

UNITED STATES DISTRICT COURT  
FOR THE DISTRICT OF COLUMBIA

SUSAN B. LONG and )  
DAVID BURNHAM )  
Plaintiffs, )  
 )  
v. ) Civil Action No. 1:00CV00211 PLF  
 )  
DEPARTMENT OF JUSTICE )  
Defendant. )  
\_\_\_\_\_ )

**DECLARATION OF MICHAEL D. MALTZ**

I, Michael D. Maltz, declare as follows:

1. My name is Michael D. Maltz. I am an emeritus professor in the Department of Criminal Justice and the Department of Information and Decision Sciences at the University of Illinois at Chicago (UIC). Over the past 35 years I have done, and continue to do, extensive work in quantitative analysis of law enforcement data. I began directing studies of the Boston Police Department in 1967, and served as a program manager for a number of studies of police operations and criminal activity at the National Institute of Law Enforcement and Criminal Justice (now the National Institute of Justice) from 1969-1972. I have been at UIC since 1972. From 1985-1989, I served as the Director of the UIC Center for Research in Law and Justice. I was editor of the Journal of Quantitative Criminology from 1996 to 2000, and I have been a consulting editor for the Journal of Research in Crime and Delinquency since 1976. From 1995 to 2000, I was a Visiting Fellow for Research and Policy Studies at the Bureau of Justice Statistics (BJS) of the Department of Justice. In addition, from 1995 to 1998, I was a member of the Department of Justice's Working Group on Criminal Justice Statistics. A copy of my Curriculum Vitae, which sets forth a more detailed description of my professional experience, is

attached.

2. My professional research has focused on making valid and useful inferences from law enforcement data and studying the significance of missing data. In this work, I am concerned with ensuring that the inferences are not attributable to biases in the data used, nor to the way the data were collected, nor to the methods used to analyze them. I have conducted in-depth analyses of the way crime data are collected and aggregated. I have also sought to identify the ways in which missing data or other problems with data collection may render certain data unreliable or useless in trying to reach particular conclusions.

3. An example of my experience in analyzing law enforcement data is the research that I did while at BJS. During my tenure there, I prepared a report on the history of the FBI's Uniform Crime Reporting (UCR) system and the problems that arise in interpreting the reports of crime and arrests in this system. The UCR is essentially a voluntary program, in which crime statistics are submitted by police agencies to the FBI. Because of this, the FBI has no control over the reliability, accuracy, consistency, timeliness, or completeness of the data in the UCR. This lack of control by the FBI over the data means that there are many gaps that need to be filled. My report analyzed the procedures used by the FBI to fill in gaps in the data and provided suggestions about how these procedures might be improved. I am currently working with the FBI and the American Statistical Association's Committee on Law and Justice Statistics on improving the FBI's Uniform Crime Reporting system; in addition, I am currently working with the Illinois State Police on improving the data they send to the FBI.

4. I have published books, book chapters and journal articles on how to analyze criminal records and law enforcement statistics, and how to use this data to evaluate the effectiveness of crime control efforts. A list of publications appears in the attached Curriculum

Vitae.

5. I am generally familiar with the data on cases handled by the United States Attorneys' Offices that is collected by the Executive Office for the United States Attorneys (EOUSA) and used by the Department of Justice to produce various statistical reports such as the Attorney General's Annual Report, United States Attorneys' Annual Statistical Reports, and management reports for use within the Department. During the early 1990s, I performed research with case management data compiled by the United States Attorneys' Offices as part of a project in which I evaluated a Department of Justice analysis of Assistant United States Attorneys' workloads that had been prepared for the Attorney General. I am also familiar with modified versions of the EOUSA data on criminal matters that are made available by the Department's Bureau of Justice Statistics (BJS) as part of the Federal Justice Statistics Program. See <<http://www.ojp.usdoj.gov/bjs/cd.htm#Federal>>.

6. I have reviewed the statements in the Declaration of Teresa Davis concerning the Department's justification for withholding "Criminal Lead Charge" information, "Internal Administrative Filing Numbers," and court numbers in the case management database files. I have also reviewed the statements in the Declaration of Kenneth L. Wainstein on the Department's reasons for withholding information in the "Criminal Lead Charge" field.

#### Public Interest in Disclosure of Tracking Numbers

7. In my opinion there is a strong public interest in the disclosure of the investigative agency numbers and program agency file numbers in the case management database files. As described in the Department's declarations, these fields contain numbers used by government agencies involved in a case to identify the case in that agency's tracking system. Below are several examples of why the public would benefit from the disclosure of agency and court

numbers.

8. Different databases may contain reports on the same activities but may not agree for a number of reasons: (1) data may be missing from one or more of the databases, (2) data may not have been properly maintained by one or more of the entities responsible for the data, or (3) different entities may use different procedures in compiling data. For example, if the Drug Enforcement Administration maintains a database of cases that include cases that it has referred to the United States Attorneys' Office, researchers may need to reconcile reports from the Drug Enforcement Administration with reports from the database maintained by the United States Attorneys' Offices. The agency/program number should permit a researcher to cross-reference the entries in the two databases to trace the source of discrepancies in the records maintained by the separate entities. Without the agency/program numbers to identify which cases in the referring agency (e.g., the Drug Enforcement Administration) database correspond to entries in the database of the agency to which the case was referred (the United States Attorney's Office), it will be extremely difficult or impossible for researchers to verify the accuracy of the data and reconcile differences.

9. As another example, suppose it was discovered that ten percent of all cases referred by the Drug Enforcement Administration to the United States Attorney's Office were dropped for insufficient evidence. Further suppose that most of these cases were of a specific type (or from a specific unit). Only by having a means of tying one database to the other would it be possible to determine if specific types of cases were disproportionately defective, and get some indication of the agencies' conduct.

11. If tracking identifiers such as the agency/program numbers are not released to researchers who are working with data like the EOUSA case management data, any statistics that

are generated are highly suspect in other ways as well. For example, in one of my studies I reviewed annual data on county homicides in Illinois in which the number of clearances of homicide cases exceeded the number of homicide cases! It turned out that the reason was that some of the homicides that were cleared had occurred in earlier years. Without a means of tracking cases, it is difficult or impossible to avoid similar situations that would provide misleading information. It is for reasons like this that the FBI has been changing its national crime reporting system to include tracking numbers, so that when an agency reports a homicide in its 2001 statistics that is solved in 2002, the disposition of that particular case can be revised.

12. An example of the public interest in disclosure of information that permits such comparisons is provided by the Federal Justice Statistics Program of the Bureau of Justice Statistics. There are several different federal governmental agencies that collect similar data describing criminal case processing events: Executive Office for the United States Attorneys, the Administrative Office of the United States Courts, Federal Bureau of Prisons, and the United States Sentencing Commission. However, the statistics published or reported by these agencies are not directly comparable because the databases used by these agencies use different criteria to tabulate and report on criminal cases. For example, in FY96 the Executive Office for the United States Attorney reported that 58,141 criminal cases were filed, while the Administrative Office for the United States Courts reported that 67,700 were filed. To account for these differences, the Bureau of Justice Statistics implemented the Federal Justice Statistics Program to provide uniform case processing statistics across different stages of the Federal criminal justice system and to track individual defendants from one stage of the process to another. See *Comparing Case Processing Statistics*, available at, <<http://www.ojp.usdoj.gov/bjs/abstract/ccps2.htm>>. The Program produces a joint statement by the federal criminal justice agencies that identifies

the primary differences in case processing statistics reported by the agencies and explains why reported statistics are not directly comparable across agencies. *Id.*; see also *Reconciling Federal Criminal Case Processing Statistics*, Federal Justice Statistics Program Methodology Report (Sept. 1999), *available at* <<http://www.ojp.usdoj.gov/bjs/abstract/rfccps.htm>>. Researchers seeking to perform similar cross-database comparisons using the EOUSA case management databases and data reported by the referring agencies or programs will need access to the agency/program numbers.

13. For the same reason, I believe that there is a public interest in the disclosure of court numbers and file names in the case management files. Disclosure of the court numbers would allow the information in the case management files to be linked to other databases that use court numbers to identify a case, most notably the databases maintained the Administrative Office for the United States Courts. Allowing researchers to identify instances in which records in the EOUSA case management database correspond to a case in the AOUSC database would serve the public interest because researchers will be able to track the case through the entire process from referral to the United States Attorney to final disposition by the courts, and will be better able to identify errors and the causes of discrepancies in the databases. Disclosure of file names permits researchers to verify the accuracy of matches obtained using the court numbers, or identify matching records where the court number is inadequate to match records because of errors or omissions in one of the data sets.

14. The Bureau of Justice Statistics has published the report, "Linking Uniform Crime Reporting Data to Other Datasets," <<http://www.ojp.usdoj.gov/bjs/pub/pdf/lucrdod.pdf>>. This report describes BJS efforts to facilitate linking crime data reported by local authorities with other databases and generally discusses why it is important to link trends in crime to

demographic, administrative, and other data. In the report, the BJS recognizes that providing a single data set that represents but one step in a process is inadequate as a means of understanding the process and, therefore, each individual data set needs to be put in context with links to other relevant data sets. This need is just as important at the federal level as it is at the state level; without an understanding of the entire process, statistics about a single step in a case have little value.

#### DOJ Claims Concerning Lead Charge Disclosure

15. I am not aware of any instance in which a law enforcement investigation has been compromised because the target of an investigation became aware that he or she was being investigated due to the disclosure of reports that do not contain information that identifies an individual person (such as name, address, etc.), but do disclose information like the lead charge information in the EOUSA case management data. Data sets concerning arrests and criminal investigations frequently contain information concerning the charges at issue. For example, the data that the Department of Justice makes available to researchers through the Bureau of Justice Statistics contains information on the suspected offenses that prompted an investigation. The information on the offense or charge, however, does not identify the individual under investigation and I have never encountered an instance where law enforcement authorities found that their investigation was compromised because a suspect learned that he or she was under investigation due to the disclosure of information from a statistical database.

16. In my opinion, the Department's hypothesis that suspects could discover that they are the target of an undisclosed investigation from the combination of the criminal lead charge field in the EOUSA case management data and information that identifies the United States Attorneys' Office responsible for the investigation is implausible speculation. The combination

of the lead charge information and the identity of the office conducting an investigation does not provide sufficient information to tip off a suspect because this information is too imprecise to permit such identification.

17. It is not reasonable to expect that suspects will be able to identify themselves as targets if this information is disclosed because the population of individuals who could be the suspect identified in an individual record is too large. The Bureau of the Census and the National Center for Health Statistics have developed confidentiality guidelines for the microdata that they release in order to prevent indirect identification of individual respondents from demographic characteristics or similar information in census and health survey data. Microdata refers to files that contain individual records with the survey responses or other characteristics for a single person, business establishment or other unit. In census data, microdata will contain information such as race, ethnicity, household type, or marital status. The smaller the identifiable geographic region identified in such microdata, the larger the risk of indirect identification. However, statistical research shows that when the size of a geographic region is increased to reduce the risk of indirect identification, the population will eventually reach a point at which a further increase in the size of a region has almost no effect on reducing the risk of identification. The Bureau of the Census and National Center for Health statistics have concluded that the geographic codes included in data sets released to the public should not identify areas with less than 100,000 persons. Federal Committee on Statistical Methodology, Office of Management and Budget, *Report on Statistical Disclosure Limitation Methodology* (Statistical Policy Working Paper 22) at p. 29, 32, 38 (May 1994), *available at*, <http://www.fcsfm.gov/working-papers/wp22.html>. For some data files with large numbers of variables (such as census surveys that include income data), the Bureau of the Census uses a

cut-off of 250,000. Id. at 29. The Bureau of Justice Statistics, which collects data through the Bureau of the Census, employs the same geographic limitations for its data. Id. at 34.

18. The specific reason for these confidentiality guidelines is to protect persons with unusual characteristics from being identified from the characteristics reported in the data set. That is, the Bureau of the Census has concluded that a population of 100,000 is sufficiently large to prevent persons with unique ethnicity, age, education level, and other individual traits from being identified through disclosure of this information in Census data. This concern parallels the concern expressed by EOUSA; in this case the unusual characteristic is the status of being under investigation for a particular offense. However, because federal United States Attorneys' Offices are responsible for large areas and populations, the potential population that may be the suspect in an investigation listed in the case management database is considerably larger than the threshold of 100,000 that the Census Bureau and National Center for Health Statistics use as their benchmark for geographic limitations. The case management database identifies the United States Attorneys' Office to which the investigation has been assigned. None of the fields listed by the EOUSA as part of the criminal master file disclose the location of the crime or the location of the suspect under investigation by city, state, or any other geographic identifier. Even if one were to assume that the potential population of suspects is comparable to the population of the district to which the investigation is assigned, the population is well above the threshold. The United States Attorneys' Office for the smallest federal district, Wyoming, is responsible for an area with a population of approximately 479,000. The second largest district, the District of Columbia, has a population of over half a million. Over a third of the 94 districts have populations that are over 3 million.

19. In addition, the entries that appear in the lead charge field of the case

management database are too general to reasonably expect that suspects will be able to identify themselves as targets. The lead charge information gives the title and section number of the statute in the United States Code that government officials have designated as the lead charge. These code provisions describe crimes in general terms and a single code section may cover a wider range of conduct. For example, 18 U.S.C. ' 1001 encompasses circumstances where a person “falsifies, conceals, or cover up by any trick, scheme, or device a material fact,” or “makes any materially false, fictitious, or fraudulent statement or representation” or “makes or uses any false writing or document knowing the same to contain any materially false, fictitious, or fraudulent statement or entry.” Moreover, this Section may be invoked “in any matter within the jurisdiction of the executive, legislative, or judicial branch of the Government of the United States.” The fact that this code section has been listed as the lead charge for a particular investigation does not provide details that would allow a suspect to determine that the conduct being investigated is the offense that he or she committed. For example, such an entry does not reveal the subject matter of the false statement, or to whom the false statement was made. Nor does it disclose whether the investigation concerns oral statements, false writings, a scheme to conceal, or any of the other types of conduct described in the statute.

20. To cite another example, if a case management record lists 18 U.S.C. ' 371 as the lead charge, this discloses only that investigators are investigating a matter in which a suspect conspired “either to commit any offense against the United States, or to defraud the United States, or any agency thereof in any manner or for any purpose.” This entry discloses no specifics about the conduct that is being investigated and, consequently, does not provide information that could reasonably be expected to allow a suspect to determine that he or she is the subject of this investigation. The idea that a suspect could determine that he or she is under

investigation because a case management database record states that a particular United States Attorneys' Office has a pending investigation in which 18 U.S.C. ' 371 is listed as the lead charge is fanciful and implausible because of the range of conduct that could trigger such an investigation.

21. The United States Code provisions cited in the Department's declarations as examples of lead charges that should not be disclosed also illustrate this point. The statute concerning a "continuing criminal enterprise," 21 U.S.C. ' 848, may be identified as the lead charge whenever government investigators conclude that they may be able to charge a suspect with violating "any provision of this subchapter or subchapter II of this chapter the punishment for which is a felony," and "such violation is a part of a continuing series of violations of this subchapter or subchapter II of this chapter" undertaken in concert with five or more other persons with respect to which the suspect "occupies a position of organizer, a supervisory position, or any other position of management," and "obtains substantial income or resources." An entry that lists this statute as the lead charge does not disclose details that would allow an individual to determine that he or she is the target of this investigation. The lead charge entry does not disclose which felony statutes prosecutors believe have been violated, what type of position the suspect occupies, or what "income or resources" are being investigated. A case management record that states that a particular United States Attorneys' Office has an ongoing investigation of conduct that may fall under this statute is not sufficient to alert an individual that his or her conduct is being investigated because the lead charge entry contains no details beyond the code section, and an enormous range of activities could lead to an investigation that anticipates charges under this statute.

22. In addition, identification of the lead charge does not provide much information to

a suspect because lead charges are not mutually exclusive. That is, a suspect may belong to more than one “charge category.” A suspect’s conduct in a particular crime may fall under several statutes, any one of which could be listed by investigators in the “lead charge field” when they initiate an investigation. This makes the EOUSA case management data quite different from Census data, where a person belongs to only one race category, only one income category, and only one educational level category. In other words, the information contained in the lead charge is so broad and indeterminate (to the potential suspect) that its use in pinpointing an individual is hardly of any benefit.

23. Even if the population of individuals that could potentially be the suspect identified in a case management database record was small enough to create a risk of identification from lead charge entries, the risk that a suspect could become aware that he or she was the target of an investigation from the lead charge information would only be reasonable for entries that list as lead charge a statute that describes the elements of the crime in very specific terms and does not include alternative descriptions of the conduct that may be charged as a crime under the statute. Moreover, the statute would need to appear in the database, but it must appear rarely. If the statute is so obscure that it is never identified as the lead charge for an investigation, disclosure of the lead charge information obviously cannot interfere with non-existent investigation, and if the statute is listed as the lead charge in several pending investigations, the conduct that it describes is not so rare that a suspect could conclude that he or she must be the target of a particular investigation in which that statute is listed as the lead charge. In addition, one must assume that suspect has not become aware from other sources that an investigation is pending. The Department's declarations do not identify an example where these conditions would exist.

24. The declaration by Kenneth Wainstein appears to implicitly assumes (at paragraphs 5-7) that if a lead charge offense appears relatively rarely in a particular location, this would be sufficient to alert a suspect who committed this offense and allow him or her to take steps to evade or obstruct the investigation. However, as explained above, a lead charge citation is too indeterminate for pinpoint identification and the standards used by the Bureau of the Census for protecting the confidentiality of microdata that does not have large numbers of variables show that such identification is a concern only if the population of individuals who might be identified by the record has been narrowed to at least 100,000. Moreover, in my opinion, a suspect who takes the trouble to find and search the EOUSA database for such an entry is already in a heightened sense of alert. Disclosure of the lead charge information will not give the suspect who examines case management data cause to be more wary than he or she already is. Furthermore, because the lead charge entries in the EOUSA data do not disclose specifics about the investigation, disclosure of the lead charge information will not provide the suspect with an additional means to interfere with an investigation. It is far more likely that suspects are tipped off when acquaintances are interviewed than that they would discover a means to interfere with an investigation by searching for lead charge entries in a database.

#### Methods for Limiting Data To Prevent Individual Identification

25. The general issue raised by the Department of Justice's claim -- namely how can one disclose data in a manner that does not permit someone to indirectly identify the subject of a record -- is an issue that frequently arises in handling data sets. For example, the Bureau of the Census regularly confronts this issue in disclosing data from the decennial Census. If a record identifies an individual who lives in a county with small population, is a member of an obscure Indian tribe, and resides in household with six members, the individual may be identifiable from

this information and releasing a data set with this information would breach the confidentiality of the census survey.

26. EOUSA's remedy for this issue is inconsistent with the approach generally taken by the Bureau of the Census and the Justice Department's Bureau of Justice Statistics to protect confidentiality in these situations and results in withholding far more data than necessary. Simple techniques for preventing unwanted disclosure in these situations include suppressing detailed geographic information. As mentioned above, the Bureau of the Census limits geographic detail by providing that a geographic identifier will not be used unless the area it covers contains at least 100,000 persons. Other simple remedies involve collapsing response categories, rounding, top- and bottom coding. If a particular value or set of values appears infrequently, confidentiality can be maintained by collapsing these values into a larger category or range of values. For example, if occupation information is so detailed that it shows an occupation of United States Senator and the state in which the individual resides, the record will point to one of two people. This infrequent occupation could be collapsed into a broader category to alleviate this problem. An analogous result would be achieved with lead charge information by removing the last digit of the section number for statutes that are rarely identified as lead charges (e.g., fewer than five per year), so that the charge is identified as one among a series of statutes, but the precise statute is not disclosed. If the risk of identification arises because the data reports income to the nearest \$10,000, and a small number of individuals show an income of \$2,000,000, the problem can be alleviated by "top-coding" to make all incomes over \$100,000 indistinguishable. These simple methods are well-established approaches to disclosure protection. See Davis, et al., CONFIDENTIALITY, DISCLOSURE AND DATA ACCESS: THEORY AND PRACTICAL APPLICATIONS FOR STATISTICAL AGENCIES at 5 (2002), *excerpt*

available at <<http://www.census.gov/srd/sdc/index.html>>; Federal Committee on Statistical Methodology, *Report on Statistical Disclosure and Disclosure- Avoidance Techniques* (Statistical Policy Working Paper 2) (1978), available at <<http://www.fcsm.gov/working-papers/sw2.html>>. Similar strategies could be used here to avoid the Department of Justice's remedy of withholding all of the lead charge information, regardless of how frequently the charges appear in the data sets.

Pursuant to 28 U.S.C. ' 1746, I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct to the best of my knowledge.

Executed on \_\_\_\_\_.

\_\_\_\_\_  
Michael D. Maltz