordham University, seltzer@fordham.edu

Abstract

The paper addresses the role of law, policy, and ethics in work on statistics and counterterrorism. It focuses on the implications of these issues for data mining efforts that attempt to identify individuals or population sub-groups as terror threats by using, at least in part, data gathered by a federal statistical agency.

Keywords: data mining, Department of Homeland Security (DHS), federal statistical system, statistical confidentiality, targeting

1. Introduction

This paper addresses the role of law, policy, and ethics in work on statistics and counterterrorism, particularly focusing on issues relevant to the use of data mining in counterterrorism applications. However, many of the principles and issues raised apply broadly to other statistical applications in the field of counterterrorism and, indeed, to other fields as well.

What I have to say is based on several decades of experience of work in and with national and international statistical agencies, particularly in the field of population statistics. It also draws on the research carried out in the past decade on the misuse of population data systems in this country and elsewhere to assist in a range of human rights abuses by targeting vulnerable population sub-groups and individuals (see, for example, Anderson and Seltzer [2004]; Seltzer [1998]; Seltzer and Anderson, [2000; 2003]). This paper is also informed by my experiences over the past five years as the Chair of the American Statistical Association's (ASA) Committee on Professional Ethics, although I hasten to add that the views I express are my own and not those of the ASA or its Committee on Professional Ethics.

2. Data Mining

Like most statistical methodologies data mining by itself is ethically and legally neutral. This is particularly so because the term data mining is a generic one referring to a wide range of procedures, involving diverse data sets, and carried out for

numerous purposes. For example, in the context of counterterrorism, data mining might be used to search out bioterrorist disease threats, to locate facilities likely to be the source of explosives of one kind or another, or to identify persons – individuals or groups of people – seen to pose threats as real or potential terrorists. In this paper I will confine my attention to data mining efforts concerned with the last category, that is, those that attempt to identify individuals or population sub-groups as terror threats because the issues that arise in this type of application are the most clear-cut.

As already noted, data mining is an umbrella concept. I will focus on two types of applications:¹ First, where the goal is to either produce detailed tabular descriptions of a population based on two or more data sets that, taken one at a time, provide a less detailed description or an equivalent approach involving multivariate analysis. Here, whether the approach is tabular or multivariate, the statistical statements made are about the population under investigation. The second type of data mining application is exemplified by efforts to construct or impute characteristics for individuals found in one data set based on values for the same individual or similar individuals found in one or more other data sets. Here the goal is to make statements about an individual, although in some applications that may be merely an intermediary step in a more traditional statistical analysis.

Both kinds of data mining raise a number of ethical, legal, and policy concerns. These concerns cluster around three different sets of issues. First, those related to the validity and appropriateness (both in statistical and substantive terms) of the statistical methods employed. Second, those related to the use of micro data pertaining to identifiable individuals or mesodata (that is, data for small geographic entities) used to target vulnerable population sub-groups. Third, those that give rise to possible disservices to the statistics profession or to statistical agencies due to public backlash over actual or perceived shortcomings of specific data mining applications or other statistical techniques used in counterterrorism work. In the balance of this paper, these three sets of issues are referred to respectively, as validity issues, statistical confidentiality issues, and responsibility issues, recognizing that the terms validity, statistical

confidentiality and responsibility are incomplete descriptors of all the issues to which they refer.

Ethical guidance is available on all three topics – validity, statistical confidentiality, and responsibility. Policy and legal guidance is also available on many aspects of statistical confidentiality, particularly with respect to data obtained from major federal statistics data sets, such as the decennial census, and policy guidance is also available on issues related to validity.

The primary source for guidance on ethics comes from the ASA's *Ethical Guideline for Statistical Practice* [1999], adopted by the ASA Board in 1999 and available on line at the ASA's website (www.amstat.org) and in print from the ASA office. The main sources for guidance on policy and law are the policies developed under the leadership of the Office of Statistical Policy in OMB and the various statutory protections provided in the US Code for statistical information collected by various federal statistical agencies.

3. Validity

Several provisions of the ASA's ethics guidelines address issues of validity. They include in section II.A:

2. Guard against the possibility that a predisposition by investigators or data providers might predetermine the analytic result. Employ data selection or sampling methods and analytic approaches that are designed to assure valid analyses in either frequentist or Bayesian approaches.
4. Assure that adequate statistical and subject-matter expertise are both applied to any planned study. If this criterion is not met initially, it is important to add the missing expertise before completing the study design.
5. Use only statistical methodologies suitable to the data and to obtaining valid results. For example, address the multiple potentially confounding factors in observational studies, and use due caution in drawing causal inferences.
7. The fact that a procedure is automated does not ensure its correctness or appropriateness; it is also necessary to understand the theory, the data, and the methods used in each statistical study. This

goal is served best when a competent statistical practitioner is included early in the research design, preferably in the planning stage.”

And in section II.C:

2. Report statistical and substantive assumptions made in the study.
5. Account for all data considered in a study and explain the sample(s) actually used.
6. Report the sources and assessed adequacy of the data.
7. Report the data cleaning and screening procedures used, including any imputation.
8. Clearly and fully report the steps taken to guard validity. Address the suitability of the analytic methods and their inherent assumptions relative to the circumstances of the specific study. Identify the computer routines used to implement the analytic methods.
9. Where appropriate, address potential confounding variables not included in the study.
12. Report the limits of statistical inference of the study and possible sources of error. For example, disclose any significant failure to follow through fully on an agreed sampling or analytic plan and explain any resulting adverse consequences.

In addition, from a policy perspective, many federal statistical programs have developed written statistical standards describing in considerable detail the data collection, data editing and cleaning, estimation, inferential and dissemination procedures they employ (see, for example, the most recent statistical standards adopted by the National Center for Education Statistics [2002] and the Bureau of Transportation Statistics [2005]). These standards are subject to independent scrutiny outside the concerned statistical unit and agency and are considered to be applicable, not only to those working in the concerned federal statistical program, but also to subcontractors and other working for and with the agency.

The question immediately arises, do the Department of Homeland Security and the Department of Defense have comparable written standards for their work on

statistics and counterterrorism in general and for their work on data mining in particular? If so, are they widely known and adhered to by those doing such work, including subcontractors?

Certainly issues of validity are highly relevant when considering the methods used to combine information from different data sets in data mining applications, given that each data set is subject to different sorts of coverage and content errors, many with different time references, and some subject to sampling variability. These issues of validity become even more acute when data mining serves as the basis for inferences about those to be detained, criminally charged or deported from the country or policy decisions about population sub-groups defined along religious, racial, ethnic, ancestry, or linguistic lines.

4. Statistical Confidentiality

The ASA ethics guidelines deal with statistical confidentiality in section II.D, "Responsibilities to Research Subjects (including census or survey respondents and persons and organizations supplying data from administrative records, as well as subjects of physically or psychologically invasive research)." Among the pertinent provisions are

1. Know about and adhere to appropriate rules for the protection of human subjects, including particularly vulnerable or other special populations who may be subject to special risks or who may not be fully able to protect their own interests. Assure adequate planning to support the practical value of the research, the validity of expected results, the ability to provide the protection promised, and consideration of all other ethical issues involved. Some pertinent guidance is provided in key references 3 - 7 at the end of this document for U.S. law, the U.N. Statistical Commission, and the International Statistical Institute. Laws of other countries and their subdivisions and ethical principles of other professional organizations may provide other guidance.

3. Avoid excessive risk to research subjects and excessive imposition on their time and privacy.

4. Protect the privacy and confidentiality of research subjects and data concerning them, whether obtained directly from the subjects, from other persons, or from administrative

records. Anticipate secondary and indirect uses of the data when obtaining approvals from research subjects; obtain approvals appropriate for peer review and for independent replication of analyses.

The roots of the concept of statistical confidentiality and the protection of harm from arising by cooperating with statistical inquiries go back to the Hippocratic oath where physicians agree not to cause harm to their patients and not to gossip about information obtained in the course of their professional work. The modern concept of statistical confidentiality in the United States evolved in the 1890s as a means of encouraging businesses to report accurately by assuring them that business rivals, muckraking journalists and populist members of Congress would not have access to the information they provided, except as statistical aggregates.

These assurances were first extended to population data in the proclamation issued by President Taft in connection with the 1910 Decennial Census. That proclamation included the following language:

The census has nothing to do with taxation, with army or jury service...or with the enforcement of any national, State, or local law or ordinance, nor can any person be harmed in any way by furnishing the information required [Barabba, 1975: 27; quoted in full in Bohme and Pemberton, 1991: 6].

Effective legal protections against disclosure were first introduced in laws relating to data collected by the US Census Bureau concerning businesses. By the time of the 1920 and 1930 Censuses effective protections were extended to cover population data obtained by the Census Bureau as well. Subsequently, with a few roll-backs [Bohme, and Pemberton, 1991; Seltzer, and Anderson, 2003; Anderson, and Seltzer, 2004], these legal protections have been extended to cover virtually all data collected for statistical purposes by federal statistical agencies [Office of Management and Budget, 1997; Wallman, 2003].

5. Responsibility

It is important to recognize the full implications for the statistics profession and federal statistical agencies of possible public backlash arising from inappropriate or irresponsible applications of data mining and other statistical procedures in counterterrorism efforts. In recent years more and more research is being

undertaken documenting the role of population data systems in assisting in human rights abuses by targeting vulnerable population subgroups and individuals. These abuses, particularly the proactive assistance provided by the US Census Bureau to the US Army in the actual round-up of Japanese Americans on the West Coast early in World War II, are being recalled by the public and journalists today.

They were cited, for example, in two such instances of backlash in 2004. In the first case, early in 2004, the *Washington Times* (a conservative newspaper) ran a story headlined, "Study used census information for terror profile" [Washington Times, January 19, 2004]. The substance of the story was about a NASA study aimed at profiling airline passengers as terrorist risks and more centrally involved the use of data provided by airlines. However, this initial story concentrated on the apparent use of data from the Census Bureau, stating early in the story,

The NASA study highlights concerns among civil-liberties advocates that the government is gathering private information and even using its own data -- contrary to repeated official assurances from the Census Bureau - - to develop a data mining system to prescreen all airline passengers.

The story went on to quote from presumed experts in the field in these terms,

Bill Scannell, president of the group DontSpyOnUs.com, called the inclusion of census information "absolutely appalling."

"Information given by American citizens for reasonable demographics information has been turned around and used to spy on people. This sounds like East Berlin, circa '74," said Mr. Scannell, a privacy advocate.

"There is a certain amount of fumbling around going on," said Barry Steinhardt, director of the American Civil Liberties Union's technology and liberty program ... However, Mr. Steinhardt, who sits on the Census Advisory Committee, said releasing information on households and individuals is "a major breach of trust."

"The advisory board specifically asked this question, whether they were providing data to any other government agency, and the answer was 'no,'" Mr. Steinhardt said. "We

will have to look carefully at what they provided NASA and why."

In fact, in that case, the census data used came from the 1990 public use sample and was used to test some of the features of the modeling work done in the NASA study and not in the profiling model directly. Rather quickly, the backlash turned from the Census Bureau to the airlines which provided the passenger data and the Department of Defense officials behind the passenger profiling effort.

In a more recent, and ongoing example, the Census Bureau has remained at the center of the controversy, with possible implications for the content and quality of the 2010 Census. The more recent example, where the Bureau's experience in assisting in the internment of the Japanese Americans was explicitly recalled, involved the provision of tabulations from the 2000 Census to the Department of Homeland Security (DHS) showing, for each 5-digit zip code in the country, the number of Arab Americans by detailed ancestry. The subject became a matter of public knowledge and controversy in late July 2004 when the *New York Times* [July 29, 2004, p. 19] ran a story headlined, "Homeland Security Given Data on Arab-Americans" based on the results of a Freedom of Information Act (FOIA) request made to the Census Bureau by the Electronic Privacy Information Center in May.

According to the *Times* story,

The tabulations were produced in August 2002 and December 2003 in response to requests from what is now the Customs and Border Protection division of the Department of Homeland Security. One set listed cities with more than 1,000 Arab-Americans. The second, far more detailed, provided ZIP-code-level breakdowns of Arab-American populations, sorted by country of origin. The categories provided were Egyptian, Iraqi, Jordanian, Lebanese, Moroccan, Palestinian, Syrian and two general categories, "Arab/Arabic" and "Other Arab."

The story included the defense of the Census Bureau that cooperation with other federal agencies was standard practice. Indeed, it was required. Although recognizing that there may be legitimate concerns about the intended use of these tabulations, a Bureau spokesman noted that "we have not been given the authority to determine which organization gets which information." The Bureau also sought to separate any

discussion of the Bureau's mandate to share information from controversies about the actual data being shared, characterizing the latter as "a societal debate, not a census debate."

Even in the initial story a number of questions were raised about the appropriateness of the actions of the DHS and the Census Bureau. For example, according to former Census Bureau Director Kenneth Prewitt,

Given the bureau's history, consideration of requests from law enforcement agencies requires more than strict parsing of legalities. The Census Bureau has a longstanding practice of being unusually cautious about such cooperation because it is difficult to explain to the public. There is an issue of principle involved as well as law.

Directly referring to the Bureau's involvement in the round up of the Japanese American population on the West Coast in the months after Pearl Harbor, Prewitt stated, "In World War II we violated our principles even if we didn't violate the law, and we assured people we wouldn't do it again."

Along the same lines, James Zogby, president of the Arab American Institute was quoted to the effect that

The data sharing was particularly harmful at a time when the Census Bureau is struggling to build trust within Arab-American communities. As this gets out, any effort to encourage people to full compliance with the census is down the tubes. How can you get people to comply when they believe that by complying they put at risk their personal and family security?

Samia El-Badry, a demographer and an Arab-American member of the Census Bureau's decennial census advisory committee, was quoted as stating "The real question is to Homeland Security. What are they hiding? Why do they need this?"

The Customs and Boarder Protection (CBP) division of DHS offered several explanations for their interest in these data. According to the emails released as a result of the FOIA request and the July *New York Times* story that quoted a CBP spokeswoman, the CBP indicated that the requests were made "to help the agency identify in which airports to post signs and pamphlets in Arabic."

After matter became a source of increasing controversy, senior officials of DHS and the CBP met

with representatives of the Arab American Institute and the American-Arab Anti-Discrimination Committee on August 13th. Several days later the Arab American Institute issued a press release that quoted a DHS statement that said that

The data "arose from two separate" and informal requests. The first request "concerned several languages, not just Arabic" and the second request "was ...based on a specific list of countries of concern designated by the U.S. Department of State" that went beyond the Arab world. The DHS also states that "neither request - and no request at any time - asked for identification of Arabic speaking people or Arabic ancestry by zip code," and that CBP "did not release any of this information to any other agency and did not, in fact, use, forward, or maintain the information and the information was deleted." [Arab American Institute, press release, 8/17/2004]

By contrast, an August 13 follow up *New York Times* story, quoted CBP Commissioner Bonner as stating that "the data request was made to help the agency determine in which airports to post signs and pamphlets in Arabic and said that similar data was sought on other ethnic groups." [NY Times, 8/13/2004, "Coalition Seeks Action on Shared Data on Arab-Americans," p. A11].

The main point of this follow up story was to describe the actions of "a coalition of ethnic advocacy groups, privacy watchdogs and civil rights and civil liberties organizations" to demand a further response from both the Census Bureau and the DHS. The story related that

A letter of complaint, drafted by the Arab American Institute Foundation and signed by more than 50 organizations and people, was sent to the Census Bureau on Thursday. The statement questioned the bureau's "judgment and discretion" in cooperating with domestic security officials and called for the bureau to announce a "plan of action" to address public concerns.²

It also quoted several statements made by participants. For example, in the words of Helen Hatab Samhan, executive director of the Arab American Institute Foundation, "Based on the number of organizations who have joined us on this, there is a serious sense of betrayal from many communities." While Wade Henderson, executive director of the Leadership Conference on Civil Rights stated, "The Census

Bureau has unwittingly played into the worst fears of all minority communities that they are being watched, cataloged and tracked for improper purposes.”

In addition to these and two subsequent stories in the *New York Times*, a number other news outlets around the country also covered the controversy. It was also the subject of considerable informal discussion at the Joint Statistical Meetings held in mid-August 2004, with current Census Director Louis Kincannon finding it necessary to repeatedly defend the Census Bureau’s position.

By the end of August 2004 both the Census Bureau and the CBP had announced changes in policies and procedures in an effort to respond to the widely expressed concerns (see, for example, US Census Bureau, Press Release dated 8-30-2004, “U.S. Census Bureau Announces Policy Regarding Sensitive Data,” and *New York Times*, 8/31/2004, p. 14 “Census Policy on Providing Sensitive Data Is Revised”).

Nevertheless, as the *NY Times* August 31 story and those appearing elsewhere indicated the matter was not closed at least far as a number of ethnic group representatives were concerned. While welcoming the new Census Bureau policy many asked about how the new policy would be carried out. For example, Samia El-Badry asked: “So now requests go to an assistant director, but what happens then? What criteria will be used to evaluate requests?” She also indicated that she, and other members of the decennial census advisory committee, would take the matter up again at the fall meeting of the committee.

Subsequently, the matter did become an important focus of a November 2004 census advisory meeting. According to a *New York Times* story about this event (“Panel Says Census Move on Arab-Americans Recalls World War II Internments”, 11/10/2004, p. 19), “representatives of Asians, Hispanics, blacks, American Indians and Native Alaskans each objected to the [Census Bureau’s] action.” The story once again stressed that the data released were publicly available on the Bureau’s internet site, but also quoted Census Director Kincannon’s acknowledgement that the release “affected the perception of the Census Bureau ... and that is a very important problem for us.”

However, missing from the public discussion was a recognition of the significance of one of the products identified in the Bureau’s response to the FOIA request: a CD disc containing detailed ancestry data in a standard format for each of the 32,038 5-digit zip codes in the United States. Given the ubiquity of postal zip codes as a geographic identifier in numerous

other data bases, the disc provided by the Bureau to DHS would be of obvious utility in any effort to link these data bases and draw inferences about specific Arab Americans populations collectively and individually through data mining techniques.

Unfortunately, almost any conclusions based on such a data mining exercise would almost certainly be highly misleading. In contrast with the many carefully-written definitions, explanations, and caveats that accompany both Census Bureau publications and the outputs available on the Bureau’s website, the material provided by the Bureau to DHS appears to have been stripped clean of such meta-data and the normal caveats they contain. Thus, anyone using the tabulations provided to DHS would not be warned that (a) the zip code tabulation areas used by the Census Bureau did not necessarily correspond to actual postal zip codes, (b) the results presented were based on tabulations of persons reporting an Arab ancestry either as a primary or secondary ancestry, (c) the counts of Arab Americans were based on sample data from the census, (d) the Census Bureau had applied statistical procedures, including data swapping of sampled households among neighboring geographic areas, to help maintain confidentiality, and (e) the census results were subject to sampling error and nonsampling error. (For example, many persons did not respond to the ancestry question in the 2000 Census or simply answered “American.”) The impact of several of these sources of error is compounded by the very small number of persons reporting any Arab ancestry that one would expect to find in each 5 digit zip code tabulation area (that is, about 38 persons).³

Accordingly, whether the goal of the DHS in obtaining this 2000 Census set of meso data was to estimate the need for airport signage at the 5-digit zip code level or, more plausibly, to use data mining to identify presumed terrorists or presumed terrorist threats, it would seem that both the Arab American community and DHS were ill-served by the Bureau’s actions in providing this output. Of course, the Census Bureau itself was ill-served as well. As the Bureau’s Director acknowledged, “We recognize that simply making sure we obey the law may not always be enough to ensure that people trust us ... Perception also affects how people view and cooperate with the census” (*New York Times*, 8/31/2004, p. 14).

6. Discussion and Conclusions

Clearly a first, important step for sound work, (that is, valid, ethical, and responsible work) on statistics and counterterrorism is the development of written

statistical standards along the lines used in other federal statistical programs. (These standards were briefly described in section III above.) Such standards are important not only in promoting sound work by having agreed on protocols for reaching conclusions where quantitative data play a major role, but also by providing an opportunity for both statisticians and others to agree on such protocols independently from an actual crisis-laid situation and the tensions and emotions that decisions in such circumstances frequently evoke.

Such standards do not mean that mistakes won't be made either because the standards are ignored or that the standards adopted were defective in some essential respect. (For example, they were insufficiently specific or structured on incomplete or faulty assumptions.) One way of improving the quality of such standards is to provide as wide as possible scrutiny of these standards at various stages in their development. Despite their possible shortcomings, such agreed on written standards provide some authority that quantitative analysts can cite in describing the basis for their conclusions in the face of those holding alternative views. It is also one way of promoting the use of the best statistical methods in work on counterterrorism.

Second, those engaged in data mining and other statistical work related to counterterrorism need to be aware of the legal, policy, and ethical constraints involved in different kinds of uses of major federal statistical data sets, including public perceptions and concerns. For example, the use of vital statistics records and related health reporting systems to track disease outbreaks in an effort to provide an early warning of a bioterrorist attack is fully consistent with the original public health purposes of these data systems. On the other hand, any effort to turn the population census from a statistical tool to an investigative instrument for targeting individuals or vulnerable population subgroups living in specific places for investigation or more egregious harms (such as detention or deportation) is likely to cause considerable public outcry. In addition, it is likely to substantially reduce the government's ability to collect reliable data on many topics in the future.

Even from this brief review, it is apparent that there is considerable overlap between good science and ethics and between law and ethics. But it is also important to recognize that ethics can go beyond both science and law. Neither science nor legality alone, or in combination with one another, can be used to justify unethical behavior. This truism was reflected in the actions of the court in the trial of those charged with

medical experiments, and the related collection of information and anthropological materials, in the so-called doctors' trial at Nuremberg after World War II. In that trial defendants unsuccessfully attempted to justify their actions on the grounds that (a) their actions were legal under current national law and (b) they were engaged in scientific work for an important and beneficent purpose. Subsequently, many national and international ethical statements, research policies and regulations as well as related laws have adopted the same perspective as the judges at Nuremberg.

Indeed, when legal and technical safeguards fail, ethics often becomes our final safeguard against many kinds of shortcomings and more serious abuses. In these circumstances, one would strongly urge all those engaged in work on statistics and counterterrorism to carry out ethical reviews of their work. Such reviews to be effective must also involve at least some persons with no intellectual, career, institutional, or personal stake in any specific outcome of the review. In this regard, many of the major past ethical failures arose in situations where none of those involved recognized that what they were doing posed an ethical problem. Accordingly, ethical reviews and discussions of ethics are one of the best preventive measures to take against serious ethical problems.

What is the bottom line of this paper. Don't outlaw data mining in the fight against terrorism or for other uses. However, in applying data mining techniques, we must be aware of, and deal with, the serious validity and statistical confidentiality issues that arise. Some approaches for doing so have been suggested.

Finally as the ASA ethics guidelines state,

All statistical practitioners are obliged to conduct their professional activities with responsible attention to: 1. The social value of their work and the consequences of how well or poorly it is performed. This includes respect for the life, liberty, dignity, and property of other people. [section I.C]

To do any thing less is to fail to act responsibly.

References

- American Statistical Association. (1999) *Ethical Guidelines for Statistical Practice*. Alexandria (VA): American Statistical Association.

- Anderson, Margo and William Seltzer. (2004) "The Challenges of Taxation, Investigation, and Regulation: Statistical Confidentiality and U.S. Federal Statistics, 1910-1965." Paper prepared for U.S. Census Bureau Symposium, Woodrow Wilson International Center for Scholars, March 4-5, 2004.
- Arab American Institute. (2004) "DHS Responds To Arab American Concerns Over Census Data." Press release dated August 17, 2004. Accessed from <http://www.aaiusa.org/pr/release08-17-04.htm> on 11-23-04.
- Banks, David. (2004) "Feasibility Considerations in Three Kinds of Data Mining." Paper presented at the NISS-ASA Workshop on Statistics and Counterterrorism, New York, November 20, 2004.
- Barabba, Vincent. (1975) "The Right of Privacy and the Need to Know," in U.S. Census Bureau, *The Census Bureau: A Numerator and Denominator for Measuring Change*, Technical Paper 37. Washington, D.C.: Government Printing Office.
- Bohme, Frederick G. and David N. Pemberton. (1991) "Privacy and Confidentiality in the US Censuses -- A History." Paper presented at the annual meeting of the American Statistical Association. Atlanta, GA (August 18-22, 1991).
- Bureau of Transportation. Statistics. (2005) "BTS Statistical Standards Manual, October 2005." Available at http://www.bts.gov/programs/statistical_policy_and_research/bts_statistical_standards_manual/
- National Center for Education Statistics. (2002) "2002 NCES Statistical Standards." Available from <http://nces.ed.gov/statprog/2002/stdtoc.asp>
- Office of Management and Budget. (1997) "Order Providing for the Confidentiality of Statistical Information." *Federal Register* 62:124 (June 27) 35044-35050, accessed 5/30/02 from www.whitehouse.gov/OMB/inforeg/statpol.html.
- Seltzer, William. (1998) "Population Statistics, the Holocaust, and the Nuremberg Trials." *Population and Development Review*, 24 (3, September): 511-552.
- Seltzer, William and Margo Anderson. (2000) "After Pearl Harbor: The Proper Role of Population Data Systems in Time of War." Paper presented at the annual meeting of the Population Association of America, March 23-25, 2000, Los Angeles, CA. Available at <http://www.uwm.edu/~margo/govstat/integrity.htm>
- _____. (2003) "Government Statistics and Individual Safety: Revisiting the Historical Record of Disclosure, Harm, and Risk." Paper prepared for presentation at a workshop, Access to Research Data: Assessing Risks and Opportunities, organized by the Panel on Confidential Data Access for Research Purposes, Committee on National Statistics (CNSTAT), The National Academies, Washington, DC, October 16-17, 2003. Available at http://www7.nationalacademies.org/cnstat/Seltzer_Anderson.pdf
- Wallman, Katherine K. (2003) "Data Access and Confidentiality – the Changing Legal Landscape." Discussion prepared for presentation at a workshop, Access to Research Data: Assessing Risks and Opportunities, organized by the Panel on Confidential Data Access for Research Purposes, Committee on National Statistics (CNSTAT), The National Academies, Washington, DC, October 16-17, 2003. Available at http://www7.nationalacademies.org/cnstat/Katherine_Wallman.pdf

¹ There are several ways that this umbrella concept may be sub-divided. In this paper I focus on the statistical objective of the procedure. By contrast, Banks [2004] sub-divides the term in terms of the investigative purpose of the application.

² The text of the letter and full list of signatories is available from the Arab American Institute at http://www.aaiusa.org/census_letter.htm

³ There were a total of just over 1.2 million persons reported with an Arab ancestry in the 2000 Census in the 50 states, the District of Columbia, and Puerto Rico, the geographic scope of the data provided to DHHS. There were a few more than 32,000 5 digit zip code tabulation areas in the corresponding area.