Extracting Confidential Information from Public Documents: The 2000 Department of Justice Report on the Federal Use of the Death Penalty in the United States

David J. Algranati[1] and Joseph B. Kadane[2]

Research is growing on methods to balance two competing public purposes: the need of the public for information about society and government, and the right to privacy and confidentiality of individual information. This article presents a case study in which a governmental agency (the U.S. Department of Justice) released a report (on federal capital cases, 1995–2000) seeking to conceal certain information (the recommendations of the U.S. Attorney and the DOJ Review Committee on whether to seek the death penalty), while releasing cross-tabulations of the concealed information with other, public information about the cases. Careful study of the released cross-tabulations led to full identification of all the variables in 386 out of 682 cases. Using other public data led to full identification of another 160. For the remaining 136 cases, the space of possible values for the missing data is drastically reduced. This article explains how the inferences were made.

Key words: Inadvertent disclosure; missing data; disclosure; linked tables; semi-linked tables.

1. Background and Motivation

1.1. Motivation

Much of the voluminous literature on statistical disclosure control is written as guidance for statisticians at a government agency who want to protect the privacy of detailed respondent information, especially, but not exclusively, in the context of surveys of businesses. (See especially Willenborg and de Waal 1996, 2001; Giessing 2001 and the articles in the JOS special issue edited by Fienberg and Willenborg 1998). Particularly in the context of a survey in which participation is voluntary, such protection is essential to persuade firms to respond. In the context of a mandatory survey the issue remains, as serious political embarrassment could ensue for a statistical agency that allows sensitive private information to become public by failing to protect privacy in what it releases.

The case study reported here is different in several respects. The governmental agency

involved is the U.S. Department of Justice (DOJ), not generally regarded as a statistical agency. The data concerns the actions of public officials, namely the decision whether to seek the death penalty in the prosecution of crimes in which federal law permits the death penalty. In the period under study, the U.S. Attorney in the federal district where the crime took place made a recommendation to a DOJ committee, which in turn made a recommendation to the attorney general, who decided on the position of the government. While the attorney general's decision is public, the DOJ took and takes the position that the recommendations of the U.S. Attorney and the DOJ review committee should not be public, presumably for fear of being embarrassed in court if the attorney general had over-ridden the negative recommendations of the U.S. Attorney and/or review committee. The legitimacy of public officials concealing their official acts from public scrutiny in order to avoid embarrassment is not further considered here.

The death penalty itself is a controversial public issue. The DOJ's purpose in assembling the report studied here was to examine whether race, either of the defendant or of the victim, was an important determinant of the DOJ's decision whether to seek the death penalty. However, the report gives only cross-tabulations without a statistical analysis to address the issue posed. In a separate work (Algranati and Kadane 2003) we have addressed these racial issues using the database whose assembly is the subject of this article.

Our purpose is thus two-fold: to offer a real example of a governmental report which revealed more data than had been intended, and to demonstrate the construction of the database we, and perhaps others, use in further analysis.

1.2.   Background

On September 12, 2000, the United States Department of Justice released a report, *The Federal Death Penalty System: A Statistical Survey* (hereafter ''the 2000 Report''). Several summary statistics emerged from the report, including that over 75% of the decisions to seek the death penalty were directed at minority defendants.

Attorney General Janet Reno stated that she was ''sorely troubled'' by the report and added, ''The survey today finds that minorities are overrepresented in the federal death-penalty system, as both victims and defendants, relative to the general population'' (Brune 2000). Deputy attorney general Eric Holder, who oversaw the creation of the report, stated that he was ''personally and professionally disturbed by the racial disparity'' (Jackson 2000). Holder added that ''I knew there would be a disparity, but I did not expect (one) that large'' (Johnson 2000). In a June 2000 press conference, President Bill Clinton voiced concern about the geographical distribution of those in the federal death penalty system: ''Almost all the convictions are coming out of just a handful of states, which raises the question of whether, even though there is a uniform law across the country, what your prosecution is may turn solely on where you committed the crime.''

While Reno and Holder clearly appeared concerned about the content of the September 12 Report, more scathing criticism came from people outside of the Justice department. NAACP chairperson Julian Bond described the federal death penalty system as, ''the worst sort of racial profiling with the worst result.'' Senator Russell Feingold (D-WI) commented, ''Just as we feared, the same serious flaws in the administration of the death

penalty that have plagued the states also afflict the federal death penalty.'' Attorney David Bruck, who is a member of the federal death penalty resource council, stated that, ''No state in America, not Mississippi, not Texas, not South Carolina, has produced such a racially lopsided death sentencing record as has the federal government'' (Lacey and Bonner 2000).

Following the 2000 Report, Reno called for further study of the federal death penalty process. To some extent, this was addressed by a subsequent Department of Justice report released on June 6, 2001 (hereafter ''the 2001 Report''). The 2001 Report contained some additional data – specifically the racial distribution of 291 ''potential'' capital cases that were not included in the 2000 Report. Mostly the 2001 Report attempted to offer explanations of some of the anomalous features of the 2000 Report data.

Any modeling of this process on the defendant level is obfuscated by the nature of the presentation of data in the 2000 Report. It is aggregated data, presented via 400 pages of tables. However, by using only the information in the 2000 Report, all covariates used in the study could be identified on the defendant level for 386 of the 682 cases. For the 296 cases with missing information, informative constraints can be inferred. Combining this information with outside sources leads to the identification of an additional 160 cases, and tighter constraints on the missing data in the remaining 136 defendants are obtained.

2. The Federal Death Penalty ''Protocol'' and Data in the September 12 Report

Each federal death penalty case originates in a U.S. federal district. There are 94 such districts. Each such district has a U.S. Attorney who supervises the handling of cases in that district. From 1988 to 1994, the U.S. government would seek the death penalty only if the U.S. Attorney handling the case recommended that the death penalty should be sought, and the attorney general concurred.

Beginning in 1995, Attorney General Janet Reno amended the United States attorneys' manual to change the process by which potential federal capital cases were handled. Under this system, the U.S. Attorney handling the case is required to submit his or her recommendation as whether or not the government should seek the death penalty once a federal capital crime is charged. Regardless of the nature of the recommendation, the case is then reviewed (without race information provided, unless it is done so by defense attorneys) by a committee of Department of Justice officials known as the attorney general's review committee on capital cases, which also makes a recommendation to the attorney general.

Regardless of the nature of the earlier recommendations, the attorney general then reviews the case (again, without race information) and makes the official decision as to whether the U.S. government will seek the federal death penalty in the case. The attorney general is free to act regardless of the earlier recommendations.

The 2000 Report contains information on 52 cases from 1988–1994, and 682 cases from 1995–2000. This article focuses on the data from the 682 cases from 1995–2000. The report contains information pertaining to eight variables for these cases:

1. Federal district. Each case must originate in a single federal district. Of the 94 federal districts, 72 submitted federal capital cases during the study period.
2. Defendant race. Race is coded into four mutually exclusive categories: ''White'', ''Black'', ''Hispanic'' and ''Other.'' ''Hispanic'' refers to all defendants of

Hispanic ethnicity, regardless of race. The September 12 Report describes this classification scheme:

> This survey refers to defendants and victims as ''White,'' ''Black,'' ''Hispanic,'' or ''Other,'' due in large part to the way in which data regarding the federal death penalty has been collected. The last category ''Other'' includes any person whose race is Asian, Pacific Islander, Native American, Aleut, Indian, or unknown. The survey uses ''Hispanic'' as a separate category to refer to persons of Hispanic ethnicity, regardless of race. As a result, the terms ''White,'' ''Black,'' and ''Other'' as used in this survey refer only to non-Hispanic members of those racial groups.

3. Victim race. Victim race is the race of the deceased. It is coded into six mutually exclusive categories, the four in defendant race, plus ''None'' and ''Multi.'' ''None'' occurs in crimes where there is no death involved.[1] Victim race ''Multi'' is used when there are multiple victims and those victims do not all fall in the same one of the four race categories.
4. Federal capital crimes. During the study period, 36 different federal capital crimes were charged. Each defendant is charged with at least one of these crimes and for each crime, either the defendant is charged with the crime or he (she) is not (e.g., for the purposes of the data, if a defendant was charged with multiple counts of the same crime, that would be represented in the same way as a defendant who was charged with only one count of that crime). The 36 federal capital crimes charged during the study period can be found in Table 1.
5. Number of victims (multiple victims/not multiple victims).
6. U.S. Attorney recommendation. This is seek the death penalty/do not seek the death penalty/no recommendation.
7. Review committee recommendation. This is seek the death penalty/do not seek the death penalty/no recommendation.
8. Attorney general decision. This is seek the death penalty/do not seek the death penalty/no decision.[2]

3. Inferring Defendant-Level Data

This section describes the process through which the tables of the September 12 Report were combined to obtain defendant-level data through a series of examples. The data in the September 12 Report are a series of two- and three-way tables of the various covariates.

A freedom of information act request was submitted in October 2000 for defendant-level data in these cases. That request was denied, because the U.S. Attorney recommendation and the review committee recommendation are considered to be confidential

---

[1] Most federal capital crimes require that a death be involved with the crime. There are five occurrences in the study period where there was a federal capital crime charged with no death involved. These were five cases involving violations of 18 U.S.C. 794 (Gathering or delivering defense information to aid a foreign government).
[2] There are two circumstances under which the attorney general would not have made a decision on a case. One is that the case had been resolved (e.g., charges were dropped) before the attorney general had an opportunity to make a decision. The second is that the attorney general did not have an opportunity to make a decision by the time the study period ended.

Table 1. Description of federal capital crimes

Crime	Times charged	Description
18 U.S.C. 1111	63	Murder
18 U.S.C. 1114	19	Protection of officers and employees of the United States
18 U.S.C. 1116(a)	3	Murder of foreign officials, official guests, or internationally protected persons
18 U.S.C. 1118	2	Murder by a federal prisoner
18 U.S.C. 1121(a)	5	Killing persons aiding federal investigations or state correctional officers
18 U.S.C. 1201(a)	35	Kidnapping
18 U.S.C. 1203(a)	15	Hostage taking
18 U.S.C. 1512(a)	57	Tampering with a witness, victim, or an informant
18 U.S.C. 1513(a)	22	Retaliating against a witness, victim, or an informant
18 U.S.C. 1716	3	Injurious articles as nonmailable
18 U.S.C. 1958(a)	34	Use of interstate commerce facilities in the commission of murder-for-hire
18 U.S.C. 1959(a)	155	Violent crimes in aid of racketeering activity
18 U.S.C. 2113(e)	22	Killing in conjunction with bank robbery and incidental crimes
18 U.S.C. 2119(3)	71	Death in conjunction with carjacking
18 U.S.C. 2241(a)	1	Aggravated sexual abuse
18 U.S.C. 2241(c)	1	Aggravated sexual abuse (with children)
18 U.S.C. 2332a	5	Use of weapons of mass destruction
18 U.S.C. 241	6	Conspiracy against rights
18 U.S.C. 242	9	Deprivation of rights under color of law
18 U.S.C. 245(b)	6	Federally protected activities
18 U.S.C. 247(d)(1)	1	Damage to religious property; obstruction of the free exercise of religious rights
18 U.S.C. 32	1	Destruction of aircraft or aircraft facilities
18 U.S.C. 33	1	Destruction of motor vehicles or motor vehicle facilities
18 U.S.C. 36	3	Drive-by shooting
18 U.S.C. 37	3	Violence at international airports
18 U.S.C. 794	5	Gathering or delivering defense information to aid foreign government
18 U.S.C. 844(d)	4	Killing in conjunction with interstate/foreign transport of explosives with knowledge or intent of use to kill
18 U.S.C. 844(f)	5	Killing in conjunction with destruction of government property via fire or explosives
18 U.S.C. 844(i)	20	Killing in conjunction with destruction of property used in interstate or foreign commerce via fire or explosives
18 U.S.C. 924(j)	186	Firearm murder
18 U.S.C. 930(c)	6	Possession of firearms and dangerous weapons in federal facilities
21 U.S.C. 848(c)	1	Continuing criminal enterprise
21 U.S.C. 848(e)(1)(A)	128	Continuing criminal enterprise
21 U.S.C. 848(e)(1)(B)	7	Continuing criminal enterprise
49 U.S.C. 46502	1	Aircraft piracy
8 U.S.C. 1324(a)	23	Bringing in and harboring certain aliens

information. The extent to which these are considered confidential is illustrated by the fact that the defendants themselves are not privy to these recommendations. (They do, however, know the attorney general's decision, as this is the official decision of the government and is recorded in court documents by a notice to seek the death penalty.)

The abundance of tables in the September 12 Report provides the opportunity to use those tables to identify covariates on the defendant level. Specifically, it is possible to identify all eight covariates for 386 of the 682 cases, and the joint distribution of five covariates (Defendant race, federal district, U.S. Attorney recommendation, review committee recommendation, attorney general decision) can be identified for all 682 cases. There are 296 cases that have missing data. These cases have missing data on up to three covariates. However, informative constraints can be inferred on the missing data. Additionally it can be demonstrated that the maximal possible amount of information has been obtained from the September 12 Report.

To illustrate how the identification of the crimes charged, victim race, and number of victims works in practice, two examples are given. An example of how external information is used to sharpen the information inferred from the 2000 Report is also provided. Full details of the procedure used to infer data, and evidence that the inference made from the 2000 Report is maximal, are given in the Appendix and in Algranati (2002).

3.1.   Example: Rhode Island

The Rhode Island federal district submitted five cases during the study period. The fully identified defendant-level data for these cases is provided in Table 2.

The entries in Table 2 are reconstructed by linking data from eight different tabular summaries from the 2000 Report. This reconstruction begins with Table 5A. That table contains U.S. Attorney recommendations (seek/do not seek/no recommendation) by federal district and race/ethnicity of defendant. Immediately, this table provides the universe of cases for Rhode Island, and the joint distribution of U.S. Attorney recommendation and defendant race. The U.S. Attorney recommended not to seek the death penalty for all five cases, four of which were against Black defendants, one against a Hispanic defendant.

Information about the crimes charged to these defendants was obtained from Table 8-2 of the 2000 Report. That table contains counts of crimes charged by capital offense, federal district, and race/ethnicity of defendant among defendants whom the U.S. Attorney recommended against seeking the death penalty. This table indicates a total of five crimes charged, all of which are 18 U.S.C. 1959(a), four against the Black defendants and one

Table 2.   *Fully identified cases in the Rhode Island federal district*

#	Crimes charged	Def. race	U.S. Attorney rec.	R.C. rec.	A.G. decision	Vic. quant.	Vic. race
1	18 U.S.C. 1959(a)	Black	Do not seek	Seek	No decision	1	Hispanic
2	18 U.S.C. 1959(a)	Black	Do not seek	Do not seek	Do not seek	1	Hispanic
3	18 U.S.C. 1959(a)	Black	Do not seek	Do not seek	Do not seek	1	Hispanic
4	18 U.S.C. 1959(a)	Black	Do not seek	Do not seek	Do not seek	1	Hispanic
5	18 U.S.C. 1959(a)	Hispanic	Do not seek	Do not seek	Do not seek	1	Hispanic

against the Hispanic defendant. Since the constraints on the reporting of these tables are that each defendant must be charged with at least one crime, and multiple charges of a federal capital crime are reported the same as a single charge, the implication of this table for the Rhode Island defendants is that they all were charged with 18 U.S.C. 1959(a) (and no other federal capital crimes).

The values of the victim race for these defendants are revealed in Table 10A-3 of the 2000 Report. This table reports U.S. Attorney recommendations by federal district by race/ethnicity of defendant *among cases with exclusively Hispanic victims*. Note that this table is a cross of the same variables of Table 5A of the 2000 Report, except that it contains data only on cases with exclusively Hispanic victims. For Rhode Island, this table indicates five cases, implying that all five of the Rhode Island cases are cases with exclusively Hispanic victims.

Number of victims in Rhode Island was inferred similarly to victim race. Table 11A-1 of the 2000 Report provides U.S. Attorney recommendations by federal district by race/ethnicity of defendant *among cases with a single victim*. This table is a cross-tabulation of the same variables as Table 5A of the 2000 Report, except that it contains data only on cases with single victims. For Rhode Island, the table indicates five cases, implying that all five of the Rhode Island cases are cases with single victims.

Reconstruction of the review committee recommendation and the attorney general decision comes from Tables 12, 18A, and 25-1 of the 2000 Report. Table 12 contains review committee recommendations by district and race/ethnicity of defendant. Table 18A contains attorney general decision by district and race/ethnicity of defendant. Table 25-1 contains the cases where the U.S. Attorney, review committee, and attorney general agree, by district and race/ethnicity of defendant. Tables 12 and 18A indicate four cases each, and Table 25-1 indicates four cases (three with black defendants, one with a Hispanic defendant). This implies that in one case with a black defendant in Rhode Island, there was no decision by the review committee and attorney general. For the remaining four cases in Rhode Island, the review committee and attorney general agree with the U.S. Attorney, not to seek the death penalty.

3.2.   Example: Eastern district of Missouri (hereafter Eastern Missouri)

There are five cases from the Eastern Missouri federal district. The defendant-level data identified from the 2000 Report are provided in Table 3. In contrast to Example 3.1, here the 2000 Report does not permit reconstruction of all the variables, specifically the crimes charged.

Table 3.   *Information inferred for the Eastern Missouri federal district*

#	Crimes charged	Def. race	U.S. Attorney rec.	RC rec.	AG decision	Vic. quant.	Vic. race
1	18 U.S.C. 924(j), 18 U.S.C. 1201(a)	Other	Seek	Seek	Seek	1	White
2	18 U.S.C. 2119(3)	Black	Seek	Do not seek	No decision	1	Other
3	Unknown	Black	Seek	Seek	Seek	1	White
4	Unknown	Black	Seek	Seek	Seek	1	White
5	Unknown	Black	Seek	Seek	Seek	1	White

Table 5A of the 2000 Report indicates that there are five cases in the Eastern Missouri district, regarding all of which the U.S. Attorney recommended seeking the death penalty. Among these five cases, four were against Black defendants and one was against a defendant of race ''Other.''

Table 8-1 of the 2000 Report provides counts of crimes charged by capital offense, federal district, and race/ethnicity of defendant among defendants regarding whom the U.S. Attorney recommended seeking the death penalty. This table indicates four charges of 18 U.S.C. 924(j) (three against Black defendants, one against a defendant of race ''Other''), two charges of 18 U.S.C. 1201(a) (one against a Black defendant, one against a defendant of race ''Other''), two charges of 18 U.S.C. 2113(c) (both against Black defendants), and two charges of 18 U.S.C. 2119(3) (both against Black defendants). Since there is only one defendant of race ''Other,'' the two crimes charged to defendants of race ''Other'' must be charged to that defendant.

Table 10A-1 of the 2000 Report provides U.S. Attorney recommendations by federal district by race/ethnicity of defendant *among cases with exclusively White victims*, while Table 10A-4 of the 2000 Report provides U.S. Attorney recommendations by federal district by race/ethnicity of defendant *among cases with exclusively victims of race ''Other.''* Table 10A-1 indicates four cases, three with Black defendant race and one with defendant race ''Other.'' Table 10A-4 indicates one case with Black defendant race.

Number of victims is completely revealed by Table 11A-1 of the 2000 Report, which indicates that there are five cases with single victims, implying that all cases in Eastern Missouri are single-victim cases.

Table 12 of the 2000 Report provides counts of review committee recommendations by district and defendant/ethnicity of defendant. This table reveals that the review committee recommended seeking the death penalty against three Black defendants and one defendant of race ''Other.'' It also indicates that the review committee recommended not seeking the death penalty against one Black defendant.

Table 14-2 of the 2000 Report provides counts of crimes charged by capital offense, federal district, and race/ethnicity of defendant among defendants whom the review committee recommended not to seek the death penalty. This table indicates one count of 18 U.S.C. 2119(3) against a Black defendant, and no other crimes.

Table 16-4 of the 2000 Report provides review committee recommendations by district and race/ethnicity of defendant *among cases with exclusively victims of race ''Other.''* This table indicates one case with a Black defendant against whom the review committee recommended not seeking the death penalty. The implication of this information is that the one Black defendant with a victim of race ''Other'' was charged with 18 U.S.C. 2119(3).

Table 18A of the 2000 Report indicates four cases, in each of which the attorney general decided to seek the death penalty. Table 25-1 of the 2000 Report indicates that there is agreement among the U.S. Attorney, review committee, and attorney general for four cases.

Thus, the one case regarding which the review committee recommended not to seek the death penalty was the one where the attorney general made no decision. Defendant-level data was fully identified for two of these five cases: the one with defendant race ''Other'' and the one case with victim race ''Other.'' For the three remaining cases, defendant-level information on crimes charged cannot be fully identified. There are other tables in the

2000 Report which contain information on crimes charged (Tables 8-2, 14-1, 20-1, and 20-2), but none of these tables can further refine the information about crimes charged to these three Black defendants. What is known is that collectively, these three defendants are charged with three counts of 18 U.S.C. 924(j), one count of 18 U.S.C. 1201(a), two counts of 18 U.S.C. 2113(e), and one count of 18 U.S.C. 2119(3) (note that since there are three charged with 18 U.S.C. 924(j) and three defendants, each defendant is *at least* charged with 18 U.S.C. 924(j)).

3.3.   *Using external information to fully identify Eastern Missouri*

Through accessing external data sources, the uncertainties involved in Table 3 can be eliminated. The federal death penalty resource council tracks data on federal capital cases. Their records indicate that Norris G. Holder and Billie Jerome Allen are Black defendants from the Eastern Missouri district. While there is a narrative describing some of the details of the case, the precise charges against Holder and Allen were not identified by their statutes. However, the case number for these defendants, 97-CR-0141, was provided.

Using either the defendant name or case number, details about the case can be accessed through public access to court electronic records (PACER). PACER is an electronic public access service that allows users to obtain case and docket information from federal appellate, district and bankruptcy courts, and from the U.S. party/case index. Case docket reports in PACER can be searched for by case number or party name. By accessing the court docket reports via PACER, the precise statutes charged against the defendants can be identified.

Through PACER, it was identified that Holder and Allen were each charged with 18 U.S.C. 924(j) and 18 U.S.C. 2113(e). Thus, the remaining defendant must be charged with 18 U.S.C. 924(j), 18 U.S.C. 1201(a), and 18 U.S.C. 2119(3).

The fully-identified cases of Eastern Missouri are detailed in Table 4.

4.   Conclusions

The 2000 Report from the DOJ provided aggregated information for eight variables for 682 federal capital cases from 1995 through 2000. Two of these variables, the U.S. Attorney recommendation whether to seek the death penalty and the attorney general's

Table 4.   Information inferred for the Eastern Missouri federal district

#	Crimes charged	Def. race	U.S. Attorney rec.	RC rec.	AG decision	Vic. quant.	Vic. race
1	18 U.S.C. 924(j), 18 U.S.C. 1201(a)	Other	Seek	Seek	Seek	1	White
2	18 U.S.C. 2119(3)	Black	Seek	Do not seek	No decision	1	Other
3	18 U.S.C. 924(j), 18 U.S.C. 2113(e)	Black	Seek	Seek	Seek	1	White
4	18 U.S.C. 924(j), 18 U.S.C. 2113(e)	Black	Seek	Seek	Seek	1	White
5	18 U.S.C. 924(j), 18 U.S.C. 1201(a), 18 U.S.C. 2119(3)	Black	Seek	Seek	Seek	1	White

review committee on capital cases recommendation whether to seek the death penalty, are considered to be confidential information.

Apparently, since the DOJ released aggregated information, and did not explicitly identify variables at the defendant level, the DOJ believes that it had preserved the confidentiality of the U.S. Attorney and review committee recommendations. However, the large number of tables (over 400 pages) provided in the 2000 Report affords an opportunity to piece together tables and infer the values of all eight variables in the report (including the two confidential variables) at the defendant level.

Using only the 2000 Report, all eight variables can be identified for 386 of the 682 (56.6%) of the cases. Incorporating information from two external sources, the federal death penalty resource council and court records, an additional 160 cases can be identified.

Willenborg and de Waal (2001, p. 17) distinguish between linked tables, ''any set of tables that has been produced from a common microdata file,'' and semi-linked tables ''. . . from different ones (files) that refer to (almost) the same population.'' We think that the example reported here is most accurately thought of as semi-linked tables, since some tables reported decision and recommendations at all levels, while others did not, resulting in tables with different totals. As yet there is scant literature on avoiding disclosure in semi-linked files, so it is not clear that the DOJ could have avoided exposure of sensitive information to this extent using currently available methods. However, in principle they could have done what we have done in the article, which would have alerted them to the issue.

The identification of these data is important for two reasons. One, it allows those interested in exploring the federal death penalty process to do so at the defendant level, using information which was not (explicitly) available previously. Two, it provides a vivid demonstration of how, in practice, confidentiality can be unintentionally breached and information mistakenly disclosed. It is imperative that any company, agency, or individual releasing data, which they consider to be sensitive, consider the totality of the information disclosed. In the case of the 2000 Report, each individual table provided aggregated information which did not compromise confidential information. However, the combination of tables revealed all the variables for over half the cases, undermining the confidentiality of the variables that the aggregated tables and entire report were to preserve.

Appendix

A.   Piecing together tables to get higher-dimensional tables

Several of the three-dimensional tables in the September 12 Report are conditional on a particular value of a fourth covariate. For example, September 12 Table 10A-1 is U.S. Attorney recommendation by federal district and defendant race for cases with White victims. September 12 Tables 10A-2 through 10A-5 provide U.S. Attorney recommendation by federal district and defendant race for cases with Black, Hispanic, Other, and ''Multi'' victims, respectively. Combining these five tables gives the four-way table of U.S. Attorney recommendation, federal district, defendant race, and victim race.

The synopsis of September 12 tables used to infer defendant-level data is given in Table 5. The rows of this table represent multi-dimensional tables and the columns

represent the variables in those tables. For convenience, labels (''Group A'' through ''Group M'') have been assigned to these tables. An important note with regard to the columns in Table 5 is that the data for the recommendations and decisions are binary (seek/no seek). For example, Group J provides the joint distribution of federal district, defendant race, U.S. Attorney recommendation, review committee recommendation, and attorney general decision, *given that all three decisionmakers made a seek/no seek recommendation/decision*. Section 2 identifies these variables as being ternary. Specifically, it is possible for each decisionmaker to make no decision. Section A.1 addresses how to use the September 12 Report to access the joint distribution of defendant race, federal district, U.S. Attorney recommendation, review committee recommendation, and attorney general decision.

A.1   Obtaining the joint distribution of the recommendation/decisions

A cornerstone of the inference of the data is that the five variables, U.S. Attorney recommendation, review committee recommendation, attorney general decision, federal district, and defendant race can be inferred for all cases.

Group J gives this five-way table, but only for the binary coding (seek/no seek) of the recommendation/decision variables. In some cases, at least one of the decisionmakers did not make a recommendation or decision. Group J accounts for 571 cases, implying that at least one of the decisionmakers did not make a recommendation or decision in 111 cases. More information must be gathered to find out the pattern of the covariates for these 111 cases.

Group K gives the distribution of defendant race by federal district by U.S. Attorney recommendation by attorney general decision (again for seek/no seek decisions only). Group L gives the distribution of defendant race by federal district by U.S. Attorney recommendation by review committee recommendation (for seek/no seek decisions only). Group M gives the distribution of defendant race by federal district by review committee recommendation by attorney general decision (for seek/no seek decisions only). These tables reveal that there are 602 cases where the U.S. Attorney and the review committee both made recommendations, 575 cases where the U.S. Attorney and the attorney general made recommendations/decision, and 572 cases where the review committee and the attorney general made recommendations and decisions.

The defendants in Group J are a proper subset of those in Group K. A comparison of these two groups reveals that there are four cases in which the U.S. Attorney and attorney general made decisions and the review committee did not. Moreover, by locating the four cells in these two groups that are different, the values of federal district, defendant race, U.S. Attorney recommendation and attorney general recommendation are identified for these four cases.

Similarly, the defendants in Group J are a proper subset of those in Group L. A comparison of these two groups reveals that there are 31 cases in which the U.S. Attorney and the review committee made recommendations, but the attorney general did not make a decision. By locating the 31 cells in these two groups that are different, the values of federal district, defendant race, U.S. Attorney and review committee recommendations are identified for these 31 cases.

Likewise, the defendants in Group J are a proper subset of those in Group M.

Table 5.   *Synopsis of variables covered by tables in the September 12 Report (''X'' indicates which table(s) include information about that variable)*

Sept. 12 Tables #	Group	Cases	Federal district	Defendant race	Crimes	Victim race	Number of victims	U.S. Attorney rec. (Seek/No seek)	Review committee rec. (Seek/No seek)	Attorney general decision (Seek/No seek)
8-1, 8-2	A	677	X	X	X			X		
10A-1, 10A-2, 10A-3, 10A-4, 10A-5	B	677	X	X		X		X		
11A-1, 11A-2	C	677	X	X			X	X		
14A-1, 14A-2	D	603	X	X	X				X	
16-1, 16-2, 16-3, 16-4, 16-5	E	603	X	X		X			X	
17-1, 17-2	F	603	X	X			X		X	
20-1, 20-2	G	576	X	X	X					X
22A-1, 22A-2, 22A-3, 22A-4, 22A-5	H	576	X	X		X				X
23A-1, 23A-2	I	576	X	X			X			X
25-1, 25-2, 25-3, 25-4	J	571	X	X				X	X	X
26A-1, 26A-2	K	575	X	X				X		X
27-1, 27-2	L	602	X	X				X	X	
28-1, 28-2	M	572	X	X					X	X
7	N	677		X	X			X		

A comparison of these two groups reveals that there is one case in which the review committee and attorney general made decisions and the U.S. Attorney did not make a recommendation. By locating the one cell in which these two groups differed, the values of federal district, defendant race, review committee recommendation and attorney general decision can be identified for this one case.

Group G indicates that there are a total of 576 cases where the attorney general made a decision (seek/do not seek). From the information laid out above, these 576 cases can be partitioned into 571 cases where both the U.S. Attorney and the review committee made a recommendation, four cases where the U.S. Attorney made a recommendation, the review committee did not, and one case where the U.S. Attorney did not make a recommendation, and the review committee did. The implication is that there are no cases in which the attorney general made a decision while the U.S. Attorney and the review committee made no decisions.

Group D indicates that there are a total of 603 cases where the review committee made a recommendation. From the information laid out above, these 603 cases can be partitioned into 571 cases where both the U.S. Attorney made a recommendation and the attorney general made a decision, 31 cases where the U.S. Attorney made a recommendation and the attorney general did not make a decision, and one case where the U.S. Attorney did not make a recommendation and the attorney general made a decision. This implies that there are no cases in which the review committee made a decision while the U.S. Attorney and attorney general made no decisions.

Group A indicates that there are a total of 677 cases where the U.S. Attorney made a recommendation. From the information laid out above, these 677 cases can be partitioned into 571 in which both the review committee and the attorney general made a decision, four cases in which the attorney general made a decision and the review committee made no recommendation, and 31 cases in which the review committee made a recommendation and the attorney general made no decision. This implies that there are 71 cases in which the U.S. Attorney made a decision, while neither the review committee nor the attorney general made a decision.

There are a series of steps to arrive at the joint distribution of the federal district, defendant race, U.S. Attorney recommendation, review committee recommendation, and attorney general decision for these 71 cases in which the U.S. Attorney made a recommendation, but the review committee and the attorney general did not.

First, construct a five-dimensional table of federal district, defendant race, U.S. Attorney recommendation, review committee recommendation, and attorney general decision which contains the following data:

- the 571 cases from Group J;
- the 4 cases in which the U.S. Attorney and attorney general made decisions, but the review committee did not;
- the 31 cases in which the U.S. Attorney and review committee made recommendations, but the attorney general did not.

Collapse this table which contains data on 606 cases over review committee recommendation and attorney general decision. This results in a three-dimensional table of federal district, defendant race, and U.S. Attorney recommendation for 677 cases. Collapsing

Table 6.   Number of cases by existence of recommendations/decisions

U.S. Attorney recommendation	Review committee recommendation	Attorney general decision	Count
Yes	Yes	Yes	571
Yes	No	Yes	4
Yes	Yes	No	31
No	Yes	Yes	1
Yes	No	No	71
No	Yes	No	0
No	No	Yes	0
No	No	No	4
Total			682

Group A over crimes charged gives a three-dimensional table of federal district, defendant race, and U.S. Attorney recommendation for 606 cases. Taking the cell-wise differences between these two three-dimensional tables results in the distribution of federal district, defendant race, and U.S. Attorney recommendation for the 71 cases in which the U.S. Attorney made a recommendation while the review committee and attorney general made no decisions.

The distribution of cases by the existence of recommendations is given in Table 6. The data in Table 6 are reasonable in the light of what is understood about the prosecutorial process. The four cases in which no recommendations were made, nor decisions recorded, the 71 cases in which the U.S. Attorney made a recommendation but neither the review committee nor the attorney general acted, and the 31 cases in which both the U.S. Attorney and the review committee made recommendations but the attorney general did not decide, are all incomplete records of the normal process. Some of them could reflect cases under-going review at the time the data were gathered; the remainder would be the result of plea bargains rendering the rest of process superfluous. The 71 cases in which the U.S. Attorney and review committee made recommendations and the attorney general made a decision represent the full exercise of the intended procedure. Hence there are only five anomalous cases of attorney general decisions, four without a review committee recommendations and one without a U.S. Attorney recommendation.

Note that, in Table 6, there are four cases in which neither the U.S. Attorney nor the review committee nor the attorney general made a decision. September 12 Tables 11A-1 and 11A-2 (which collectively are Group C) contain columns on ''Total Defendants Submitted by U.S. Attorneys'' as well as columns on ''U.S. Attorney recommendations to seek the death penalty'' and ''U.S. Attorney recommendations not to seek the death penalty.'' These columns are broken down by federal district and defendant race. When the sum of the latter two columns does not equal the former, that indicates that the U.S. Attorney did not make a recommendation. There are five such cases. One of these is accounted for by the one case in which the U.S. Attorney did not make a recommendation and the review committee and attorney general did make decisions. The remaining four are the four cases in which neither the U.S. Attorney, review committee, nor the attorney general made a decision. The federal district and defendant race for these cases is identi-fied by identifying in which cells the ''Total defendants submitted by the U.S. Attorneys''

is not equivalent to the corresponding cells in ''U.S. Attorney recommendations to seek the death penalty'' and ''U.S. Attorney recommendations not to seek the death penalty.''

A.2.   Identification of crimes charged, victim race, and number of victims
While Section A.1 describes how the joint distribution of federal district, defendant race, U.S. Attorney recommendation, review committee recommendation, and attorney general decision is obtained from the September 12 Report for all 682 cases, it is not possible to obtain the joint distribution of all eight variables for all 682 cases based only on the information provided in the September 12 Report.

As Table 5 reveals, there is no group of tables containing information on crimes charged, victim race, and number of victims with more than one recommendation/ decision. Moreover, there is no group containing joint information with any two of the variables: crimes charged, victim race, and number of victims.

Inference of the values (or bounds on the values) of crimes charged, victim race, and number of victims can be arrived at by looking at the marginal distributions of these variables given one of the recommendations/decisions, and the fact that the joint distribution of the pattern of recommendations/decisions is known.

Groups A, D, and G contain information on crimes charged only if a recommendation/ decision was rendered by the U.S. Attorney, review committee, and the attorney general, respectively. To identify the crimes charged for the four defendants that had no recommendation or decision, Group N is used. Group N contains information on crimes charged by defendant race by U.S. Attorney recommendation. This lists the number of times a crime is charged, and then categorizes that number by displaying counts of that crime by U.S. Attorney recommendation and defendant race. When the sum of the elements are not equal to the total, that implies that there were instances of that crime which were charged to someone with whom the U.S. Attorney did not make a recommendation. This process was able to uniquely identify the crimes charged to these four individuals.

Since Groups B, E, and H have similar total columns, the victim race of these defendants with no recommendations/decisions was able to be identified. Likewise, Groups C, F, and I contain total columns which can be used to identify the number of victims (single/multiple) of those who have no recommendation or decision.

B.   *Quality of inferred values*

The methodology outlined in this appendix results in 386 fully-identified cases and 296 cases with up to three variables with missing data. In the 296 cases with missing data, informative constraints were inferred for these values. This section demonstrates that the values inferred are the maximal amount of information that could be inferred from the September 12 Report.

To demonstrate that the values inferred are maximal, this section illustrates that the tables used to infer the defendant-level covariates can be recreated from the constructed dataset.

Since the five-variable joint distribution of federal district, defendant race, U.S. Attorney recommendation, review committee recommendation, and attorney general decision is known, any table which is a function of some combination of these five variables can be created. This accounts for the recreation of Groups J, K, L, and M from Table 5.

The four-variable joint distribution of federal district, defendant race, crimes charged, and any one of U.S. Attorney recommendation, review committee recommendation, and attorney general decision is known. This enables the recreation of Groups A, D, and G.

The four-variable joint distribution of the federal district, defendant race, victim race, and any one of the U.S. Attorney recommendation, review committee recommendation, and attorney general decision is known. This enables the recreation of Groups B, E, and H.

The four-variable joint distribution of federal district, defendant race, number of victims (multiple victims/not multiple victims), and any one of the U.S. Attorney recommendation, review committee recommendation, and attorney general decision is known. This enables the recreation of Groups C, F, and I.

This accounts for the creation of Groups A through M, the tables used to construct the inferred data.

The remaining tables in the DOJ report about the federal death penalty cases from 1995–2000 can either be arrived at by simply collapsing over one of the groups (this is true of Tables 5A, 10A, 11A, 12, 13, 16, 17, 18A, and 22A) or be recreated from the inferred dataset (1A, 2A, 13, and 19).

Throughout most of the September 12 Report, number of victims is treated as being either single or multiple victims. Three tables in the September 12 Report, 9A, 15, and 23A, deal with the number of victims as an actual count. The count of victims in a given district could be less than or greater than the number of defendants. For example, if there are two defendants in a district, and they are both charged with killing the same individual, then that district has one victim. An example of there being more victims than defendants would be if there was one defendant who is charged with killing six people. Then that district had one defendant and six victims.

Most of the tables in this report pertaining to victims in the September 12 Report deal with either single or multiple victims. The information inferred about number of victims for this data set is ternary (single/multiple/none). Using that ternary variable, it is not possible to recreate the victim count data of September 12 Tables 9A, 15, and 23A. Note, however, that these tables with victim counts were used on three occasions when they were helpful in reducing uncertainty about the matching of victim race and number of victims. For example, consider a district where there were three defendants, and it is known that collectively, one had Hispanic victim(s) and two had Black victim(s), and that collectively, two had single victim(s) and one had multiple victim(s). But the matching of victim race and number of victims could not be inferred from the information in Groups A–N. If the victim count table revealed that this district had one Black victim and four Hispanic victims, then this implies that the multiple victim case is the case with Hispanic victim race.

Whether or not information from September 12 Tables with victim counts is used, Groups A–N can be recreated from the inferred data.

5. References

Algranati, D. (2002). Exploring Racial and Geographical Effects in the Decision to Seek the Federal Death Penalty, 1995–2000. Ph.D. Thesis, Carnegie Mellon University.

Algranati, D. and Kadane, J. (2003). Racial Effects in the Federal Decision to Seek the Death Penalty. Forthcoming.

Brune, T. (2000). Death Penalty Disparities/Federal Study Finds Minorities Overrepresented on Death Row. Newsday, September 13, A05.

Fienberg, S. and Willenborg, L. (eds.) (1998). Special Issue on Disclosure Limitation Methods for Protecting the Confidentiality of Statistical Data. Journal of Official Statistics, 14, 337–345.

Giessing, S. (2001). Nonperturbative Disclosure Control Methods for Tabular Data, Chapter 9 in Confidentiality, Disclosure and Data Access: Theory and Practical Application for Statistical Agencies, P. Doyle, J. Lane, J. Theeuwes, and L. Zayatz (eds.), Elsevier Science.

Jackson, R.L. (2000). Study Finds Racial Gap on Death Row: More Black Convicts Face Capital Punishment than Whites. Activists Call for a Federal Moratorium on Executions. Los Angeles Times, September 13, A5.

Johnson, K. (2000). Big Disparities in the Federal Death Penalty. U.S.A. Today, September 13, 4A.

Lacey, M. and Bonner, R. (2000). Reno Troubled by Death Penalty Statistics. The New York Times, September 13, A17.

U.S. Department of Justice (2000). The Federal Death Penalty System: A Statistical Survey (1988–2000). http://www.usdoj.gov/dag/pubdoc/dpsurvey.html

U.S. Department of Justice (2001). The Federal Death Penalty System: Supplementary Data, Analysis, and Revised Protocols for Capital Case Review, June 6. http://www.usdoj.gov/dag/pubdoc/deathpenaltystudy.html

Willenborg, L. and de Waal, T. (1996). Statistical Disclosure in Practice. Lecture Notes in Statistics #111. Springer Verlag, New York.

Willenborg, L. and de Waal, T. (2001). Elements of Statistical Disclosure in Control. Lecture Notes in Statistics #155. Springer Verlag, New York.