INSTITUTE FOR SOCIAL RESEARCH

New Mexico Statistical Analysis Center

Quality of the Department of Public Safety's Disposition Data

Prepared by:
Callie Dorsey
Kristine Denman

Formatting and editing support:
Ashleigh Maus

November 2019

This project was supported by Grant # 2017-BJ-CX-K018 from the State Justice Statistics program. The State Justice Statistics program is a component of the Office of Justice Programs, which also includes the Bureau of Justice Statistics, the National Institute of Justice, the Office of Juvenile Justice and Delinquency Prevention, and the Office for Victims of Crime. Points of view or opinions in this document are those of the author and do not represent the official position or policies of the United States Department of Justice.

Acknowledgements

We would like to thank the New Mexico Department of Public Safety for their willingness to provide these data to us and for their participation in this project. We would also like to thank the New Mexico Sentencing Commission (NMSC) who stores copies of the data used in this report, and provided portions of the data directly to us. We would also like to thank Linda Freeman, Executive Director for the NMSC, for her valuable comments as we drafted this report. Finally, we would like to thank the Bureau of Justice Statistics for providing the funding for this study through their State Justice Statistics Program.

Executive Summary

Introduction

DPS maintains criminal history data for individuals arrested in New Mexico including disposition information. However, the quality of this data can vary over time. We assessed the completeness of this disposition information in 2007 and found it is most often unknown. As such, we have not used DPS as a source of disposition information in our research or policy studies. Incomplete data has consequences beyond research: these data populate criminal history background checks and noncriminal background checks used for employment and licensing decisions. In this study we assess more recent disposition data from 2012-2016 for completeness, validity, uniqueness/duplication, consistency and accuracy.

Key strengths of DPS disposition data

- Validity is high.
    - Only 3% of charges with dispositions lack validity (have a code not included in the table of codes we received or have "See Remarks" with no corresponding information in the notes field).
    - Almost all (>99%) invalid dispositions result from a "See Remarks" disposition with no corresponding disposition information recorded in the notes field.
- Duplication is minimal.
    - When aggregated to SID, FBI number, date of arrest, ORI number and name, and offense description, the rate of duplication of charges is less than 1%.
        - Most commonly, multiple lines of data were due to discrepant information.
            - Most often, differences were due to varying case information (e.g., state tracking number, ORI case number) and personal identifiers (e.g., name, date of birth) along with some other difference.
    - Approximately 4% of incidents involve two or more arrests on the same day by different originating agencies.
- Overall consistency rates are very high.
    - Most arrest incidents (90%) with a corresponding court case have at least one matching disposition between the DPS data and automated Administrative Office of the Courts data.
    - 79% have all matching dispositions.
    - Consistency across data sources increases over time; in 2012, 76% of charges matched and in 2016, 87% matched.
- Accuracy rates are high.
    - Based on a random sample of 50 cases with inconsistencies between DPS and AOC, very few DPS recorded dispositions (<1%) were actually wrong.
        - Inconsistencies typically occur for two reasons:
            - Coding of the disposition varies by source, most often in cases involving a conditional discharge or deferral.
            - The data are dynamic: charges and dispositions change over time.

Key weaknesses of DPS disposition data

- Completeness rates are low.
    - The rate of charges with unknown disposition is higher now (83%) than the last time we assessed this data (77%).
    - Based on a small random sample, approximately 53% of arrest incidents with unknown dispositions had been disposed by the court.
- Specific consistency varies.
    - Consistency of dispositions from DPS and AOC data ranged across counties from a low of 45% of cases with one or more matching dispositions to a high of 100%.
    - The recording of conditional discharges and deferred sentences varies.
        - DPS sometimes records these as "guilty plea;" this may be due to the way DPS receives the data and not the result of a data entry error.
        - AOC sometimes includes these as a "disposition" but not always.

Conclusions and suggestions for improvement

Overall, the quality of the disposition data from DPS, when populated, is good. The primary deficit is the amount of missing information. This is an ongoing problem with these data, and the results of this analysis indicate that the situation has not improved over time. Importantly, New Mexico is embarking on an effort to integrate data systems. We anticipate that successful integration would result in the population of disposition data.

Besides completion, we found several potential areas for improvement. First, though the disposition type "See remarks" is a valid code, the information is missing or the notes do not include disposition and/or sentence information 45% of the time. In some cases, the information populated reflects existing codes, and could have been populated there rather than in the notes. In these cases, storing and updating disposition information in the correct disposition variable instead of the notes variable could increase completeness and accuracy of the data. Importantly, the notes/remarks field also captures charge resolutions that do not fit in the standard codes used by DPS. For example, DPS uses this field to capture pleas to a lesser charge, case transfers, offender extradition, and fines. Adding codes for these categories could help increase the amount of valid dispositions.

Second, the SID number is sometimes associated with multiple FBI numbers on the same arrest date. While this accounts for a very small percentage of cases overall (.01%) it is important to be aware that this happens, especially as the state moves towards an integrated system. The overall number of instances (N=411) is rather minute, so while it would take some effort to resolve these discrepancies, it would likely not involve extensive resources to do so.

Finally, we found some inconsistencies within the data. First, the recording of conditional discharges and deferred sentences is inconsistent both within the DPS and AOC datasets. This problem may not originate with DPS since they receive the data from different sources. Furthermore, in a separate assessment of the AOC data, we found inconsistency with recording these outcomes was a common issue within the AOC data. This is a particularly important issue for those cases where the judge has granted a conditional discharge, since there is no adjudication of guilt. Second, the accuracy of dispositions varied by county. Further investigation into the source of these inconsistencies could improve the overall accuracy of the DPS disposition data.

Introduction

We rely on information populated in administrative data for a variety of purposes, including informing procedures, studying outcomes, and guiding policies. However, the quality of these data can vary over time, by jurisdiction, and by the specific data fields. Given the implications that these data can have, it is imperative that we assess the quality and completeness of the administrative data we, and others, access to conduct our research and make policy recommendations.

We use data from the New Mexico Department of Public Safety (DPS) often in our research. DPS maintains criminal history data for individuals arrested in New Mexico. One field available in that data is case disposition. DPS receives disposition information from the Administrative Office of the Courts and the Administrative Office of the District Attorneys, as required by the statute (NMSA 29-3-8 §F and §G). DPS' IT staff manually enter the information, and may supplement with additional information provided by other agencies.

As part of an assessment of data completeness conducted in 2007, we found that while the disposition field was always populated, it was most often (77%) populated with "unknown." Thus, we have not used DPS data as a source for the disposition of cases for research or policy studies. Incomplete data has real-world consequences. In 2016, DPS performed an estimated 12,100 noncriminal justice background checks to inform licensing and employment decisions (Goggins & Debacco, 2016: Table 14). Without conviction information, these background checks can only indicate arrests (accusations), not determinations of guilt. However, the extent to which the data are missing currently is unknown. The current study seeks to address this concern through an assessment of DPS' disposition data.

Sample

DPS provided five datasets with arrests that occurred between 2012 and 2016, with each dataset corresponding to a year. The variables include: the FBI number; SID number; date of arrest; ORI number and name; ORI Case number; state tracking number; arrest location; offender information (dob, race, sex, ethnicity, name, last four ssn); flags for certain characteristics (sex offender, alcohol related, drug related, domestic violence related): charge code and description of charge; disposition, and notes related to the disposition/sentence. DPS has provided a copy of this information except dispositions to the NMSAC and the New Mexico Sentencing Commission on a quarterly basis for many years. They graciously added the disposition data for this study. In addition to the arrests that occurred in or after 2012, DPS staff enter arrest data from prior years from hardcopy fingerprint cards, typically as people acquire new arrests. Thus, the earliest arrest date in the sample dates back to 1930.

The NMSC stores all data received in a data warehouse. When they receive new data from DPS, NMSC staff append the new data to the data already in the warehouse. The data include all arrests entered in that quarter, regardless of arrest date. In addition to the original variables from DPS, NMSC adds some variables to standardize portions of the data (e.g., collapsing crime type into broader categories).

Methods

For this project, we focused on assessing the quality of the disposition data in the DPS dataset. We assessed the following five measures:

1. How *complete* is the information?
2. How *valid* is the information?
3. To what extent is there *duplication* of records?
4. How *accurate* are the records?
5. How *consistent* is the information?

Completeness focuses on the extent to which the disposition data are missing from charges in the dataset. Validity assesses data entry errors, examining whether the codes recorded are valid codes used by DPS, as well as assessing the proportion of cases with values outside the expected range. We calculated duplication, or uniqueness, by calculating the rate of duplicate records. We used all available years of data to check for *completeness, validity*, and *duplication* of records.

For the last two measures, *accuracy* and *consistency*, we limited the data to the years 2012 to 2016. We compare DPS disposition data with dispositions in automated data from the Administrative Office of the Courts (AOC). In cases where these do not match, we assess accuracy by examining hardcopy J&S documents. Besides comparing dispositions across data sources, we assessed consistency by looking at whether the above measures (completeness, validity, uniqueness, and accuracy) vary over time and by location.

Since each line of data from DPS corresponds to a charge on a given date by arresting agency, there may be multiple lines of data for a single incident. Thus, we analyze the data in multiple ways. First, we analyze the data by charge; this includes every line of data. Second, we analyze the data by incident and arresting agency. Individuals arrested on the same day by two or more different agencies would have two or more lines of data for that date. We found that 4.1% of incidents had charges from two or more agencies on the same date between 2012 and 2016, and 3.6% overall. While it is possible that these arrests arise from unique incidents, these typically refer to the same incident. The two most common scenarios are that an individual violates probation due to a new offense and as such gets fingerprinted by parole/probation and law enforcement or that an individual commits a new offense while detained and as such is fingerprinted by a detention center/corrections department and law enforcement. Finally, we analyzed this data by all charges for an offender on a given date regardless of differing ORI's. For each different assessment, we calculated the overall missing disposition rate for all years of data available, then limited it to 2012 to 2016, then annually from 2012 forward.

We merged AOC data with the DPS arrest data. Finding the correct match between a particular arrest in the DPS data and the corresponding court case in the AOC can be difficult. We utilized a few methods to find the correct match. First, we limited the data to those that have a disposition populated in the DPS data, then we began to match cases. Next, we searched for matches by individual. Since we have merged DPS and AOC data for other projects, we had some matches already identified for calendar years 2012 and 2013. We flagged those matches. For the remaining cases, we matched individuals using various combinations of personal identifiers (name, date of birth, last four digits of SSN). Once we made a good match on an individual, we identified matches between arrests and corresponding court cases. To link an arrest to a particular court case, we matched by the offense date recorded in the AOC data and the arrest date in the DPS data, which are typically the same or very close (within a few days).

We were able to find a corresponding court case for 19.2% (n=16,635) of cases between 2012 and 2016 with DPS dispositions. However, it is important to note that some of the cases with no corresponding court case found genuinely have no corresponding court case. For example, charges may be

downgraded to misdemeanors and not go through district court, the case may not end up in court at all, or the charges came from a previous ongoing court case. In order to further explore why we were not able to find corresponding court cases for all DPS cases with dispositions, we selected a random sample of 30 relevant cases and searched Odyssey to determine a reason we were unable to find a court case. For 76% (n=23) of the charges checked we were able to find a reason for the inability to match with a court case. The most common reason accounting for 30% (n=9) of the cases in our sample was that the cases were not tried in District court, either because the charges were lowered to a misdemeanor or because the felony charges ended at the Magistrate level through plea or dismissal. A handful (13.33% n=4) of the cases in our sample have a court charge date that is before the DPS arrest date and therefore were not picked up in our merges that look for a charge date on or after the arrest date. Some (16.67% n=5) of the charges were not new charges, but were connected with ongoing court cases involving other charges. For example, being held in contempt of court or a failure to follow requirements of a conditional discharge. Next, 16.67% (n=5) of the charges have a corresponding court case but this court case was not in the AOC court dataset we used to merge. For one case we were able to find a court case by hand but unable to establish a reason this case was not matched automatically. For the remaining 20.00% (n=6) of cases we were unable to find a corresponding court case in Odyssey.

Finally, using the AOC and DPS data, we assessed whether the disposition populated in the DPS dataset was accurate to the automated AOC data. While most of dispositions are merged into the DPS data from AOC, some are manually entered which can lead discrepancies between the DPS and AOC dispositions. Given the complexity of matching each charge in the DPS data with the corresponding charge in the AOC data, we opted to assess consistency by aggregating each set of data to the incident level. We then created dichotomous variables which indicate whether a given disposition was populated for that incident and by which data source. For example, when a disposition of convicted sent was found for any charge in the AOC data, we created a variable called “convictedsent_AOC” and coded it as “1.” A disposition of convicted sent in the DPS data for that incident would be considered a match.

We checked a sample of cases with discrepancies between DPS dispositions and AOC court dispositions against hardcopy J&S documents. We selected a random sample of 50 cases with disparate dispositions to assess the source of the disparity (AOC or DPS). This allowed us to further assess accuracy of these data.

We used descriptive and bivariate statistics to summarize the five aforementioned measures. We checked for completeness by calculating the rate of missing data in each dataset, and validity by assessing the proportion that have values outside the expected range. We tested for uniqueness by calculating the rate of duplicate records. We also provide a rate of accuracy, which was based on the proportion of disposition information that matches between the AOC data and the DPS data and whether this varies by jurisdiction and over time. We further looked at accuracy through our assessment of the hardcopy J&S documents. We assessed consistency by looking at whether the above measures (completeness, validity, uniqueness, and accuracy) vary over time.

Results

Completeness of disposition data

The first objective of this study is to assess the completeness of the disposition data in the DPS dataset. As described in the methods section, we analyzed the data in multiple ways (by charge, by date and agency, and by date). While the overall numbers differ, the rates are approximately the same and the trends are the same. Thus, we present only the data by charge in Table 1 below. Appendix A illustrates the annual completeness rates by incident date and agency, and by date. As can be seen below, when we analyze the data by charge, we find 83% of charges in the entire dataset do not have a disposition. This increases to 86% when we limit the data to 2012 forward. As may be expected, the proportion of missing data increases each year. This may reflect the fact that some cases may not yet have been disposed, and therefore do not have a disposition. The rate of missing dispositions is notably higher than we found in the past. As noted previously, in our prior study (Denman, Broidy, Wadsworth, and Albright, 2007) which included 2001 to 2006 data, we found a missing rate of 77%; the overall rate for 2012 to 2016 is now 86%.

Table 1: Disposition completeness by year

Disposition	All available years (1930 to 2016)	2012 to 2016	2012	2013	2014	2015	2016
No Disposition Populated	83.73%	86.31%	81.37%	84.48%	86.44%	88.82%	92.81%
Disposition Populated	16.27%	13.69%	18.63%	15.52%	13.56%	11.18%	7.19%
Total N	750,200	612,553	134,645	140,887	132,232	106,874	97,915

Since such a large proportion of charges did not include a disposition, we selected a random sample of 30 felony level arrests without dispositions and checked for corresponding court cases. We were able to find a corresponding court case with pre-2016 disposition information for 53% (n=16) of the cases we checked. We were able to find a further 16% (n=5) post-2016 dispositions but these would not have been included in the data we received. For the remaining 30% (n=9) we were unable to find a corresponding court case in order to check for disposition information, and therefore likely did not result in a court case. These results suggest that the disposition data may be available for up to 50% more cases than are currently populated.

Validity of information

Next, we examine the validity of the disposition information populated. DPS provided us with a list of disposition codes they use (see Appendix A). Very few charges included a disposition code that was not included in the disposition table. Among cases with a disposition, less than .05% have a disposition that is not included on the code list. However, though the disposition type "See remarks" is a valid code, the notes sometimes are not related to the disposition and/or sentence. Instead, this field sometimes contains corrected personal identifiers, corrected charge descriptions, corrected charge levels, social

security numbers, and notifications that the individual has multiple arrest cards. Sometimes information is simply missing from the notes field. Therefore, we also examined the data by whether or not the notes indicated a valid disposition. When we include that information, we found that 3% of charges between 2012 and 2016 had an invalid disposition. This proportion varied over time with a low of 2% in 2014 and a high of 4% in 2013.

Table 2: Validity of disposition by year

	Disposition	All available years (1930 to 2016)	2012 to 2016	2012	2013	2014	2015	2016
Valid dispositions	By category	94.33%	93.66%	92.59%	95.31%	95.02%	93.08%	89.94%
	By remarks	2.93%	3.22%	4.04%	0.47%	3.26%	4.20%	7.06%
Total valid dispositions		97.26%	96.88%	96.63%	95.78%	98.28%	97.28%	97.00%
Invalid dispositions	By category	0.03%	0.03%	0.03%	0.02%	0.06%	0.02%	0.06%
	By remarks	2.72%	3.09%	3.34%	4.20%	1.67%	2.69%	2.95%
Total invalid dispositions		2.74%	3.12%	3.37%	4.22%	1.72%	2.72%	3.00%
N		124,056	85,751	25,718	22,537	18,016	12,216	7,264

Duplication of information

Our third measure of data quality is record duplication. We checked for duplication first by determining whether the variables we expect to be unduplicated were, indeed, unduplicated. These include SID, FBI number, ORI number, ORI name, date of arrest, and offense description. We aggregated to these fields and chose the minimum and maximum of all remaining variables. This resulted in 743,203 lines of data. We then checked whether there was duplication by alternately omitting ORI name and number, FBI number, and SID. By doing this, we found that there was duplication on these primary keys only by SID. This occurred in just .01% of cases (n=411). In other words, different SIDs were associated with the same arrest incident in a small proportion of cases. This also means that in these cases, the individual for that arrest incident had two or more SIDs listed with a single FBI number.

Next, we checked for duplicated charges using SID, FBI number, date of arrest, ORI number and name, and offense description. Although some of the SID's were duplicated, not every charge has an FBI number (2.5% n=18,754) so it is necessary to use a combination of both in order to avoid false duplications. We found that .6% (n= 3,917) of the charges between 2012 and 2016 were duplicated on the variables above. Next, we aggregated these duplicated charges together with minimum and maximum values for all other variables in order to assess whether the charges were perfect duplications or contained discrepant information.

SAC staff grouped DPS variables into categories: personal identifiers, flags, and case information. The personal identifier category is comprised of the last four digits of the Social Security number, first name, middle name, last name, and date of birth. While no charges between 2012 and 2016 were discrepant only on personal identifier information, 69% (n=816) of the charges with multiple discrepant categories include a discrepancy between personal identifiers. In other words, personal identifiers along with a mismatch in some other category accounted for the majority of duplicated data.

The flag category is comprised of a drug flag, alcohol flag, and a domestic violence flag. All of the discrepancies in the drug and alcohol flags indicate a genuine discrepancy between a "yes" flag and a "no" flag. However, 83% of discrepancies for domestic violence flags are between a "no" flag and missing information. While the missing information might mean the charge has no flag we have no way of knowing this for sure.

The case information category is comprised of ORI case number, state tracking number, agency type, charge code, disposition, and DPS notes. The most common discrepancy is this category originated with the state tracking number and the ORI case number. Some charges have discrepancies in multiple categories. Finally, 11% of charges overall include true duplications with no discrepant information. This could mean that there are multiple lines of data for the same charge (a data entry error). Conversely, it may be that some individuals had multiple counts of the same charge and all of the other data was the same for each charge. These results are summarized in Table 3 below.

Table 3: Source of discrepant information among duplicated cases by year

	2012-2016	2012	2013	2014	2015	2016
Discrepancy source:	%	%	%	%	%	%
Flags	8.48%	15.79%	6.48%	5.00%	5.09%	5.93%
Case Information	50.40%	50.93%	48.86%	44.76%	55.19%	60.17%
Multiple Categories	30.10%	22.90%	32.57%	36.19%	29.16%	30.79%
No Discrepancies/duplication only	11.03%	10.37%	12.08%	14.05%	10.57%	3.11%
Total N	3,917	1,070	1,142	840	511	354

Finally, there are times when different agencies arrest and/or fingerprint an individual on the same day. We assessed how often this occurs. We examined the data at the case level and checked for duplication by SID, FBI number, and arrest date. We found that 3.6% (n=16,021) of cases have multiple incidents on the same day involving different originating agencies. These incidents could result when an individual an agency arrests an individual then transfers the person to another agency, or less likely, the person is involved in two completely separate arrest incidents on the same day. Although all the same day cases have different originating agencies not all have different *types* of originating agencies (e.g., Probation and Parole vs. State police). Most (79% n=12,665) cases have different types of originating agencies but some (20.9% n=3,356) of the cases have multiple originating agencies that are the same type (e.g., Santa Fe Police Department and Los Alamos Police Department). These results are summarized in Table 4 below and expanded upon in Appendix C.

Table 4: Duplicated cases by originating agencies

Originating Agency	Percent
Multiple Law Enforcement	18.25%
Multiple Probation/Parole Agencies	1.45%
Multiple Correction Facilities	1.25%
Different Agencies	79.05%
TOTAL	16,021

Consistency of information

We then compared the consistency of the disposition information found in DPS to that recorded in AOC. We limited the data to DPS incidents between 2012 and 2016 with valid dispositions for which we found a corresponding court case also with disposition data (n=14,542). We calculated a rate of matching dispositions using the incident as the unit of analysis rather than the charge. Since a single case can have multiple charges we calculated a rate of matching that ranges from 0 for no dispositions match to 1 for all dispositions match. Most (89.67%, n=13,040) cases with a corresponding court case have at least one matching disposition, and a majority (78.9%, n=11,473) of these cases have all matching dispositions. Just over 10% (n=1,502) of cases with a corresponding court case have no matching dispositions.

However, these results are not stagnant across years and counties. First, there is a general trend towards more consistent disposition information over time. Particularly notable is the increase in the proportion of perfect matches, from 76% in 2012 to 87% in 2016. Table 5 below illustrates this trend.

Table 5: Rate of court case with matching dispositions by year

	2012	2013	2014	2015	2016	2012-2016
All Dispositions Match	76.21%	79.87%	78.66%	80.31%	87.15%	78.90%
Some Dispositions Match	13.02%	10.41%	10.98%	7.83%	7.94%	10.78%
No Dispositions Match	10.77%	9.72%	10.35%	11.86%	4.91%	10.33%
N	3,779	4,526	3,651	2,158	428	14,542

Next, we analyzed consistency across counties. We found a much wider variation in the rate of consistency across counties ranging from a low of 42% of cases with all matching dispositions up to 100% matching. However, the number of cases per county varies and clearly contributes to these rates: all of the counties with entirely accurate case dispositions had fewer than 10 cases total. Table 6 details these results.

Table 6: Rate of court case with matching dispositions by county

County	All Dispositions Match	Some Dispositions Match	No Dispositions Match	N
Mora	100.00%	0.00%	0.00%	2
Union	100.00%	0.00%	0.00%	4
Unknown	100.00%	0.00%	0.00%	6
Roosevelt	92.81%	4.32%	2.88%	139
Eddy	92.61%	5.00%	2.39%	460
De Baca	91.67%	8.33%	0.00%	12
Sierra	87.50%	12.50%	0.00%	8
Luna	86.82%	5.73%	7.45%	349
Guadalupe	85.71%	14.29%	0.00%	14

Lea	84.77%	12.30%	2.93%	512
Cibola	84.51%	0.00%	15.49%	142
McKinley	84.04%	3.19%	12.77%	188
Curry	83.58%	6.80%	9.62%	603
Bernalillo	82.46%	13.60%	3.94%	5,580
Chaves	82.30%	12.92%	4.78%	565
Socorro	82.22%	4.44%	13.33%	90
San Juan	81.56%	12.94%	5.50%	1,437
Rio Arriba	80.94%	9.63%	9.43%	509
Grant	80.59%	12.94%	6.47%	170
San Miguel	80.58%	14.03%	5.40%	278
Otero	78.68%	10.00%	11.32%	530
Quay	77.98%	12.84%	9.17%	109
Torrance	75.86%	0.00%	24.14%	58
Santa Fe	75.42%	12.79%	11.79%	603
Lincoln	69.75%	21.85%	8.40%	119
Valencia	69.54%	5.23%	25.23%	325
Colfax	68.00%	4.00%	28.00%	150
Dona Ana	63.10%	5.35%	31.55%	729
Hidalgo	62.07%	24.14%	13.79%	29
Los Alamos	61.11%	30.56%	8.33%	36
Taos	56.84%	3.16%	40.00%	190
Sandoval	42.45%	2.52%	55.03%	596
Total				14,542

Accuracy of information

Given that we found inconsistencies when comparing the data sources, we next explored the accuracy of the data in order to assess potential causes of the discrepancy between DPS dispositions and the court dispositions. We created a random sample of 40 cases with no matching dispositions and 10 cases with at least one unmatched disposition for analysis. Next, we checked these 50 cases against hard copy J&S documents and AOC's online court case entries to ascertain a reason and a source of disparity. It is important to note that while we name a source of the disparity, it often does not mean that the data are wrong. Rather, it is often a reflection of the nature of comparing diverse datasets.

The majority of discrepancies found in the sample (60%, n=30) is due to differences in the way DPS records, codes, or tracks information relative to the data we receive from AOC. In these cases, the actual information matches but the way DPS houses/receives the information inhibits direct matching. The most common cause of discrepancy (n=12) between cases was because the cases had initial charges that were either dismissed or lowered before proceeding to District Court. DPS tracks disposition data by charge, so when a charge is changed, DPS records "dismissed" as the disposition but there is information about the altered charge in a note. Since the AOC data included only District Court cases, the original (dismissed) charges were not in the dataset. Rather, the new charge appears with the associated disposition. While this was the more common scenario, sometimes the prosecutor added

charges that did not appear initially. These are not present in the DPS data for that arrest date since they were not the original charges.

A second challenge arises from the way DPS and AOC record the disposition versus the sentence. In nine cases, DPS recorded the plea as a disposition while AOC has the sentence that results from said plea. Specifically, if an individual pleads guilty and the judge orders a deferred sentence or conditional discharge, the DPS data sometimes records guilty plea as the disposition while the court has "deferred" or "conditional discharge" as the disposition. Importantly, DPS does use the codes "deferred" and "conditional discharge," so we consider this an erroneous entry since those options are available. Our assessment indicates that in the remaining eight cases in this category, the DPS data were erroneous. In four of these cases, the disposition changed but not updated in DPS. For example, in one case, the court amended the disposition from "deferred" to "convicted," but DPS had the "deferred" disposition. Note, though, that DPS receives disposition data from the Administrative Office of District Attorneys as well as the AOC, and this discrepancy may be the result of the source of information rather than DPS. One case had both a charge mismatch as well as a conditional discharge coding error. However, DPS may have updated this information after we received the data. In another four cases, there was truly an error in the recording of the disposition. For example, one case has "See Remarks" as a disposition but no note with disposition information and the rest of the charges have "unknown" dispositions. In another case, prosecutors filed two cases at the district level with identical charges for the same incident. One case was dismissed altogether since it was duplicated. DPS has "dismissed" as the disposition instead of the ultimate disposition from the second District Court case.

In six cases, we attributed the discrepancy to the AOC data. Four have mismatched dispositions because of AOC coding issues. In three of these cases, the court convicted the individual, but the judge deferred the sentence or ordered a conditional discharge. However, the disposition in the automated data is "convicted." The reason this occurs is that the NMSC receives only charges with a "disposition." In these cases, court clerks entered the conditional discharge or deferral under the titles "conditional discharge" or "deferral" rather than "disposition." Therefore, the dispositions were not in the database we received. Importantly, court clerks often record deferred sentences and conditional sentences under the title "disposition." Thus, there is inconsistency in the way court clerks enter this information. In the last case, it appeared that court clerks updated the charge information and disposition, so we received two lines of data for the same charge. The last two cases that we attribute to the AOC did not have updated dispositions. However, we expect that subsequent data pulls will have that information.

In eight cases, we attribute the differences to both AOC and DPS. One source of discrepancy in this category is simply a difference in options for recording dispositions across datasets (n=6). In five of these cases, the judge allowed the defendant to participate in a pre-prosecution diversion program. DPS has a disposition category for entering a pre-prosecution program that automatically couples with a dismissed sentence (pre-prosecution/dismissed). On the other hand, the AOC records these as either deferred or conditional discharges with the requirement that an offender attend a program, consistent with the language in the J&S documents. In another case, the DPS recorded "dismissed with prejudice" as an acquittal whereas the AOC records this as "dismissed." Besides coding mismatches, we found two cases where there was a coding issue originating with DPS but the AOC data was either wrong or missing. In one of these cases, the charges changed over time (thus, we attribute that issue to DPS data), but the automated data from AOC also does not include all of the charges (i.e., those that were dismissed). In the last case, DPS recorded the disposition in the notes rather than noting it as a

conviction. In this same case, the automated AOC data are wrong: there are three charges in the case, all of which resulted in an acquittal but the automated data show these all as convictions.

Finally, 12% (n=6) of the cases are mismatched with neither DPS nor AOC being the source of the error. In all of these cases, multiple court cases are associated with the same individual and we matched the wrong case. For all but one case, the arrest date was the same in each case. In two cases, we should have determined that we had identified the wrong case because the charges differed. However, as noted above, charges change over time: the initial arrest charges may not be the same charges the prosecutor pursues. In another mismatch by us, prosecutors filed two cases with nearly identical charges, but one ended in magistrate court and the second proceeded to District Court.

Overall, true errors were minimal. One case involved an error originating in the AOC data. Four cases had either missing or wrong information originating from the DPS dataset. In an additional ten cases, the disposition indicated the defendant was guilty, but the sentence of conditional discharge or deferred was not reflected in the DPS data. While technically wrong, there was some finding of culpability in these cases, but the sentence was incorrectly recorded. Generally, the most common reason for the inconsistency across data sources is not error, but a difference in the way the agencies record the data (n=19). The second reason for inconsistency is due to the dynamic nature of the data: the charges (n=12) or disposition (n=6) change over time. Table 7 below summarizes the results.

Table 7: Source discrepant dispositions

Source	Discrepancy reason	n	N	%
DPS			**30**	**60%**
	Coding of conditional discharge	9		
	Charge mismatch	12		
	Error- missing or wrong information	4		
	Not the final disposition	4		
	Multiple reasons	1		
AOC			**6**	**12%**
	Coding	4		
	Not the final disposition	2		
Both			**8**	**16%**
	DPS coding, AOC error	1		
	DPS coding, AOC missing	1		
	Coding mismatch	6		
Neither			**6**	**12%**
	Wrong case	6		
Total		50	50	100%

Conclusion and suggestions for improvement

The purpose of this study was to assess the quality of disposition data in the DPS dataset, using a variety of measures. DPS receives hard copy disposition data from two sources: the Administrative Office of the Courts and the Administrative Office of District Attorneys. DPS staff then manually enters the data.

The first measure we examined was completeness. Since the NMSAC last assessed the DPS data in 2007, we found that the rate of incomplete data has increased rather than decreased. Between 2001 and 2006, 77% of the charges had dispositions that were unknown (Denman et al.). The current analyses indicate that the unknown/missing rate has increased to 86% for the years 2012 to 2016. Importantly this missing rate is extremely high in relation to the rest of the country as well. Of the 42 states and territories that reported their rate of felony charges with final dispositions in 2016, New Mexico has the 4th lowest rate of reported final dispositions at 25% (Goggins & DeBacco, 2018: Table 1).

However, DPS staff indicated that they did not expect many cases to have disposition information. Importantly, some of these charges will not have any disposition associated with them. When we checked a random sample (n=30), we found disposition information from the online secured Odyssey system for 53% of the sample. Some cases, then, were likely disposed at the municipal level or never forwarded to the ADA's office for prosecution. Delays also play a role. In New Mexico the elapsed time between the occurrence of a final felony court case disposition and its receipt by the data repository is 31-90 days (Goggins & DeBacco, 2018: Table8). After this initial wait, it takes on average a further 181-365 days to enter this information into the state's criminal history record database (ibid). This delay likely contributes to the missing disposition information observed, as even after the disposition event has happened it could take over a year for this information to make its way into the DPS data.

The second data quality measure was validity. We found that very few (less than 3%) of the dispositions lacked face validity. Among those that were invalid, the vast majority (99%) are associated with the "See Remarks" disposition. In these cases, while the disposition is "See Remarks," there is no corresponding disposition information in the notes/remarks field. Revisiting the use of "See Remarks" as a disposition could greatly reduce the rate of invalid disposition information. Importantly, the notes/remarks field captures the resolution of charges that do not fit in the standard codes used by DPS. For example, DPS uses this field to capture pleas to a lesser charge, case transfers, offender extradition, and fines. Adding codes for these categories could help increase the amount of valid dispositions.

Third, we explored the extent of duplication within the records. Less than 1% of the arrest incidents had any duplication when examined by SID, FBI number, arrest date, ORI number and agency, and charge. True duplication occurred in only 11% of the cases in which we flagged a duplication. This is likely an overestimate as some people will have multiple counts of the same charge, and these would count as duplicated here. Thus, much of this analysis illuminates the extent to which there is mismatched information rather than actual duplication. For example, there were duplicates because the state tracking number or ORI case number differed from line to line within an arrest incident. Recall that the NMSC receives quarterly updates and appends that new information to the old data. It is likely that some of the duplication we detected is due to updated/revised information.

In addition to checking the data by SID, FBI number, arrest date, ORI number and agency, and charge, we also checked for duplications arising from different agencies. Notably, approximately 4% of cases have multiple arrests on the same day by different originating agencies. This could occur when one

agency transfers an individual to another agency. While not common, it could also occur when someone is arrested for two separate incidents on the same day.

Thus, duplication is relatively uncommon (less than 1%) and typically arises from mismatched/discrepant information provided on two lines of data. It is more common to have multiple arrest incidents on the same day for the same charges due to different arresting agencies (4%). In a very small number of cases (n=411, .01%), we found different SIDs attributed to the same person (the same FBI number was populated for both). This has implications for the planned data-sharing network mandated by House Bill 267. The state will need to be able to link records from disparate systems together and these duplications could hinder this process.

Fourth, we examined consistency between DPS dispositions and court data. For this assessment, we compared the disposition recorded in the DPS data to that from an automated dataset originating with the courts. Consistency rates were high; approximately 90% of incidents included one or more matching disposition and in nearly 80%, all dispositions matched. Furthermore, we found consistency has been improving every year from 2012 to 2016. However, we also found significant differences by county. The average rate of cases with no matching dispositions with the court data is 10%. Five out of the 31 counties included in the dataset have rates of inconsistency above 25%, and in one county, the majority (55%) of cases have dispositions that are different from the court data. Further investigation into why some counties have relatively high rates of inaccuracy could greatly improve the overall consistency of the DPS disposition data.

Another source of inconsistency is likely due to the dynamic nature of these data. For example, in a court case where a person pleads guilty and is convicted and sentenced, the DPS disposition field might contain either “guilty plea” or “convicted sent(enced).” This is likely the result of the timing of the receipt of the information. If the sentence is not known, then “guilty plea” would be recorded; if the sentence is known, then “convicted sent” would be recorded. Similarly, a judge-ordered conditional discharge may be recorded as a “guilty plea,” a “conditional discharge,” or “conditional/dismissed” upon successful completion of the sentence.

Finally, we examined accuracy. We undertook this analysis in order to understand the reason for the discrepancies observed between DPS and the court dispositions. In examining a random sample of 50 cases with disparities between DPS and AOC, we found that very few (less than 1%) of the DPS recorded dispositions were actually incorrect. A common source of disparity was the recording of dispositions in cases where the judge deferred the sentence or ordered a conditional discharge. This is not a straightforward disposition as the ultimate disposition will change. The charges can be dismissed pending successful completion of the terms of the agreement or the opposite could occur if the defendant violates the terms of the agreement. As evidenced here, both DPS and AOC record this outcome in an inconsistent way. DPS sometimes uses the disposition of “conditional discharge” but not always. AOC may record this under the “disposition” category or under a separate entry labeled “conditional discharge” or “deferred.” If it is the latter, the dataset the NMSC receives will not include that disposition and will only include the plea. Furthermore, some of the mismatch between DPS and AOC dispositions occurs unnecessarily because the correct DPS disposition information is stored in the notes variable. Of the charges between 2012 and 2016 that have valid dispositions, 53% (n=43,700) have notes. Updating the disposition field with corrected information instead of recording this information in the notes could increase the accuracy of the disposition data.

The purpose of this study was to assess the quality of DPS disposition data between 2012 and 2016. To this end, we identified several key strengths and weaknesses of the data. The greatest weakness identified is that the majority (83%) of charges do not have a disposition associated with them and that this rate has increased since 2007. Furthermore, the reported rate of felony dispositions for New Mexico is well below the average for the other states and territories. However, our findings indicate that when charges do have dispositions populated, very few are invalid and accuracy is high. Moreover, the consistency of the DPS disposition data with the AOC data is quite high, and has improved every year.

Appendices

Appendix A: Completeness by incident and date

Table A.1 Completeness at the incident level by year

	Disposition	All available years (1930 to 2016)	2012 to 2016	2012	2013	2014	2015	2016
No Dispositions by Incident	No Dispositions Populated	83.15%	85.12%	80.22%	83.46%	85.49%	87.08%	91.36%
	All See Remarks (Without Disposition Information)	.42%	.47%	.93%	.63%	.32%	.27%	.07%
No Disposition Total		83.57%	85.59%	81.15%	84.09%	85.81%	87.35%	91.43%
Dispositions by Incident	At Least One Valid Disposition	15.9%	14.04%	18.22%	15.85%	13.84%	12.29%	8.11%
	See Remarks (With Disposition Information)	.53%	.36%	.63%	.06%	.35%	.36%	.46%
Disposition Total		16.43%	14.41%	18.85%	15.91%	14.19%	12.65%	8.57%
N		451,169	364,442	80,038	82,572	77,419	64,756	59,657

Table A.2 Completeness by date and year

	Disposition	All available years (1930 to 2016)	2012 to 2016	2012	2013	2014	2015	2016
No Dispositions by Date	No Dispositions Populated	82.86%	84.84%	79.85%	82.91%	85.08%	86.92%	91.29%
	All See Remarks (Without Disposition Information)	.5%	.47%	.94%	.63%	.32%	.27%	.07%
No Disposition Total		83.36%	85.32%	80.79%	83.53%	85.4%	87.19%	91.36%
Dispositions by Date	At Least One Valid Disposition	16.18%	14.32%	18.58%	16.41%	14.26%	12.46%	8.18%
	See Remarks (With Disposition Information)	.46%	.36%	.63%	.05%	.33%	.35%	.46%
Disposition Total		16.64%	14.68%	19.21%	16.47%	14.6%	12.81%	8.64%
N		434,836	349,465	75,677	77,909	73,817	62,995	59,067

Appendix B: Dispositions

Table B.1 Dispositions by year

Disposition	All	% 2012	2012	2013	2014	2015	2016
ACQUITTED	0.10%	0.10%	0.03%	0.07%	0.16%	0.22%	0.08%
ADJUDICATED JUVENILE	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APPEALED	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CLEARED BY ARREST	4.10%	4.10%	3.45%	3.67%	4.56%	4.08%	6.59%
CONSENT DECREE COMP/DISMISSED	0.01%	0.01%	0.01%	0.01%	0.01%	0.00%	0.00%
CONSENT DECREE	0.02%	0.02%	0.02%	0.02%	0.01%	0.02%	0.01%
CHARGE NOT FILED	0.05%	0.05%	0.02%	0.01%	0.00%	0.07%	0.29%
CONDITIONAL/DISMISSED	0.04%	0.04%	0.06%	0.03%	0.03%	0.05%	0.00%
CHARGE DROPPED	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%
DISMISSED	10.66%	10.66%	11.42%	10.01%	9.56%	10.11%	13.66%
CONDITIONAL DISCHARGE	7.98%	7.98%	7.66%	8.34%	9.52%	8.72%	2.92%
DEFERRED/DISMISSED	0.06%	0.06%	0.04%	0.08%	0.11%	0.04%	0.00%
DEFERRED	4.92%	4.92%	4.02%	4.94%	6.26%	6.29%	2.40%
DISMISSAL BY COURT	0.22%	0.22%	0.40%	0.17%	0.10%	0.16%	0.17%
DIRECTED VERDICT	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%
EXTRADITED	0.01%	0.01%	0.01%	0.02%	0.00%	0.00%	0.00%
GUILTY	48.31%	48.31%	45.39%	50.33%	49.18%	47.81%	51.05%
GUILTY BUT MENTALLY ILL	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
GUILTY BY PLEA	3.60%	3.60%	5.98%	3.14%	1.72%	2.06%	3.80%
GUILTY AT TRIAL	0.06%	0.06%	0.10%	0.03%	0.06%	0.02%	0.04%
HUNG JURY	0.01%	0.01%	0.01%	0.01%	0.01%	0.00%	0.00%
HABITUAL OFFENDER SAME PERSON	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%
MISTRIAL	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%
NOT GUILTY	0.20%	0.20%	0.25%	0.13%	0.18%	0.24%	0.23%
NOT GUILTY BY REASON INSANITY	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%
NOLLE PROSEQUI	3.34%	3.34%	4.13%	2.87%	1.95%	2.82%	6.32%
PRE-PROS PROGRAM/CHG DISMISSED	0.73%	0.73%	1.27%	0.99%	0.39%	0.04%	0.00%
PROSECUTION DECLINED	0.35%	0.35%	0.18%	0.37%	0.42%	0.47%	0.55%
ACCEPTED INTO PRE-PROS PROGRAM	0.69%	0.69%	0.45%	0.96%	1.02%	0.65%	0.00%
PRE-PROS PROGRAM/UNSUCCESSFUL	0.25%	0.25%	0.31%	0.41%	0.18%	0.09%	0.00%
RELEASED/NO CHARGES FILED	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
SUSPENDED SENTENCE	7.95%	7.95%	7.35%	8.67%	9.58%	9.12%	1.79%
MISCONDUCT BY ATTORNEY	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
VACATED CONVICTION	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
SEE REMARKS (WITH DISPOSITION INFORMATION)	3.22%	3.22%	4.04%	0.47%	3.26%	4.20%	7.06%
SEE REMARKS (NO DISPOSITION INFORMATION)	3.09%	3.09%	3.34%	4.20%	1.67%	2.69%	2.95%
INVALID CODE	0.03%	0.03%	0.03%	0.02%	0.06%	0.02%	0.06%
TOTAL N	12,405	85,751	25,718	22,537	18,016	12,216	7,264

Appendix C: Multiple Originating Agencies on Same Day

Table C.1 Type of originating agencies when multiple arrests on same day

Originating Agencies	Percent	N
Law Enforcement and Parole/Probation	56.83%	9,104
Multiple Law Enforcement	18.25%	2,924
Law Enforcement and Corrections	13.73%	2,200
Corrections and Parole/Probation	7.53%	1,207
Multiple Parole/Probation	1.45%	232
Multiple Corrections	1.25%	200
DA and Law Enforcement	0.37%	60
Law Enforcement and Other	0.25%	40
DA and Parole/Probation	0.14%	22
Parole/Probation and Other	0.06%	10
Corrections and Other	0.04%	7
Court and Parole/Probation	0.04%	6
Court and Law Enforcement	0.02%	4
Law Enforcement and School	0.02%	3
Court and Corrections	0.01%	1
Corrections and School	0.01%	1
Total		16,021

Works Cited

Goggins, B. R., & DeBacco, D. A. (2018). Survey of state criminal history information systems, 2016: A criminal justice information policy report.

Criminal Justice Reforms, House Bill 267, 2019 Regular Session (New Mexico) (U.S.A).