State of Oregon: NCJRP Phase III Data Infrastructure and Integration Plan

1. Introduction

This Data Infrastructure and Integration Plan describes Phase III of Oregon's NCJRP grant. Phase III includes implementing several jail database reforms in Oregon, as well as the subsequent research projects this database will support. Specifically, this document details the steps that will be taken to build this jail data infrastructure, the data collection processes, the subsequent fidelity checks, and the resulting research project that will focus on examining the causes and effects of pretrial detention in Oregon. These efforts are undertaken by the Oregon Criminal Justice Commission (CJC), which houses Oregon's Statistical Analysis Center (SAC), in cooperation with researchers at Portland State University (PSU). This Data Infrastructure and Integration Plan will describe the current condition of jail data in Oregon, the limitations that these data pose for criminal justice research in the state, how the reforms to these data collecting practices will help alleviate these limitations, the preliminary methods for evaluating the resulting data gathering processes, and the utility of the ensuing research projects. This document was written by the CJC, with the cooperation of researchers at PSU, where the point of contact for grant administration is Bridget Budbill (bridget.budbill@oregon.gov), and the point of contact for technical inquiries is Ken Sanchagrin (ken.sanchagrin@oregon.gov).

2. Background

The primary policy goal of this NCJRP grant is a significant improvement in Oregon's criminal justice data systems. Following these improvements, evaluative studies and policy analyses will be completed using the new data, which, in turn, will inform potential future policy and regulatory reforms. Enumerating these policy and regulatory reforms at this juncture, however, would be premature, as it is impossible to say with a significant degree of certainty what broad patterns exist across Oregon's 31 county-operated jail institutions and what policies should be improved and/or enacted. The SAC is aware of pretrial and other jail reform efforts that have proven effective in other states, but without Oregon-specific data that accounts for Oregon's unique policy framework and population patterns it remains unclear whether these same reform efforts will have similar results. Further, because Oregon is a geographically large state, with varying frameworks of local control, a comprehensive jail database is necessary to identify problems and construct solutions across the state.

Phases I and II of this NCJRP grant included a preliminary research study as well as planning for Phase III. The preliminary research study focused on the impacts of pretrial detention on sentencing outcomes in nine Oregon counties. The CJC developed non-violent, felony conviction cohorts for each county for the years 2016-2017 based on data from the Department of Corrections database. The identifying information for each individual in these cohorts were sent to each county with the request that the counties return to the CJC (1) the length of pretrial detention for each individual and (2) whether or not the individual was released at some point prior to trial. When the data was compiled, two outcomes were evaluated using matching regression techniques: (1) the type of sentence received (jail, prison, or probation), and (2) the length of the sentence. The PSU researchers found that being detained

for the full pretrial period was associated with 2.1 times greater likelihood of receiving an incarceration sentence, but had little impact on the length of sentence, conditional on the type of sentence.[1]

Oregon currently has statewide data systems for many facets of the criminal justice system. These databases include the Statistical Analysis of Policing (STOP) database, which tracks the characteristics of non-service call police stops; the Law Enforcement Data System (LEDS) database, which tracks all arrests in the state; criminal and specialty court databases that are tracked by the Oregon Judicial Department (OJD); and the incarceration and felony probation database kept by the Department of Corrections (DOC)[2]. Data integrity and consistency are ongoing concerns and areas of focus for the CJC in working with these data sets, but these databases are generally reliable and known commodities.

The notable gaps in data coverage, however, include detention and release information prior to and during trial, case processing information of lower-level misdemeanors, and criminal cases that are never transferred to DOC for processing. These data are tracked by the jail and/or county level administrators, but Oregon has 31 local jails across 36 counties that have highly variable populations and resources. Additionally, Oregon jails operate 17 different jail management systems (JMS), and even jurisdictions operating the same JMS (or using the same JMS vendor) are unable to extract data sets with ease or consistency or connect systems to share data jail-to-jail. The data fields and data quality, thus, are inconsistent across these jurisdictions.

Further, there is no infrastructure to support data sharing with the SAC at regular intervals. Therefore, the SAC may request singular data draws from each individual jail jurisdiction, but statewide data is impossible to obtain on a reasonable timeline or at regular intervals and there are severe challenges with measurement errors or inconsistencies in the final, aggregated data.

Given these data limitations, past SAC research concerning both pretrial case processing and jail systems have been limited in both scope and scale, similar to the preliminary research completed in Phases I & II. Oregon policymakers, criminal justice stakeholders, and community members are increasingly concerned with questions pertaining to the causes and effects pretrial detention, the social cost of incarceration, the patterns of jail capacity and jail overcrowding, and sentencing outcomes for lower-level statutory crimes. The current limitations in jail data precludes the Oregon SAC from answering these questions at a systematic, statewide scale that informs policy discussions. Without an Oregon jail database, any policy questions pertaining to jail systems may only be answered with speculation based on small-scale, idiosyncratic research or on findings from other states.

The first step of Phase III is initiating jail data sharing relationships between the SAC and each of the 31 local jails across the state. This effort will be followed by effort to build the infrastructure to standardize a statewide jail data system with regular updates through a data portal. Initially, the SAC is reaching out to each jail commander across the state to get a single data draw of their jail information for the 2018-2019 period. The SAC will then analyze these

---

[1] The full research study may be found on the CJC website here: https://www.oregon.gov/cjc/CJC%20Document%20Library/EffectofPretrialDetention.pdf.
[2] Only a subset of more serious misdemeanors are tracked in this data set. All other misdemeanors are not tracked.

data to determine existing overlap between systems, significant shortcomings, and potential areas of data integrity failures. The information gathered in this initial stage will be the foundation for the SAC to establish an ongoing database, inform how the SAC may best provide technical assistance to the counties and jails. The information will also guide how the new data system may best be structured to minimize the burden on state and local resources.

3. Evaluation Data Requirements

As described in previous NCJRP documentation, this project does not address a set of specific, pre-determined policy questions. Instead, the additional data capacity created through this project will facilitate further discussion of all jail-related policy discussions. To this end, the formative and outcome evaluations related to the new data systems will be used to improve these data systems in the future, rather than ensuring that these data conform to a previously unmeasured baseline in Oregon. For example, some jail data sets do not currently log each jail entry and exit event, but rather just the initial booking and the final release, whether this final release was caused by a pretrial release event or occurred at the end of a sentence. This is a shortcoming in *data entry*, but is not a *data integrity* issue. The CJC will work directly with jails where anomalies or potential integrity issues arise, but the best that can be done is to improve the data entry systems and/or process going forward and use the historical data as best as possible.

The other key "outcome" will be the next stage of the pretrial detention analysis completed by PSU. The CJC does not have expectations on the scale or scope of the conclusions formed by the research team, as these expectations would preclude the independence of the PSU research team. The CJC does, however, expects the research to conform to professional standards in criminological research, be presented to and subject to the scrutiny of the relevant advisory and legislative committees, and be subjected to the rigors of the academic peer-review process. Concerning the outcomes measured in this research itself, PSU will further explore the patterns of pretrial detention and release and, in particular, explore the antecedents of pretrial detention decisions.

4. Governance and Policies

All data infrastructure governance and policies are determined by the CJC. Stakeholders, namely jail commanders and staff and individuals representing groups that have frequent interactions with the jail system, have been peripherally involved in the initial data gathering process and will continue as a jail advisory committee. These groups will also be relied on for data validation checks, as needed. The CJC is responsible for infrastructure development. A limited number of data fields received by the CJC contain sensitive personal information. This data is universally removed prior to generation of any data products as this information does not serve any research purposes. Further, prior to releasing information via, say, a FOIA request, data is scrubbed of all name and identifying information. Unique individual or case identifiers are generated.

A primary challenging in sharing this data with 3rd party researchers will be in the evaluation of the 3rd party's research proposals. As described elsewhere in this document and in

prior NCJRP documents, the data generated through this grant will have shortcomings, despite being a vast improvement over previously non-existent data systems. These shortcomings will preclude answering a range of research questions that are likely to be of interest to several researchers. The CJC will evaluate these requests and determine if the jail data set can be used to answer the question and, if not, work with researchers to adjust the research question and/or direct them to the data resources more likely to serve their purposes.

5. Data Sharing Architecture and Management Environment

Thus far, the CJC has received data from jails through a combination of email submissions and SFTP submissions. In the future, the CJC hopes to receive the data sets from jails at regular, to be determined, intervals. These data will be submitted through a standard SFTP site, similar to that used for the initial data draw. CJC staff will then clean, organize, and manage the data on their internal CJIS certified servers. Once data is fully cleaned and organized it will be shared with research partners via an online, cloud-based data portal. Data will be posted to this environment as singular, one-off data sets to serve single analyses, unless more regular data updates are required for an yet unforeseen project. Data is typically stored and shared in .csv format, but CJC adapts this to the needs of the researchers if .csv format does not suffice. Data use agreements are in place for all data sharing partners and all infrastructure conforms to standard data security protocols. All CJC research staff have Level 2 CJIS security awareness certification.

6. Data Curation

The CJC has already started data quality analysis on the first set of jail data submissions. This process has led to several forms of consolidation and data normalization in multiple dimensions, as jails have differing practices for data tracking and utilize a total of 17 different jail management systems across the 31 county-level jails. Some systems are set up as case management systems where there is a single row for each booking event with information that is updated and overwritten as the case progresses. Other jails have a different row for each discrete stage of the case (e.g., pretrial, sentence). Still others have a different row for each charge related to a single booking. The second dimension by which jails vary is through the contents of each data field. Data fields generally have a similar range of underlying definitions, but have different labels and coding conventions. For example, most jails have a booking reason variable that includes information such as "warrant", "probation violation", or "original offense." How this information is coded, however, varies by jail. Probation violation, for example, may be coded as "PROB VIOL", "PV", "PROBATION VIOLATION", "VIOL", and others. Sometimes there is variance within jails or non-intuitive values (i.e., one jail has codes corresponding with most letters of the alphabet, but provided no data dictionary). All variables have at least some degree of variance across jails.

The CJC normalizes the data by first cleaning and recoding the data, to the extent possible. The CJC then appends each jail's data set into a single file. The first set of data products the CJC produces are each individual jail's cleaned data and a unified dataset that preserves the maximum information from each jail. This data set, without further processing, is not directly useful for *statewide, cross-jail* comparisons since the information contained in each row varies

by jail, but is useful for the fullest-information *within*-jail analyses and does facilitate some cross-jail analysis where systems and data practices are similar.

The second data product collapses this data to the relevant information *by unique original booking*. Where, for example, pretrial release dates are present we will calculate the pretrial detention duration and sentence duration (if any) and keep this information on the same row. Similar processing will occur with multiple charges and multiple booking and release reasons. A key component of this data set will be the construction of individual (rather than case) identifiers. This will enable analyses of individuals who repeatedly enter jail by demographics and crime type for outcomes such as incarceration time, processing time, and sentencing outcomes.

7. *Data Analysis*

Considering the Phase II findings that pretrial detention increases the likelihood of an incarceration sentence, it is possible the observed relationships may be related in part to courtroom scheduling, procedures, idiosyncratic case and defendant details, and discretion. The Phase II conclusions, however, were not able to exceed the limitations of the available data. Thus, many questions remain related to the complexities and informal decision-making that occurs across courtroom actors, judges and prosecutors, in particular. Identifying the causes, procedures, and consequences related to these complexities requires more data. Most importantly, additional quantitative measures are required (e.g., failure to appear information), more qualitative data is needed to depict what is happening in these quantitative data patterns, and a larger length and breadth of data to ensure that the current information is not solely driven idiosyncratic policies, economic conditions, or other factors.

Subsequently, at this stage, we will examine the decision-making processes of pretrial detention in Oregon counties in more detail. Specifically, because the preliminary analysis showed that pretrial detention is strongly related to sentencing outcomes, the evaluative stage of Phase III will examine the factors that initially lead to pretrial detention. Doing so will help the state better understand reasons for why pretrial detention is so influential in increasing the likelihood of incarceration. In addition to the qualitative interview schedules being formulated to capture pretrial decision-making, it is recognized that county differences may be rather vast depending on factors such as resource availability for pretrial services or risk assessment use. Compounding those differences were policy changes taking place over the last few years. Throughout the Phase II examination of pretrial detention effects on sentencing, counties were simultaneously rolling out efforts to reform pretrial practices. As a result, the CJC and PSU have partnered to launch a digital survey of pretrial practices as they exist in each county. The survey will collect information on jurisdictional priorities, tools, and protocol used in the pretrial detention/release process. Data from this survey will be used to identify major differences between counties that are taking steps to reforming pretrial services and practice, and those that are not.

During and upon collection of the survey data, PSU researchers will devise and the qualitative portion of this study, which involves in-depth, face-to-face (including video-conference) interviews of judges, prosecutors, defense attorneys, jail/detention officials, and pretrial services personnel. PSU aims to include a representative voice from at least the original

nine counties, though giving precedence to ensuring that half of our qualitative observations are from counties *with* pretrial programs, and half are from counties *without* such programs. PSU's ultimate aim is to recruit five to six prosecutors, judges, defense attorneys, jail/detention and pretrial staff from each of an approximated nine counties for about 50 total interviews.

Upon execution of the pretrial jail database and development of PSU's qualitative pretrial decision-making project in Phase III, Oregon will, for the first time, have information and data components necessary to evaluate a statewide pretrial program reform effort during Phase IV. This will include comparing jurisdictions with pretrial programs with those that do not have pretrial programs and ascertaining which formulations of pretrial programs may be most successful in Oregon and how to scale successful pretrial programs across jurisdictions without programs in place.

8. Summary

Phase III of this generous NCJRP grant will lay the groundwork for data-driven criminal justice system evaluations in Oregon and inform legislative policy discussions regarding pretrial reform in the near term as well as future jail population issues. Phase III encompasses the groundwork to implement data submission processes and ensure final database fidelity, which will support reliable analysis of the resulting data. Phase III also initiates the process of the next phase of pretrial research in Oregon, which will be the first test of our new jail data capabilities. Subsequently, Phase IV's prospective pretrial program assessment work will rely on the data and information acquired during Phase III.