

## DATA APPENDIX

### *Construction of the Linked Dataset:*

The primary data for this analysis are restricted-use data collected by the Bureau of Justice Statistics (BJS) and made available to researchers via the National Archive of Criminal Justice Data.<sup>1</sup> Using the restricted-use dyadic linking files provided by BJS, we link case records across agencies as they go through the criminal justice process, enabling us to follow defendants from arrest through to sentencing. To that end, we link cases files across the following agencies: the U.S. Marshals' Service (USMS), the Executive Office for the U.S. Attorneys (EOUSA), the Administrative Office of the U.S. Courts (AOUSC) and the U.S. Sentencing Commission (USSC). USMS data cover all persons arrested by federal law enforcement agencies, persons arrested by local officials and then transferred to federal custody, and persons who avoid arrest by self-surrendering.<sup>2</sup> The EOUSA data come from the internal case database used by federal prosecutors. It contains a variety of charging and investigation-related information for every case in which any prosecutor at a U.S. Attorney's office opens a file.<sup>3</sup> The AOUSC data are from the Federal Courts and contain data on all criminal cases heard by federal district judges, and any non-petty charge handled by a federal magistrate judge. The USSC collects data on any case that results in conviction and sentencing for a non-petty offense.

The term "petty offense," as used here, refers to anything other than a Class A misdemeanor or felony.<sup>4</sup> By definition, such offenses carry statutory maximum penalties of six months or less; in practice the great majority (75 percent in our sample) result in non-incarceration sentences. Although the USSC does not collect data on these cases, their

---

<sup>1</sup> Descriptions of the files are available at <http://www.icpsr.umich.edu/icpsrweb/content/NACJD/guides/fjsp.html>. Full citations are included below under Data Sources.

<sup>2</sup> Although the last group are not technically "arrests," we refer to these data collectively as "arrest data."

<sup>3</sup> The U.S. Attorney's offices are the main federal prosecutors' offices, organized by district. They handle 95% of all federal prosecutions; the remainder are handled by specialized offices in the Department of Justice.

<sup>4</sup> In the federal system, Class A misdemeanors carry one-year statutory maximum prison sentences, and felonies are anything with a higher statutory maximum. Magistrate judges may enter convictions and sentences for Class A misdemeanors if the parties consent (28 U.S.C. § 636), and records of such cases are expected to be submitted to AOUSC and USSC. They may enter convictions and sentences for petty offenses even without the parties' consent (28 U.S.C. § 636), and hence such charges rarely appear before district judges.

dispositions and in 98% of the cases their sentences are recorded in the EOUSA suspect investigation files and/or the AOUSC data. Our main sample includes only 279 petty offenses.

Our objective in constructing the main sample was to include all cases (within offense-type, date, and offender restrictions discussed below) in which (1) an offender was arrested or self-surrendered, (2) the case was referred to a federal prosecutor, and (3) the case proceeded to some kind of substantive conclusion, whether it was declination by the prosecutor, dismissal, acquittal, or conviction and sentencing. The goal of this study is to quantify the extent of any post-arrest racial disparity in sentencing and to understand the role of prosecutorial discretion in generating that disparity. We therefore did *not* include cases in which the subject of an investigation was never apprehended, cases in which a person was arrested but immediately released without ever being referred to prosecutors, or cases in which the defendant died before the case ended. We also excluded cases in which the arresting officers or the prosecutor transferred the case to a different authority for prosecution (e.g., state prosecutors or military prosecutors).<sup>5</sup> It was also necessary to exclude the 0.25% of cases that had not reached a conclusion by the end of FY2009, the last year on which we have data. In order to avoid double-counting cases, we also excluded any record in which the final disposition was superseded by a new filing or transfer to another federal district; in such cases, new records are created by the new district or filing. These new records contain all the original arrest information and already appear in our sample under the new filing or in the district that ultimately handled the case.

The BJS linking files have high, but not perfect match rates. The linking algorithm is dyadic, such that agencies' files must be linked in the following order: from USMS to EOUSA investigation and case files, from EOUSA to AOUSC case files, and from AOUSC to USSC.<sup>6</sup>

---

<sup>5</sup> Such cases could not reasonably be coded as favorable dispositions, because the charges brought by the other authority could be severe; furthermore any disparities in the outcomes of those cases would come from the disparities present in the post-arrest process in the other level of government.

<sup>6</sup> More specifically, there are multiple types of "Standard Analysis Files" (SAFs) from EOUSA ("Matters Out" files on criminal suspects, "Cases In" files on cases filed in district court in each year, and "Cases Out" files on cases terminated in district court in each year) and from AOUSC ("Cases In" and "Cases Out," for cases filed and terminated each year, respectively). BJS's linking algorithm offers two possible linking pathways, one that connects the "Cases In" files between EOUSA and AOUSC, and one that connects the "Cases Out" files; each pathway is then supplemented with a separate set of intra-agency links so that all SAFs from each agency could be used whenever possible. Both pathways were used to maximize total linking rates. SAFs were appended across years before being linked across agencies. The Cases In and Cases Out files for each agency contain redundant fields; Cases Out is simply an updated version created when the case is terminated that adds the terminal-stage information (e.g., the disposition of each charge). The AOUSC Cases in file was our preferred source of information on initial charges (except in unusual cases where only Cases Out was available due to an intra-agency linking problem),

The estimation sample is limited to cases that could be linked from the USMS records to the first of the EOUSA files: the prosecutor’s investigation file, which the prosecutor opens for any case referred to her, even if she then declines to pursue charges. Approximately 83% of USMS records linked through to EOUSA investigation files; non-links could either be because the case was (for reasons discussed above) never referred to an assistant U.S. Attorney for potential prosecution or because of a failure of the BJS linking algorithm.<sup>7</sup> Conditional on arrest offense, district, and age (the covariates from the main analysis that are contained in the USMS data), black arrestees’ files were slightly more likely to link to a prosecutor’s investigation file (a potential, but likely minor, source of downward bias in our main estimates of racial disparity).<sup>8</sup>

Among the cases linking through to the EOUSA, we excluded 493 people ) whose cases were declined by the federal prosecutor for reasons *not* reflecting a favorable disposition of the case: transfers to other districts or prosecution by other authorities and files deemed mistakenly opened due to administrative error. The probability of a case being declined for such a reason did not vary significantly by race, conditional on the observable covariates.<sup>9</sup> Other declinations (i.e. those that resulted in the defendant not being prosecuted by anyone for the arrest offense) were included in the sample and coded as favorable dispositions (zero sentence, zero charge severity).

---

because that was the source that was available for cases not yet terminated by the end of FY 2009. Coding of final charges was based on AOUSC Cases Out.

<sup>7</sup> All linking percentages are within a pool of cases that tracks the main sample used for analysis as closely as possible without using fields that are only available for cases that successfully link. Thus, this percentage is within cases identified by USMS as black or white U.S. citizen males, excluding drug and immigration crimes, petty traffic and liquor-related arrest codes, and cases from districts lacking sufficient racial diversity; these exclusion criteria are detailed below. Cases that did not link at this stage could have been due to failures of the linking algorithm or to the case not being referred to the U.S. Attorney’s offices.

<sup>8</sup> This is based on logistic regressions of an indicator of linking failure within the same pool of cases for which linking percentages are reported, with standard errors clustered at the district level. The odds ratio for “black” was 0.893 (0.046), i.e., black suspects were slightly less likely to *fail* to link. To the extent that failure to link might reflect a positive outcome—for instance, law enforcement letting the suspect go—this difference might introduce downward bias in our main black-white disparity estimates, although as discussed in the paper, we do not purport to be able to measure disparities in treatment by law enforcement. In any case, linking failures might also indicate failures of the BJS linking algorithm. Because the USMS data contains no disposition fields, we cannot differentiate the explanations for linking failures at this stage. In any event, the difference in link-through rates, while statistically significant, is small enough that the resulting selection bias should be minor—for an average defendant, black race is associated with an increase in link-through rates from 80% to 82%.

<sup>9</sup> This is based on a logistic regression within a sample that added this subset of declined cases back to the main sample. The indicator “non-favorable declination” was regressed on the variables available for the declined cases: race, arrest offense, district, age, and the multi-defendant case indicator, with standard errors clustered at the district level, the same level used in the main analysis. There was also no significant difference in the sample mean rates of non-favorable declination for blacks and whites.

The data meeting these sample criteria were next linked to the AOUSC records. It was not necessary to exclude all of the cases that could not be linked at this stage. Non-linking cases that are recorded in the EOUSA files as declinations or as being disposed of by magistrate were not excluded from the sample because the non-linking was interpreted as a substantive result of the case's disposition. For magistrate cases that did not link to AOUSC records, charge and outcome data were obtained from the EOUSA files. Thus, only two subsets of the cases that failed to link at this stage were excluded from the sample: those that represented apparent linking algorithm failures (2.5% of total cases were excluded for this reason) or those that did link to an initial AOUSC charge record but were not terminated in time for final disposition information to be available (0.25%). Conditional on arrest offense, district, age, and multi-defendant case structure, there were no significant racial differences in the rate at which these cases were excluded from the sample.<sup>10</sup>

Lastly, cases were linked, when possible, to the USSC data. Among cases recorded by the AOUSC data as resulting in non-petty convictions (the cases that USSC is supposed to keep records on), 95% successfully linked through to the USSC data. The 5% of non-petty convictions that failed to link to the USSC were interpreted as linking algorithm failures and were excluded. Conditional on arrest offense, district, age, and multi-defendant case structure, there were no significant racial differences in the probability of linking to the USSC data.<sup>11</sup>. There is no theoretical reason to expect the imperfections in the linking algorithms constructed by BJS to substantively vary by race and therefore bias the estimates of black-white racial disparity.

Ultimately, 89.4% of arrests meeting our initial sample criteria linked all the way through to the USSC data, reflecting conviction and sentencing on a non-petty offense.

*Additional Sample Selection Criteria:*

**Timing:** The sample was limited to individuals who were initially charged or arrested (whichever was later) between fiscal years 2006 and 2008 and whose cases were resolved by the

---

<sup>10</sup> This is based on logistic regressions of an indicator of linking failure within the same pool of cases for which linking percentages are reported, with standard errors clustered at the district level.

<sup>11</sup> This is based on a logistic regression of an indicator of linking failure within the pool of cases recorded by AOUSC as resulting in a nonpetty conviction, with standard errors clustered at the district level.

end of fiscal year 2009 (the most recent year available).<sup>12</sup> These time constraints were chosen so as to produce a sample that was as recent as possible, processed entirely after the Supreme Court's 2005 decision in *United States v. Booker*, and sufficiently large for meaningful analysis, while still allowing enough time for nearly every case to reach conclusion by the end of fiscal year 2009. Few of the above-described linking failures appear to have been the result of cases not having been terminated in time. The link rates at each stage are not substantially different for cases initiated in fiscal years 2006, 2007, and 2008, and among the cases that had AOUSC case initiation records ("Cases In") during those three years, 99.75% were disposed of by the end of FY 2009 and thus had case termination records ("Cases Out").

**Defendant Type:** Most federal felonies render non-citizens deportable (often automatically), and therefore the effective severity of charges in non-citizens' cases is not readily comparable to that of citizens, who were accordingly excluded from the main analyses (although the results are robust to their inclusion; Table 4). U.S. citizenship was identified based on the USMS data. Women were also excluded based on information on arrestee gender contained in the USMS data. The sample was also limited to black and white arrestees.

**Case Type:** Due to the deportation stakes, weaker linking rates,<sup>13</sup> and the very different procedural framework many districts apply to immigration cases,<sup>14</sup> immigration cases were also excluded (almost all of which would have been excluded in any event due to the citizenship requirement).

Drug cases were excluded from the principal analysis sample, although included in robustness checks, due to the lack of reliable drug quantity information at the charging and arrest stages. Quantity is crucial to understanding the severity of drug crimes, and its absence makes the arrest offense a more limited measure of the crime alleged. EOUSA's data do include a "quantity seized" field representing the prosecutor's recording of the amount of drugs seized at arrest; this field is initially recorded in the suspect investigation file, prior to charging, and thus might serve as arrest-stage data (although its recording may involve some prosecutorial discretion). However, we discovered drastic changes in the apparent quantity distribution in this

---

<sup>12</sup> In cases in which no formal indictment was brought (meaning the parties agreed to allow the case to proceed on the "complaint" underlying the arrest warrant), the arrest date was treated as the relevant date.

<sup>13</sup> The linking files do not produce as high link rates for immigration cases, presumably because of the lower quality of identification data in many immigration cases—for example, absence of Social Security Numbers.

<sup>14</sup> Most immigration cases are processed in "fast-track" proceedings in which defendants waive many ordinary procedural protections.

field from 2003 to 2004 (the year EOUSA adopted a new data entry system) as well as large inconsistencies in quantity between this field and the sentencing-stage quantities recorded by USSC beginning in 2004, suggesting that the problem is with the post-2003 EOUSA data.<sup>15</sup> We could identify no reason for the shifts other than data entry problems, and the problems were not uniformly applicable nor confined to particular drug types or districts. There was no way to identify which specific cases had the wrong quantities recorded. We analyzed quantity data from 2001 through 2003 and found no racial disparities in quantity seized, nor did adding controls for quantity change the estimate of racial disparity in sentencing in that period.

Child pornography cases were also excluded from the main sample, although included in additional analyses, due to our inability to distinguish between key sub-provisions of the charging statutes listed by AOUSC. In particular, simple possession cases could not be distinguished from receipt or distribution cases, a distinction critical to the application of a mandatory minimum sentence. For this reason, it was necessary to exclude these cases from the main analysis. Traffic and liquor offenses, which are typically disposed of via citations, were also excluded.

These case type exclusions were based exclusively on the USMS arrest code, which is based on the arresting officer's characterization of the principal offense in the case, not on how the prosecutor chose to charge the case.<sup>16</sup> Although a person can be arrested for multiple offenses in a case, under federal sentencing law most sentences run concurrently, and the principal (most severe) offense is of special importance because it is expected to be the main and usually the only driver of the ultimate sentence. Defining the sample based on the arrest stage data alone (rather than the nature of subsequent charges in the case) avoided potentially serious sample selection issues that could have emerged had the exclusions been based on the

---

<sup>15</sup> Although USSC records drug quantity, this field is the product of sentencing fact-finding and influenced by charging and plea negotiations.

<sup>16</sup> There is no single arrest code for child pornography. The main sample excluded several arrest codes related to obscenity and child sexual exploitation. The written arrest description makes clear that at least 80% of cases with these codes were child pornography-related, and these codes cover over 95% of the cases with child pornography charges. Because there are only a small number of black arrestees with these codes (198 in our broader sample used in Table 6, or approximately 0.5% of the black cases; 96% of the arrestees with these codes are white), the decision to include or exclude cases with these codes has relatively little effect on estimated racial disparities conditional on arrest offense.

prosecutor's discretionary decisions.<sup>17</sup> Because criminal cases sometimes involve multiple types of criminality, this approach did not eliminate the possibility of secondary offense conduct related to the excluded categories. In the case of immigration and child pornography, this was rare: 0.04% and 0.2% of the cases in the sample involved any ancillary immigration and child pornography charges, respectively. However, a drug charge was brought in 9.8% of the cases, 83% of which were principally coded as weapons cases (a drug nexus is often the basis for federal criminal jurisdiction over gun possession cases). The estimates are robust to the inclusion of a flag for any mention of drugs in the written arrest description.

Cases with arrest codes indicating a reason for detention other than a criminal offense (material witness warrants and violations of the conditions of parole or probation) were also excluded from the sample. Finally, since our estimation strategy relies on comparing the sentences of black and white men arrested for the same crime, we excluded individuals arrested for extremely rare crimes. We excluded cases with grouped arrest offense codes (discussed below) that had fewer than fifteen cases in the entire period. This exclusion affected only 57 cases (0.15% of the sample before the exclusion).

The sample was also limited to the 50 US States and the District of Columbia; the territorial districts (Puerto Rico, Guam, U.S. Virgin Islands) were excluded.

#### *Construction of Key Independent Variables:*

The USMS data are the source of the following fields: arrest offense, race, age, gender, police notes describing the arrest offense, U.S. citizenship status, arrest date, the federal judicial district, and the arresting agency. The dates investigations were opened and charges were filed were drawn from the EOUSA's data, as were the number of defendants in each case; and the outcomes (guilt and sentence) of cases that reached final disposition but did not appear in the AOUSC and USSC data (principally minor cases resolved by magistrates). The EOUSA data were also used to determine whether the seizure of drugs upon arrest was mentioned in the prosecutor's investigation-stage records.

The initial charge information used in the analysis relied principally on the detailed charge information collected by the AOUSC from district courts. The AOUSC data identify five

---

<sup>17</sup> For instance, if prosecutors disproportionately declined to pursue a gun charge in guns-and-drugs cases involving white defendants, white defendants in such cases would disproportionately disappear from the sample if it were defined based on the charges rather than the arrest offense.

initial charges in each case by the U.S. Code subsection corresponding to the crime. The data also list up to five final charges and the disposition of each charge, and this information was used to identify the charges of conviction. County and counsel type are also come from the AOUSC data. USSC data were used as an alternate source of information on the final mandatory minimum.

**Race:** Race is drawn from the USMS data, and is coded as white, black, Asian, Native, and Other/Unknown. The last three groups together constituted about 4% of the cases otherwise satisfying the sample requirements, and were dropped from the sample. The USMS does not include a separate category for Hispanic; rather, Hispanics are included within other racial groups. The USSC does record Hispanic ethnicity, but this field is missing in 11% of the cases that linked to the USSC and in all of the cases that did not. Among the cases that do have this field recorded, about 19% of white defendants and 1% of black defendants are identified as Hispanic.

**Arrest Offense:** There are 430 unique arrest offenses listed in the USMS data. However, over 95% of the cases in our sample fall under just 71 arrest offenses. The original arrest offense codes included many very similar offense descriptions, including some that were slightly more detailed versions of others (for instance, “vehicle theft” and “vehicle theft by bailee”). Often the more detailed ones were rarely used. Therefore, the smallest categories were combined with the other categories that could describe the same legal offense, leading to 98 offense groups within the main analysis sample. No single numerical cutoff was used to determine when cases would be combined, because the combination depended on the legal assessment that the crimes were sufficiently similar. For the purpose of the robustness checks that included drug cases, we also subdivided some of the large drug arrest offense codes based on more specific information about the type of drugs seized, which was available from the prosecutor’s initial investigation file. The results are robust to the use of the original offense codes.

**Criminal history and education.** The criminal history variable used was the defendant's criminal history category, which ranges from 1 to 6 and forms the basis of the Guidelines sentencing grid. Education was coded in four categories: high school dropout, high school diploma, GED or vocational training, and at least some college. Criminal history and education



are only recorded by the USSC, and are missing for 11% and 13% of cases, respectively. We accordingly imputed the missing values using a “hotdeck” method, as described in the paper.

**Charge category and disposition.** The disposition of each charge filed (i.e., conviction, dismissal, or acquittal) was determined based on AOUSC records whenever possible. AOUSC also clearly labels each charge as petty, Class A misdemeanor, or felony. As noted above, for magistrate cases not included in the AOUSC, disposition and charge information was drawn from EOUSA data, and where the charge field was missing, such cases were assumed to be petty offenses.

**Sentence:** About 9.9% of cases in the sample resulted in declinations, dismissals, or acquittals, in these cases the sentence was coded as zero. For almost all of the other cases (89.4% of the total sample), the sentence information was taken from the USSC data. Sentence information for the remaining 0.7% of cases taken from AOUSC data (if available) or from EOUSA data.<sup>18</sup> We top-coded sentences at 540 months and assigned that value to all life sentences. This length is longer than the highest non-life statutory maximum found in federal law (480 months), and corresponds approximately to the remaining life expectancy of an American of the sample-average age of 36 years. Only 0.4% of cases were affected by this top-coding. Our sentence outcome of interest was incarceration sentence length, so we coded probation- and fine-only sentences as zeroes (17.7% of the sample).

**Charge Severity Measures:** The raw charge data in AOUSC consist of the statutory provision associated with each charge. This information, combined with legal research, was used to identify the statutory maximum and minimum sentence.

The core coding challenge was that the AOUSC charge fields are not always very specific—they might, for instance, refer to a particular statutory provision that contains two distinct subparagraphs with different sentencing schemes. We researched the most common ways in which these statutes are charged in order to be able to make realistic assumptions in the face of such ambiguities. In general, for instance, we assumed the defendant had no prior convictions of the exact same crime, thus avoiding special penalties that a few statutes apply to recidivists and focusing on the severity of the particular offense in question.<sup>19</sup> If none of the

---

<sup>18</sup> Just four petty offense cases had no sentence information and were coded as zeros by default, based on the typical pattern for petty cases.

<sup>19</sup> This assumption is not the same as assuming that defendants have no prior criminal history at all; criminal history was a directly observable variable for sentenced cases and was used in the main sentencing specifications.

arrest offenses was homicide-related and no listed charge fell under a homicide statute, we assumed the defendant did not kill anyone (an aggravating factor in a large number of statutes in which death is a rare result, from violations of maritime rules to health care fraud). Similarly, we assumed that defendants in non-assaultive property or regulatory offense cases did not physically injure anyone.

When possible, we resolved ambiguities by reference to the other charges in the case, when the legal elements of those charges revealed additional facts that the prosecutor must have been alleging. In these cases we made exceptions to our default assumptions. For instance, suppose Charge 1 is a burglary offense that usually has a maximum sentence of 10 years, but has a 20-year maximum if someone is seriously injured in the course of the burglary. Charge 2 is an aggravated assault charge, with a 15-year maximum, in which aggravated assault is defined to require that serious injury be proven. In that case we would flag Charge 2 with a “serious injury” indicator, and that flag would trigger an enhancement to the coded statutory maximum for Charge 1, raising it to 20 years.

Implementing this approach required constructing a number of flags for every federal criminal statute. We constructed flags to indicate whether each of the following key facts were built into the required elements of the crime: death, injury, serious injury, drug crime, sex crime, fraud, official victim, minor victim, terrorist motive, an assault, use of a weapon, use of a gun specifically, a “crime of violence,” obstruction of justice, taking a person for ransom, and whether the crime was a predicate offense for the crime of felony murder. For each statute, we also indicated any adjustments to the statutory sentence that would be triggered by the presence of particular facts as identified by the flags for the other charges in the case. We followed this basic approach for each of the legal measures. Remaining ambiguities were resolved according to reasonable assumptions grounded in legal research concerning the most common ways in which statutes are charged. If significant ambiguity remained, we assumed there was *no* statutory minimum; many criminal statutes ordinarily have no minimum except when some

---

However, prior convictions of the exact same offense can safely be assumed to be exceptional even among defendants with criminal history; indeed, most prior offenses on the records of federal defendants are not federal crimes at all, but state crimes.

special circumstance is triggered.<sup>20</sup> In contrast, we erred on the high side when coding the statutory maximum in cases in which more than one was possible.<sup>21</sup>

Once the severity of each individual charge was coded, we then combined them into the overall statutory minimum and maximum for the case as a whole. Concurrent sentencing is the default federal rule for sentencing on multiple charges. This means that the total sentence is driven (often completely) by the most serious single charge. Therefore, unless the statute specifically required a consecutive sentence we assumed concurrent sentencing, such that the combined severity of the charges was the sentence associated with the charge carrying the highest sentence. When the statute specifically requires consecutive sentencing, the sentences for the charges carrying consecutive sentences were added to the one for the most serious concurrent charge. This approach could underestimate the true way a judge would combine sentences, because secondary charges can sometimes increase the sentence even when consecutive sentences are not required by statute, particularly if the charges are based on distinct incidents or patterns of conduct. However, such cases cannot be identified from the available data, so the default assumption is the best approximation. We did, as described above, use information drawn from secondary charges to adjust the requisite statutory sentences for the primary charge. Our approach to combining charges follows the method specified in the U.S. Sentencing Guidelines (see U.S.S.G. § 5G1.2).

For all cases other than declinations, we replaced zeros on the statutory maximum scale with half a month—half of the lowest nonzero values otherwise calculated—to reflect the fact that no criminal charge truly has zero severity, even if no incarceration is imposed. This adjustment affected only 0.04% of cases.

For magistrate cases that did not link through to AOUSC, we assumed the charges were petty, which means that the statutory maximum was coded by default as 6 months, and the statutory minimum was zero. The sentencing guidelines do not apply to petty offenses, so we

---

<sup>20</sup> A detailed spreadsheet showing these flags and the assumed base statutory sentencing range and Guidelines offense level for each federal crime is available on request.

<sup>21</sup> We chose to construct our own measure of the statutory maximum rather than use the existing AOUSC “severity” field, which is ostensibly based on the statutory maximum. The AOUSC coding appears to automatically be based on the very highest maximum contained *anywhere* in the statute cited, even when that maximum is only triggered by an exceptional circumstance that rarely applies. For instance, charges under 18 U.S.C. § 1347 (health care fraud) are coded by AOUSC as having a statutory maximum of life, even though that maximum only applies when the fraud leads to a death; the standard statutory maximum is ten years. Our approach therefore uses more realistic assumptions.

also coded the statutory maximum for these as 0.5 months, the lowest possible value for any charged case. Declinations (that is, cases with no charges) were coded with zeros for all charge measures.

After following the coding methods above, the statutory *minimum* for the combined charges in 88.6% of the cases in our sample was zero. We constructed a binary variable for whether any charge carried a nonzero statutory minimum, and this was the basis for our main analyses of the role of mandatory minimums. This avoided the need to resolve certain ambiguities in the AOUSC charge coding, because some statutes provide differing lengths of mandatory minimum depending on the facts of the case.

We coded the severity of the charges of *conviction* using the same measures described above, by combining the information for all terminal charges for which the disposition field indicated a conviction. On the statutory maximum scale, we coded as “true zeros” the charges that were dismissed or resulted in acquittal; the scale for convictions began at half a month, just as it did for the initial charge measures.

### **Data Sources:**

U.S. Census Bureau, 2000 Census of Population and Housing, Summary File 3. Available at: <http://www.census.gov/census2000/sumfile3.html>

United States Department of Justice. Office of Justice Programs. Bureau of Justice Statistics. Federal Justice Statistics Program: Paired-Agency Linked Files, 2009. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2011-11-11. ICPSR Study Number 30701-v3.

United States Department of Justice. Office of Justice Programs. Bureau of Justice Statistics. Federal Justice Statistics Program: Arrests and Bookings for Federal Offenses, 2001-2009. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. ICPSR Study Numbers 24126-v2 (2011-03-08), 24145-v2 (2011-03-08), 24164.v2 (2011-03-08), 24181.v2 (2011-03-08), 24216.v2 (2011-03-08), 24199.v2 (2011-03-08), 24211.v2 (2011-03-08), 24226.v2 (2011-03-08), 24231.v2 (2011-03-08), 29428.v2 (2011-03-08), 30794-v1 (2011-07-22). Original Data Source: U.S. Marshals’ Service (“USMS”).

United States Department of Justice. Bureau of Justice Statistics. Federal Justice Statistics Program: Suspects in Federal Criminal Matters Concluded, 2001-2009 [United States]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. ICPSR Study Numbers: 24120-v2 (2011-03-08), 24139.v2 (2011-03-08), 24158.v2 (2011-03-08), 24175.v2 (2011-03-08), 24193.v2 (2011-03-08), 24210.v2 (2011-03-08), 24225.v2 (2011-03-

08), 29424.v2 (2011-03-08), 30790.v1 (2011-06-03). Original Data Source: Executive Office of U.S. Attorneys (“EOUSA Matters Out”).

United States Department of Justice. Bureau of Justice Statistics. Federal Justice Statistics Program: Defendants Charged in Criminal Cases Filed in District Court, 2001-2009 [United States]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. ICPSR Study Numbers: 24121.v2 (2011-03-08), 24140.v2 (2011-03-08), 24159.v2 (2011-03-08), 24176.v2 (2011-03-08), 24194.v2 (2011-03-08), 24211.v2 (2011-03-08), 24226.v2 (2011-03-08), 29426.v2 (2011-03-08), 30791.v1 (2011-06-03). Original Data Source: Executive Office of U.S. Attorneys (“EOUSA Cases In”).

United States Department of Justice. Bureau of Justice Statistics. Federal Justice Statistics Program: Defendants in Federal Criminal Cases -- Terminated, 2001-2009 [United States]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. ICPSR Study Numbers: 24122.v2 (2011-03-08), 24141.v2 (2011-03-08), 24160.v2 (2011-03-08), 24177.v2 (2011-03-08), 24195.v2 (2011-03-08), 24212.v2 (2011-03-08), 24227.v2 (2011-03-08), 29433.v2 (2011-03-08), 30792.v1 (2011-06-03). Original Data Source: Executive Office of U.S. Attorneys (“EOUSA Cases Out”).

United States Department of Justice. Bureau of Justice Statistics. Federal Justice Statistics Program: Defendants in Federal Criminal Cases Filed in District Court, 2001-2009 [United States]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. ICPSR Study Numbers: 24114.v2 (2011-03-08), 24133.v2 (2011-03-08), 24152.v2 (2011-03-08), 24169.v2 (2011-03-08), 24186.v2 (2011-03-08), 24204.v2 (2011-03-08), 24221.v2 (2011-03-08), 29402.v2 (2011-03-08), 30781.v1 (2011-06-03). Original Data Source: Administrative Office of the U.S. Courts (“AOUSC Cases In”).

United States Department of Justice. Bureau of Justice Statistics. Federal Justice Statistics Program: Defendants in Federal Criminal Cases in District Court -- Terminated, 2001-2009 [United States]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. ICPSR Study Numbers: 24115.v2 (2011-03-08), 24134.v2 (2011-03-08), 24153.v2 (2011-03-08), 24170.v2 (2011-03-08), 24187.v2 (2011-03-08), 24205.v2 (2011-03-08), 24222.v2 (2011-03-08), 29242.v2 (2011-03-08), 30784.v1 (2011-06-03). Original Data Source: Administrative Office of the U.S. Courts (“AOUSC Cases Out”).

United States Department of Justice. Bureau of Justice Statistics. Federal Justice Statistics Program: Defendants Sentenced Under the Sentencing Reform Act, 2001-2009 [United States]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. ICPSR Study Numbers: 24127.v2 (2011-03-08), 24146.v2 (2011-03-08), 24165.v2 (2011-03-08), 24182.v3 (2011-03-08), 24200.v3 (2011-03-08), 24217.v3 (2011-03-08), 24232.v2 (2011-03-08), 29381.v2 (2011-03-08), 30795.v1 (2011-06-06). Original Data Source: U.S. Sentencing Commission (“USSC”).

United States Department of Justice. Office of Justice Programs. Federal Bureau of Investigation. Uniform Crime Reporting Program Data: County-Level Detailed Arrest and Offense Data, 2007-2009. Ann Arbor, MI: Inter-university Consortium for Political and Social

Research [distributor]. ICPSR Study Numbers: 30763-v1 (2012-01-25), 27644-v1 (2011-04-21), 25114-v1 (2009-07-31).