Pilot study preliminary report on statewide judicial data for NJRP

Submitted by:
The Urban Institute
2100 M Street, NW
Washington, DC 20037
(202) 833-7200

Submitted to:
Bureau of Justice Statistics
810 Seventh Street, NW
Washington, DC 20531
July 2, 2010
Preliminary report

This project was supported by Grant No. 2008-BJ-CX-K052 awarded by the Bureau of Justice Statistics. The opinions, findings, and conclusions or recommendations expressed in this document are those of the authors and do not necessarily reflect the views of the U.S. Department of Justice.

Authorships and Acknowledgments:

This report was written by Rob Santos, Senior Institute Methodologist, and Molly M. Scott, Research Associate.

MEMORANDUM

TO: Thomas Cohen, BJS

FROM: Robert L. Santos and Molly M. Scott, Urban Institute

RE: Pilot Study report on statewide judicial data for NJRP

DATE: July 2, 2010

1. Introduction

The purpose of this memorandum is to report on Urban Institute's research of nine states in support of the larger re-design of the National Judicial Report Program (NJRP). This research represents an initial exploratory (pilot) effort on the feasibility and resources needed to determine the existence, nature and availability of standardized court data from state level administrative data systems. Specifically, the Bureau of Justice Statistics (BJS) wants to know the extent to which NJRP and other judicial data could be readily obtained from states rather than from the historical approach of gathering data at the county level.

As a first step, BJS and the Urban Institute (UI) identified a group of nine states and began calling and collecting information about the availability and nature of state level judicial data. The results of our data collection are used to shed initial light on the primary research questions as well as assess the feasibility and costs of extending our efforts from these nine states to all 50 states plus the District of Columbia.

We begin with a review of the primary research questions for this specific re-design effort. Then we discuss our methods and present findings in the context of the each specific research question. The last section of the memo summarizes our findings and provides insights into the NJRP redesign effort.

1.1 Research Questions & Pilot Objectives

Our research project supports the larger NJRP redesign effort being conducted by BJS. To the extent that resources permit, we address the following **research questions**:

1) What is the capacity of states to generate NJRP data at the state level?
2) What is the capacity of the states to integrate criminal histories into NJRP data?
3) What kinds of "pre-sentencing" data are available at the state level?
4) How feasible is it for NJRP to expand its data collection to include other dispositions and misdemeanors.

Pilot study. Due to the inherent uncertainty associated with the availability of state level administrative judicial data, we adopted (under the direction of BJS) a research design that begins with a pilot study to test methods of efficiently and effectively soliciting information from states to establish the existence and availability of NJRP and other judicial records at the state level. The pilot involves a subjective sample of nine states that span what we believe to be the heterogeneity of scenarios across states in terms of the existence, availability, and sophistication of statewide administrative judicial data systems.

The ***pilot study research objectives*** are to:

1) develop and test methods of identifying the appropriate state entities to efficiently elicit data on the availability and characteristics of statewide judicial records on sentencing and convictions (i.e., NJRP data), case histories and other data (e.g., misdemeanors).
2) establish the extent to which the availability of judicial data can be discerned with certainty (or whether that cannot be completed without actual extraction and processing of data);
3) develop some preliminary responses to the principal research questions posed by BJS.

This study represents a first step in determining the feasibility of statewide data collection of NJRP data. Relatively little is known about existing statewide judicial record systems, including their:

- expanse (how many states have them),
- depth (what items collected),
- coverage (all versus some counties; all versus some types of felony convictions),
- quality (missing data, standardized coding),
- availability (restricted access issues),
- currency (frequency of updates), and
- acquisition/timing process (necessary approvals and duration of approval and data preparation processes).

The pilot study explores and tests methods of soliciting and assessing who should be contacted to inquire about the existence, availability, quality, etc. of statewide NJRP-type judicial data from state level data systems. We assess the methods to establish whether or not states can supply state level data and the utility of those data for fulfilling NJRP judicial reporting needs. And we make some assessments (though we unable to be definitive given the nature and limited resources of this pilot) about the potential for providing NJRP and related data for the data systems we encountered. We do this in the context of the actual research questions posed by BJS to determine the feasibility of statewide data gathering.

Background. Since 1983, the National Judicial Report Program has provided case-level data on felony convictions and sentencing in state courts. Conducted biennially on even numbered years, the NJRP dataset includes information on offenses, convictions, penalties, and socio-demographic characteristics such as age, race and gender. The data can be analyzed both at national and county levels. Data are use by BJS to report national estimates of convictions and sentencing as well as characteristics of convicts. The NJRP data are used by federal state and local government agencies, researchers and policy analysts. As such, the program represents a national resource.

BJS initiated a review of the NJRP design in light of societal, demographic and technological changes over the past quarter century. One prominent design feature of special interest involves the NJRP's clustered, two-stage sample design. *Counties* are selected at the first stage of sampling; sentencing and conviction data are collected from the first stage counties at the second stage. In more recent iterations of the study, field staff has noticed that county courts increasingly retrieved their NJRP reporting data from centralized systems designed and maintained at the *state level* and covering all or most of the state. It is reasonable to expect that the trend towards state level judicial reporting systems will continue. As such, it is appropriate to consider the extent to which NJRP data elements can be (1) captured through state level centralized data base systems, and (2) whether or not sentencing and conviction data covering the *entire state* can be collected at least as (if not more) easily than under the current practice of county specific data requests. Clearly, sizeable gains in both cost and statistical efficiency could be realized by gathering sentencing and conviction data for entire states from centralized, standardized state data systems.

2. Methodology

2.1 State selection

Nine states were purposively identified for this research effort. Our selection criteria sought a collection of states that spanned the national range of technological sophistication and existence of state sentencing commissions. We also sought states that had at least one county sampled in the 2004 NJRP, experienced high volume court activity, and spanned all regions of the country.

To identify candidate states we reviewed the report entitled "State Sentencing Guidelines" issued by the National Center for State Courts to identify a subset of states whose state sentencing commissions "regularly report on guideline compliance." We then performed internet searches on these states to find those that published annual reports using statewide data. From this sub-group, we chose 4 states: Washington, Virginia, Minnesota, and Pennsylvania.

We then examined notes from the implementation of the SCPS survey to identify states without sentencing commissions that nonetheless maintained some kind of centralized

state data.[1] We expected these states to represent the midpoint of a continuum between those with highly centralized/standardized court data systems at one extreme and those with only decentralized (i.e., county level) data systems at the other. The states we selected were: Arizona, New York, and Texas.

Lastly, we identified 2 states without sentencing commissions *that did not have any sampled counties in the SCPS data collection*. These were thought to have limited statewide judicial statistics systems. These states were Georgia, and Idaho.

2.2 Protocol development

In order to systematically capture the same types of information from all respondents, we constructed a semi-structured protocol. The protocol originally was designed to determine the availability and content (if available) of state level judicial data, its quality, and ease of access for an agency like BJS. During the course of collecting information we found it beneficial to revise the protocol to make it adaptable to staff from various types of agencies (i.e., case management vs. research departments). We also added questions to explore data availability when the agencies themselves did not gather certain types of data. We structured the questions to facilitate more detailed information about the source of the data and the types of offenses included. The final protocol appears as Appendix A.

2.3 Overview of the interview process

In almost all cases, identifying and obtaining contact information from relevant state agencies required internet searches followed by a series of trial-and-error "cold calls" to identify most knowledgeable staff. And after making the *initial contact* with the state, it was also necessary to ask respondents for referrals to other agencies in the state at the end of each interview.

In most cases, conducting a comprehensive inventory of state judicial data required contact with two or more agencies in each state. At a minimum, calls were issued to (a) the state court administration and (b) the criminal records depository. For states with an active sentencing commission, we also needed to speak to the agency charged with studying compliance with corresponding guidelines.

We conducted interviews with a total of 21 different agencies by phone between April 5th and May 28th. The states and types of agencies that were contacted appear in Table 1. There was only one agency -- the Pennsylvania State Police -- that we were unable to interview within our study period.

In many cases, interviewers needed to speak directly to several staff within agencies to provide the information we requested in the interview protocol. For example, it was common that the person with the best global knowledge of the data and its day-to-day use (e.g., Research Analyst or Criminal Records Supervisor) was not the best person to ask

[1] For most of these states, only criminal history data was held and maintained by state agencies.

about more technical issues such as the size of the database, data access protocols, or the availability of a codebook or comprehensive list of data fields.

Table 1. Agencies Contacted in the NJRP Pilot Study by State and Type

	Agency Type		
State	Court Administration	Criminal Records Depository	Sentencing Commission
AZ	x	x	
GA	x	x	
ID	x	x	
MN	x	x	x
NY	x	x	
PA	x		x
TX	x	x	
VA	x	x	x
WA	x	x	x

3. Findings

In this section we present the findings of our qualitative analysis. We first examine the number of agencies we encountered and categorize them by *overall* quality of the NJRP-type data they reported having. We then consider how agencies rated in terms of three dimensions of quality – coverage, completeness and standardization.

Focusing on those state agencies with *high* and *medium* data quality assessments, we then consider issues of accessibility, usability (in terms of documentation) and presence of NJRP-type data items, criminal history items and other data.

We begin by presenting our framework for three quality dimensions that will drive our later analyses, findings and discussions.

3.1 Data Quality Framework

Capturing statewide data under a NJRP redesign requires first and foremost that conviction and sentencing records exist in one or more state-level databases. However, *data existence* is necessary but not sufficient. Such data should (1) cover the entire state, (2) include all current/historical NJRP-related data items, and (3) either be standardized to or easily mapped to NJRP coding conventions via batch processing (as opposed to visual inspection and coding of each data element which would be intractable). This common sense paradigm drives the framework we will use to assess agency level data quality for NJRP. Accordingly, we adopted the following three dimensions to characterize state agency data availability and developed three 'ratings' (i.e., high, medium, low) to characterize the 'quality' according to each dimension:

1. ***Coverage: Reporting compliance among jurisdictions***
   - *High*: All jurisdictions report
   - *Medium*: A few large volume jurisdictions are excluded
   - *Low*: At least half of jurisdictions do not report
2. ***Completeness of NJRP data items***
   - *High*: Data on both conviction and sentencing are complete
   - *Medium*: Some missing data mostly because of matching issues
   - *Low*: A lot of missing data due to lack of agreement on required fields
3. ***Standardization of data items***
   - *High*: All data are cleaned and standardized and put into analytic datasets
   - *Medium*: Administrative data exist with standardized fields and coding systems, but coding practice standardization is not verified
   - *Low*: No standardized fields or codes (e.g., each jurisdiction employs their own schemes to code convictions and sentencing records)

All agencies' data are classified along these three dimensions. For the sake of parsimony, we used the three dimension ratings to develop an ***overall quality index*** in order to identify the agencies are the best candidates for NJRP data collection. The overall quality groupings we identified were:

- ***high quality – sentencing & conviction:*** this group of agencies appears best positioned to provide statewide judiciary records with some/all NJRP items; agencies in this group earned high quality ratings on all three dimensions;
- ***medium quality – sentencing & conviction:*** These agencies appear to have the data but there may be some "nonfatal" issues of geographic coverage (e.g., most but not all counties included), completeness (e.g., some notable issues with missing data), and/or standardization (e.g., requiring coding/processing); agencies in this disposition display a mixture of 'high' and 'medium' ratings across the three dimensions;
- ***low quality – sentencing & conviction:*** agencies in this group have nontrivial limitations/gaps, and accordingly earned a 'low' quality rating on one or more dimensions;
- ***incomplete – conviction but no sentencing data:*** these agencies fail to provide a key group of NJRP items; the quality of the available data in this disposition group was 'high' or 'medium' but these agencies all lacked sentencing data which is critical to the NJRP;
- ***absence of case-level data;*** no judicial records data are available at the state level; these agencies provided no data useful to NJRP.

Our subsequent analyses then focus only on those agencies considered to be a viable data source for NJRP data – those whose data are rated as having collectively across the three dimensions "high" and "medium" overall quality.

Limitation of our scale. We acknowledge that this three dimensional qualitative rating (high, medium, low) is a subjective, quasi-qualitative assessment. Time and resource limitations prevented a more concise characterization of data quality (which we recognize would have been ideal). For instance, we provide a rating of 'high quality' on completeness for state agency data systems that appear to contain all or virtually all data items historically collected in NJRP. But in order to validate that all NJRP data items exist in this data base, we would need to acquire a test file, develop a crosswalk between the each NJRP item and its corresponding state data base counterpart, then verify that theses data are actually being entered (as opposed to being listed on a coding sheet/dictionary but exhibiting missing data rates of 80-90 percent). Given the volume of agencies/entities that we encountered and the limited documentation that was made available, such a quantitative assessment was beyond the resources and scope of this pilot. We chose instead an intermediate strategy – developing a good sense of the potential for NJRP availability from statewide data using our qualitative ratings scales spanning three dimensions. This provides a broader brushstroke, allowing us to paint a general portrait of the quality of the data provided by state agencies within a reasonable time/resource window. It also leaves the option for us or others to delve more deeply into specific quality issues should BJS so choose.

We now consider each research question in turn and use our qualitative assessments to give insight into statewide NJRP data collection as well as address our pilot study objectives.

3.1.1 Capacity of states to generate NJRP data at the state level

Overall availability by state. We examine the capacity of states to generate NJRP state level data by first presenting the "big picture" of overall data quality, as defined in the preceding section. We consider state agency data to be "available" to meet NJRP needs when the overall quality ratings across coverage, completeness and standardization were uniformly deemed of "high" or "medium" quality. To be clear, any designation of "low" quality (or unavailable data) on any of the three quality dimensions would result in an "unavailable" designation.

Table 2 presents the availability and overall quality of the data for the 21 state agencies contacted in this pilot study. The first two columns of Table 2 (with headings "high' and 'medium") denote those agencies that we determined had statewide judicial data *available*. Over half (14 of 21) of the agencies we contacted have both conviction and sentencing data of nominal quality (i.e., with a "high" or "medium" rating across our three quality dimensions – coverage, completeness, standardization). About a third of the agencies did not have data available or had data of sufficiently low quality (in terms of coverage and/or completeness and/or standardization) to deem NJRP data *unavailable* from these sources.

Virtually all states in our sample offer one or more agencies with *potentially* useful NJRP data. Only four agencies provide consistently high quality data including all felonies and likely to require only minimal processing for NJRP use. Three of these states (MN, VA,

and WA) have sentencing commissions; the other is NY, a state with highly centralized court reporting system. The majority of states offer at least some data on conviction and

Table 2. State Agency* Sampling Criteria, Availability and Overall Quality of Conviction and Sentencing Data by State

	NJRP Available **		Unavailable		
States by Sampling Criteria	**High Quality**	**Medium Quality**	**Low Quality**	**Unusable**	***Total***
Sentencing commission states					
MN	1	2	0	0	***3***
PA	0	2	0	0	***2***
VA	1	1	0	1	***3***
WA	1	1	0	1	***3***
SCPS states w/o sentencing commissions					
AZ	0	0	2	0	***2***
NY	1	1	0	0	***2***
TX	0	1	0	1	***2***
Non-SCPS states					
ID	0	1	1	0	***2***
GA	0	1	0	1	***2***
Total	***4***	***10***	***3***	***4***	***21***

* *Cell entries denote the number of agencies with the cell-designated attributes*

** *NJRP-available means that SOME BUT NOT NECESSARILY ALL NJRP-data elements are available from the agency data base.*

sentencing, regardless of sampling criterion (i.e., sentencing commission states, SCPS states, and non-SCPS states). Unfortunately, most of the databases are problematic in some way. They include only a subset of felonies or require substantial cleaning/processing. Surprisingly, Arizona, an "SCPS-state" with some judicial record data, has overall low data quality, rendering it *unavailable* to NJRP.

Quality of data by individual agency. The first four rows of Table 3 show that Sentencing Commissions represent three of the four agencies holding highest quality ratings across the ratings dimensions.[2] New York's Division of Criminal Justice (NYDCJ) is the only exception. The main difference between NYDCJ and its counterpart is that NYDCJ and NY Office of Court Administration share each other's data and the two agencies actively collaborate to clean and standardize data in real time. They even produce quarterly analytical datasets.

[2] Virginia's sentencing guidelines exclude about 5% of felonies in the state; however, we still

Table 3. Agency Ratings on Three Quality Dimensions for Conviction & Sentencing Data

State	Agency	Quality Dimensions		
		Coverage	Complete-ness	Standar-dization
High Quality -- Sentencing & Conviction				
NY	Division of Criminal Justice Services	+	+	+
MN	Sentencing Commission	+	+	+
VA	Sentencing Commission	+	+	+
WA	Sentencing Commission	+	+	+
Medium Quality -- Sentencing & Conviction				
GA	Bureau of Investigation	+	+	+
ID	Idaho State Judiciary	+	+	+
MN	Admin Office of Courts	+	+	+
MN	Bureau of Criminal Apprehension	+	+	+
NY	Office of Court Admin	+	+	+
PA	Sentencing Commission	+	+	+
PA	Admin Office of Courts	+	+	+
TX	Bureau of Investigation	+	+	+
VA	Supreme Court	+	+	+
WA	Admin Office of Courts	+	+	+
Low Quality -- Sentencing & Conviction				
AZ	Dept of Public Safety	+	o	+
AZ	Admin Office of the Courts	o	o	o
ID	Bureau of Criminal Identification	o	o	o
Incomplete: Conviction But No Sentencing Data				
VA	State Police	+	+	+
WA	State Police	+	+	+
Absence of Case-Level Data				
GA	Admin Office of Courts	n/a	n/a	n/a
TX	Admin Office of Courts	n/a	n/a	n/a

*** key to table cell entries: + = high quality; + = medium; o = low; n/a = unavailable**

Ten agencies with "medium" quality ratings provide some NJRP data but ultimately have deficiencies. Coding of conviction and sentencing items can vary widely and are non-standardized across different courts or counties within a given state. Moreover, criminal records agencies in this group are subject to potentially nontrivial fractions of missing data. In order for disposition and sentencing data to appear in the dataset, they must be matched to an existing arrest record. But if the arrest record is absent or otherwise fails to match a court record (e.g., because of a data entry error), then all corresponding conviction and sentencing data are in effect missing. Some "medium" quality agencies also suffer from coverage problems. In Virginia, three populous jurisdictions are excluded (Fairfax County, Alexandria, and Virginia Beach) from the state's court administration database. And despite its high completeness and standardization, Pennsylvania's Sentencing Commission does not include data from the high volume Philadelphia Municipal Court, which holds preliminary hearings for felony cases.

The last three groups in Table 3 (Low Quality, Incomplete and Absence of Case-Level Data) contain agencies whose data have serious quality deficiencies (i.e., severe lack of standardization, lots of missing data, and/or significant geographic coverage issues), do not include data on sentencing, or are only available in aggregate form.

These findings reinforce the big picture portrait we viewed in Table 2. From a logistical/implementation perspective, we also affirm the field protocol strategy of commencing contact with state sentencing commissions to explore the availability of statewide NJRP-relevant data items.

3.1.2 Data Accessibility

It is of no use to have statewide data if the length of time to acquire it is inordinately lengthy. This section explores how long it can take to obtain the data once it has been requested. We investigate accessibility (timing) by asking the agencies about the acquisition process. ***We restricted attention to situations that held the promise of providing NJRP data – the 14 agencies whose quality dispositions were designated "high" and "medium" in Table 3.***

Most agencies were reluctant to estimate the amount of time needed to process a data extract request for NJRP data. The reasons were twofold. First, some state court administration offices experience seasonal workload flows. When state legislatures are in session, for instance, agencies sometimes endure a much heavier workload due to data requests. In turn this leads to a longer turnaround time for all other lower priority data requests. Second, agencies understandably are uncomfortable estimating approval and processing time without detailed specifications for the data pull. So while we are able to establish the feasibility of extracting NJRP-type data from state agencies in this pilot study, it was not possible for agency staff to estimate turnaround time without detailed data specifications. Nonetheless, we managed to elicit more general responses from agencies about turnaround time from agencies.

The results of our inquiries are shown in Table 4, which presents procedures and rough estimates for approval and processing times required to pull NJRP-type data from data

systems. The agencies with "high" quality data ratings are generally able to respond to data requests within roughly the same time horizon, even though the request procedures varied by agency (i.e., formal request forms vs. informal email or phone requests). Most estimate it would take a few days to get a request approved and a few weeks to get it processed. The exception is the Virginia Sentencing Commission, which was unwilling to estimate the time needed to process a request.

Table 4. Accessibility of NJRP Data Items Among Agencies with High & Medium Quality Ratings

State	Agency	Request Procedure	Approval Time	Processing Time
High Quality Candidate Agencies				
NY	Division of Criminal Justice Services	request form	days	weeks
MN	Sentencing Commission	request form	days	weeks
WA	Sentencing Commission	informal request	days	weeks
VA	Sentencing Commission	informal request	days	no estimate
Medium Quality Candidate Agencies				
VA	Supreme Court	informal request	1 day	1 week
TX	Bureau of Investigation	informal request	days	2 weeks
GA	Bureau of Investigation	informal request	days	weeks
MN	Admin Office of Courts	informal request	weeks	up to a year
WA	Admin Office of Courts	request form	days	10 days
PA	Sentencing Commission	request form; $70/year of data	days	weeks
PA	Admin Office of Courts	request form	days	weeks
NY	Office of Court Admin	request form	months	months
MN	Bureau of Criminal Apprehension	request form	months	no estimate
ID	Idaho State Judiciary	***there is no policy for releasing micro data***		

Turning to the agencies whose quality disposition group was "medium," there is a noticeable degree of variation in the request procedures as well as processing and approval time. Like the previous group, data requests are made either informally or through the submission of a completed request form. One agency (the PA Sentencing Commission) also requires a payment of $70 for one year's worth of judicial data.

Variation is considerable with regard to the time required to respond to data requests. The approval process for requests was as little as one day (VA Supreme Court) and up to "months" (e.g., NY Office of Court Administration). And the processing time ranged from 10 days (WA Administrative Office of Courts) to as much as a year (MN Administrative Office of Courts).

The state of Idaho poses a unique situation. The Idaho State Judiciary *has no policy* for sharing case-level data with outside agencies. The Idaho respondent is not aware of any data requests similar to NJRP being fulfilled by the agency. All of the other high and medium quality candidates are accustomed to receiving and filling such requests.

Surprisingly, the existence of an updated centralized database does not guarantee timely data request processing. The Minnesota Administrative Office of the Courts (AOC) established a centralized database for its unified court system fairly recently, in 2008. Consequently, it is reasonable to expect that MN to be among the easiest to request and process data extractions. Yet the opposite is true. The agency reports that is has been swamped with data requests and emphasizes that the it does not have the capacity to fill them. When asked how long it might take to get a data extract from them, the staff person responded that it could take up to a year. Gaining approval to access to the MN Bureau of Criminal Apprehension's database can be time-consuming because it is governed by statute. All requests have to be approved at multiple levels of government before being actionable. Similarly, accessing data from the New York Office of Court Administration can also be a long process because all data requests must be reviewed and approved by a judge according to state law. This often requires a back and forth process that can take months. Once the request is approved, processing time varies from weeks to months, depending on the scope and complexity of the request.

Our findings with regard to accessibility have direct implications on the NJRP redesign effort. Based on our inquiries, the acquisition timing could pose a challenge and possibly a barrier to obtaining statewide judicial records data. Availability of statewide data is necessary but not sufficient for gathering NJRP data statewide. Availability needs to be coupled with a timely or well-planned data acquisition process that must be tailored to each state-specific agency.

3.1.3 Role of Data Documentation

Even when sentencing and conviction data are housed in a centralized system *and* are readily available, there is the challenge of understanding exactly what data are being stored and how they map to NJRP items. State court administration and criminal records agency staff do not use or create 'codebooks' the way that researchers, data analysts or IT staff do. Their data systems are designed to facilitate practical case management and/or respond to case-specific requests from law enforcement, legislators, or employers. Most of the information about the data system is "on screen." State employees learn data availability by interacting directly with the data systems as part of their work. As such, most state agency employees can be experts at using the system but know little about the underlying data structure, data items, coding conventions and other attributes of the data system they use. And when BJS or others ask about technical specifications of the data

system, they are unable to respond accurately and typically need technical terms explained to them. It has proven very difficult to obtain formal technical documentation. The information we have been able to gather is summarized in Table 5.

Table 5 presents the documentation availability and characteristics of statewide judicial data systems. Like Table 4, we restrict attention to the 14 agencies whose overall quality ratings were judged "high" or "medium" in Table 3.

Table 5. Documentation Availability and Characteristics Among State Agencies With High and Medium Quality Ratings

State	Agency	Documentation Provided	Fields Documented	Codes Documented
High Quality Candidates				
NY	Division of Criminal Justice Services	codebook	x	x
MN	Sentencing Commission	Statistical package printout	x	x
WA	Sentencing Commission	Statistical Package printout	x	x
VA	Sentencing Commission	Sentencing worksheet, statistical package printout	x	x
Medium Quality Candidates				
PA	Sentencing Commission	codebook	x	x
VA	Supreme Court	training manual	x	x
MN	Bureau of Criminal Apprehension	online guide	x	--
MN	Admin Office of Courts	sample docket	x	--
PA	Admin Office of Courts	sample docket	x	--
GA	Bureau of Investigation	database diagram	x	--
TX	Bureau of Investigation	training manual	x	--
ID	Idaho State Judiciary	--	--	--
NY	Office of Court Admin	--	--	--
WA	Admin Office of Courts	--	--	--

Table 5 generally echoes the data accessibility findings from Table 4. The documentation spans a variety of presentation formats, including simple diagrams of relational data bases, training manuals, sample dockets, printouts from statistical programs, and conventional codebooks. Only two agencies furnished traditional comprehensive codebooks and other supporting documentation. Another 4 provide documents that contained some information about variable labels and value codings,

though this information was often not fully documented and difficult to comprehend. Five agencies furnished documentation that allowed us to identify possible fields that overlap with NJRP but did not capture their coding schemes at all. Three agencies failed to provide any documentation whatsoever, attesting to the difficulty that agency staff experience in digesting the technical requests being made regarding NJRP-type data.

Finally, Table 5 reveals a mild relationship between documentation availability (fields, codes, codebooks) and agency type. Sentencing commissions do tend to have better documentation of fields and codes, but only 2 out of 4 maintain formal codebooks. On the flip side, NY's criminal records depository maintains a formal codebook that it updates quarterly along with its analytic dataset.

Documentation Reviews. A review of the coding documents supplied by agencies revealed that the documentation is often incomplete. Deciphering criminal charge codes routinely requires that reference be made to *state statutes* which, inconveniently, are not included with most documentation. Consequently, in order to generate NJRP-standardized conviction items, additional research is required to assess the completeness of documentation and then conduct additional research to address all the gaps that are uncovered. For instance, consider the agencies from Table 3 that were rated as having 'high' quality *standardization*. For these agencies it would be a simple task to process the data into NJRP date items (i.e., re-coding offenses, summing charges, and ranking offenses by level of seriousness). In contrast, when agencies within a state vary widely in their coding conventions, the level of complexity and effort needed to develop the same set of NJRP data items would be enormous, requiring detailed, extensive data cleaning and processing by county and/or by court.

Turning to NJRP *sentencing items*, the biggest challenge will likely be that of locating data on the minimum and maximum incarceration sentence, as well as the maximum incarceration term for the most serious offense. Of the 11 agencies that provided some type of documentation, it appeared that only 2 of the agencies recorded data in this way. The other agencies simply had a field for incarceration term imposed with no range. Sentencing Commissions often have fields for a "recommended" range, but it is not clear whether this range comes from court proceedings or out of the commissions' own calculations to ensure guideline compliance. Other areas where the NJRP sentencing data are tenuous include sentences to community service or treatment as well as flags for consecutive and suspended/deferred sentences.

In states where the primary sources of NJRP data are likely to be the *criminal record depositories*, like Texas and Georgia, another important data integration issue may arise. Most respondents from the agencies in these states consider their data systems "offender" based since they are used mostly as look-up systems of individuals for law enforcement and employers. As a result, additional processing may be necessary to transform the data into a *case-level format* that would be usable for NJRP analytical purposes. An important exception to this rule is the NY Division of Criminal Justice Services, which maintains an analytic dataset which is structured very much like the ones produced by state sentencing commissions.

3.1.4 Implications for the NJRP Re-design

We began this subsection by posing the overarching research question about what capacity the states have to provide historical NJRP-type data. While our pilot study was not designed to answer this question definitively, a number of insights emerged. There is reason to be optimistic that currently a number of states have high-quality NJRP-type data available statewide. We found that four states (NY, MN, VA,WA) likely collect statewide data that appear to map into historical NJRP sentencing and conviction items. Five other states (GA, ID, PA, TX) collect at least some statewide sentencing and conviction data that would be useful to NJRP. Unfortunately, there is no guarantee that these data cover all NJRP-items. We uncovered a number of challenges to data integrity (e.g., non-standardized coding, partial geographic coverage, collection of only a subset of all NJRP data items) that would limit the utility of the data and/or involve an intensive effort to process the data into NJRP-usable formats. Another barrier is an acquisition process that could require months (up to a year) to secure approvals and process the statewide data extractions.

We found considerable variation in coverage, completeness, standardization, acquisition timing, and documentation among statewide judicial data systems. Taken together, these findings suggest that, at least in the near future, it would not be realistic to expect exclusive statewide data collection for NJRP. It is reasonable to expect that a number of states may be in a position to provide statewide NJRP data within the next few years. And, over time, the proportion of these states should increase. In the larger NJRP re-design perspective, it may be prudent to conceive of statewide data capture as a multi-year (decade or more) endeavor, involving a transition process over which successive iterations of NJRP increasingly exploit the existence of statewide data as states themselves migrate towards fully integrated centralized judiciary data systems. This simply will not occur in a single year or even in a five-year period (unless federal legislation is passed requiring such centralization, though that would introduce other problems).

The process of acquiring statewide data is of especial concern. The capricious nature of the approval process and the timing of request fulfillment suggest aggressive planning action by BJS well in advance of NJRP data collection. Specifically, we recommend that state agencies be identified and contacted a year or more in advance of NJRP data collection to identify and overcome access barriers and more generally to establish a solid working relationship. This will facilitate the data acquisition process once the NJRP is launched.

Another recommendation is more along the lines of *strategic planning*. Given that standardization is often lacking in state systems, it would be worthwhile to engage state agencies and their staff with access to judicial data systems at conferences, workshops and other venues *on a continuous, ongoing basis*. BJS could promote standardization, full geographic coverage, completeness of data, etc. and possibly provide technical assistance with the expectation that subsequent requests for NJRP-type data extracts could be handled easily and quickly. Moreover, BJS could instill and promote the value

of the NJRP to states so that state agencies become stakeholders rather than entities upon which a burden (of NJRP data collection) is placed.

As expected, states with sentencing commissions offer a good starting point in the search for statewide standardized data. However, New York, a large, populous state appears to have a sophisticated, integrated statewide judicial data system. This suggests that other large states (e.g., CA, FL, IL) may have statewide data availability.

Our pilot study was highly informative in terms of testing and adapting our interview protocols and calling strategies to elicit information from state agencies. We quickly realized the futility of attempting to *definitively* establish the availability of NJRP data items at a statewide level. This would have required data extractions, processing and analyses. We chose instead to adopt quality ratings on three dimensions (e.g., completeness, coverage, standardization, effort needed to recode etc). If we retain our coarse quality ratings, it will be possible to conduct a data collection operation on all 50 states and DC in order to get a "good sense" of the availability of statewide NJRP data.

3.2 Capacity of states to integrate defendant criminal histories into NJRP data

The second research question focuses on expanding the breadth of NJRP data to include criminal histories (in addition to the historically collected data on sentencing and convictions). To address this, we discuss our findings from the pilot study with separate subsections devoted to data offered by *sentencing commissions* and *state criminal records depositories*.

3.2.1 Sentencing Commissions

All four sentencing commissions (i.e., MN, PA, WA, VA) were found to maintain datasets that *included* criminal histories. However, we sampled from only those commissions that gathered and maintained judicial data. As such, this finding may simply reflect our subjectively selected sample and not generalize to sentencing commissions in other states.

While the "good news" is that criminal histories were gathered by all four state sentencing commissions, the 'not-so-good news' is the considerable variation in criminal history items captured by state. For example, Washington fails to include calculated items such as the *count of offenses* across all offense types. Instead, WA records information that is much more detailed: the *state statutes associated with each prior offense* along with *qualifier codes* to mark offenses that might be subject to special sentencing considerations. These include flags for attempted offenses, juvenile offenses, and "washed" (expunged from record) offenses as well as offenses involving deadly weapons or other firearms, conspiracy, solicitation, or sexual motivation.

MN and PA state sentencing commissions maintain counts of different types of offenses in their criminal history data bases. But the data are not standardized. Minnesota's database provides variables for the *total number of prior felonies, misdemeanors, and*

juvenile offenses as well as tallies of *offenses by sub-type* (person, property, drug, criminal sex, and other types of offenses). While Minnesota's database only captures prior convictions, it distinguishes between "true" priors that occurred as a part of a separate *previous incident* from other priors that include offenses occurring as part of the *current incident* but for which a sentence had already been issued. Pennsylvania's criminal history data is much more detailed.

Pennsylvania's maintains a complex scoring system based on criminal history in order to determine appropriate sentencing ranges. The PA data include separate counts by specific *type of offense* (i.e., murder, burglary, endangering the welfare of a child) as well as by *event* (adjudication vs. conviction). All variables are arranged by the number of points assigned under the criminal history scoring system. There are other aggregated variables available, such as fields for the total number of prior felony III, misdemeanor 1 and 2 adjudications.

The required sentencing guidelines worksheets in Virginia ask questions about criminal history; however, the data is stored in terms of points under a scoring system—not as clear counts or codes for prior offenses. Points are assigned in several different ways. Some questions ask for a sum of the maximum penalties (in years) for the offender's 5 most recent and most serious offenses and then the resulting sums are used to assign points based on pre-determined ranges. Other questions ask for the total number of prior convictions or adjudications of a certain type and similarly the totals are used to assign points using fixed ranges. The questions also vary substantially by the type of offense for which the worksheet is being completed. In other words, the same criminal history information is not collected from everyone; it is tailored to their qualifying offense.

The bottom line is that the four sentencing commissions in our pilot study collect and maintain criminal history data. However, as we might expect given our experience with NJRP data, criminal histories are not being captured in a uniform, systematic way. Instead, each state tailors the data to their own needs. The implication to the NJRP re-design effort is that considerable processing will be required to format and code criminal histories so that they can stored and maintained as part of NJRP. A bigger issue is the extent to which our findings from these four subjectively selected state commissions can be extrapolated to other state sentencing commissions. Technically, such a generalization is inadvisable; more research is needed.

3.2.2 State Criminal Records Depositories

Beside a state sentencing commission, another natural starting point to explore the statewide availability of criminal histories is a ***state criminal records depository***. Generally, criminal records depositories are structured differently than sentencing commission data. Depository system data tend to be *offender*- or *arrest*-based; they typically lack separate fields for *counts* such as the "number of prior convictions" because each arrest/adjudication is its own record. In contrast, state sentencing commissions tend to have *summary criminal history profiles* (e.g., counts) for a given *court case* resulting in a felony conviction. The NY Division of Criminal Justice Services, however, is an exception. It has an analytic dataset that has separate criminal

history variables that are comparable to the ones used in sentencing commissions' databases. For example, there are counts of prior offenses by seriousness (misdemeanor and felony), by event (arrest and conviction), and by age of offender (juvenile and adult).

Table 6 exhibits the states and agencies housing their criminal records depositories, showing how often they are updated. Criminal records depositories located in states that also have sentencing commissions (MN, VA, and WA) all pull their disposition and/or sentencing data from centralized court system databases. In contrast, only 2 of the 5 other states get their judicial data from a centralized source. This almost guarantees problems with standardization and difficulty in creating a standardized set of criminal history items for NJRP.

Table 6. Criminal History Depositories in our Pilot Study by State* Showing the existence of a State Sentencing Commission and the Percentage Electronically Updated

State	State has Sentencing Commission?	Criminal History Depository	Type of source for judicial data	% Electronic Updating
MN	Yes	Bureau of Criminal Apprehension	centralized	90%
VA	Yes	State Police	centralized	98%
WA	Yes	State Police	centralized	53%
NY	No	Division of Criminal Justice Services	centralized	99%
GA	No	Bureau of Investigation	decentralized	92%
TX	No	Bureau of Investigation	decentralized	80%
AZ	No	Dept of Public Safety	decentralized	unknown
ID	No	Bureau of Criminal Identification	centralized	"most"

*** PA is not included because we were unable to converse with the Pennsylvania State Police.**

Most criminal records depositories use court data to electronically update their arrest records. (Washington is a notable exception to this because only cases with single charges can be electronically linked; all cases with multiple charges must be entered into the system by hand.) However, despite fairly robust links between courts and criminal records agencies, state sentencing commissions in our sample do *not* obtain their criminal history data from their respective state data repository. The histories are generated using *sentencing guideline worksheets* completed by judges, or they tap the pre-sentencing worksheet that probation and correctional officers administer to offenders. It is unclear why this occurs. Regardless, it leaves open a greater possibility of human error and/or a mismatch between worksheet and depository-record-based criminal histories.

To summarize, we find that criminal histories can be obtained from either state sentencing commissions or from criminal records depositories. The former are "court case based" while the latter are "offender or arrest" based. It is reasonable to expect that capturing statewide data for some or all 50 states would require a combination of one or the other data source. Both would require considerable processing in order to standardize and/or transform the data into items usable for NJPR, including the transformation into a common unit of analysis. Thus, principal issue in the NJRP redesign effort would be the

level of effort required to conduct such a transformation. Such an assessment is beyond the scope of this pilot but deserves to be acknowledged.

As for the feasibility of incorporating criminal histories statewide into a re-designed NJRP, we conclude from the pilot study that "the jury is still out." It appears that statewide criminal history data are available from most states but that considerable processing would be required to put it in a standardized form suitable for use by NJRP.

3.3 Accessibility of pre-sentencing report data

The NJRP redesign effort is considering the capture of pre-sentencing data to expand its analytic capacity and value. We explored this issue and now report our findings. Only 2 of the agencies in our sample reported obtaining "pre-sentencing" data: the Minnesota and Virginia Sentencing Commissions. Both receive data from *Pre-Sentencing worksheets* administered by probation or correctional officers to offenders. Virginia actually maintains a separate PSI database for all convicted felons either sentenced to probation or state prison time (those serving jail sentences are excluded). This accounts for approximately 80% of felony offenders. Unfortunately, the PSI database is only accessible with a court order which may render this data inaccessible. In its regular sentencing guidelines database, however, Virginia includes regular employment and marriage status as part of a pilot nonviolent risk assessment for drug, fraud, or larceny offenses that a few jurisdictions administer. Regular employment and education level are included as part of a sex offender risk assessment, though it is unclear what percentage of felons this affects.

Minnesota also retrieves data from the pre-sentencing forms. However, detailed demographic, economic, and social status data are not available. Only the conventional date of birth, race, ethnicity, and sex variables are present.

The respondent at the Pennsylvania's Sentencing Commission did not specifically mention getting "pre-sentencing" data. However, their codebook reveals several variables related to the employment status of the offender as well as their hourly wage rate.

For our sample of states, agencies report that the demographic data that they collect *is unreliable or missing altogether*. For court case management systems, demographic data are often not required and serve little purpose for day-to-day operations. Criminal records agencies have these issues multiplied twofold since they house data from at least 2 different sources—law enforcement and the courts. *They often receive complaints from individuals who check their records that their gender or race is incorrect*.

Data on arrest (i.e. arrest date which appears on the current NJPR dataset) is generally available only from criminal records depositories. *None of the sentencing commissions contain this information*, though they often have the date of the incident itself. We suspect the same is true for state court administration data, though this is difficult to verify without quality codebook information.

Our findings suggest a pessimistic outlook for incorporating pre-sentencing data into NJRP at the state level. Only a few states out of the nine we studied reported having such data. And those that did have the data reported problems with missing and illogical entries. This does not bode well for attempting to gather statewide data for NJRP, and we believe that these reports are sufficiently pessimistic that *we recommend further exploration of this issue be dropped.* At this time it does not seem feasible to collect statewide pre-sentencing data.

3.4 Accessibility of other dispositions and misdemeanor data

3.4.1 Other dispositions

Besides presentencing data, the NJRP redesign effort is considering other judicial data such as misdemeanors to incorporate into statewide data collection. We also explore this issue with our pilot study states. With the exception of sentencing commissions, all agencies that maintain case-level data with other dispositions available besides just convictions. However, because of the difficulty of obtaining quality documentation from these agencies, we only have two examples of the types of dispositions available. Table 7 provides a summary of our findings from these two state agencies: NY Criminal justice Services and the VA Supreme Court.

Table 7 entries suggest that some dispositions overlap, most notably the codes for dismissed cases and acquittals. The NY Criminal Justice Services office (left hand column of Table 7), however, is much more specific in terms of the type of conviction as well as adding in codes unique to "Youthful Offenders."

The findings and data quality from our limited exploration of "other dispositions" is sketchy at best. Clearly, standardized uniform "other disposition" data are not being routinely captured by the state agencies in our pilot sample. That portends a low likelihood that such data is available statewide for other states. And given the variety of items that fall into this 'residual' category of data, we see little promise that NJRP could benefit from additional exploration of availability for these types of items. Thus, *we are comfortable recommending that this part of the NJPR redesign effort be suspended.*

3.4.2 Misdemeanors

The final research question for our study involves consideration of adding misdemeanor data to a redesigned NJRP. We summarize our findings from questions about the availability of misdemeanor data in Table 8. Unlike earlier findings for NJRP-type data, here we see that few agencies have misdemeanor conviction and sentencing data. New York is the only state with high-quality items for misdemeanor offenses. In contrast, 6 states had low quality data on misdemeanors and another 7 had no usable data at all.

Standardization re-emerges as a challenge and this is shown in Table 9. We also see that agencies do not uniformly capture the same *types* of misdemeanors. Some restrict their data holdings to only "finger-printable" misdemeanors, while others specify A & B or 1 & 2, or "targeted" misdemeanors which could serve as enhancements for sentencing.

Table 7. "Other Dispositions" Captured by Two State Agencies

NY Criminal Justice Services	VA Supreme Court
conviction, type unknown	Guilty
conviction by verdict	Not Guilty/Acquitted
conviction by plea	Dismissed
Youthful Offenders (YO) adjudication, type unknown	Not True Bill
YO adjudication - verdict	Mistrial
YO adjudication - plea	Nolle Prosequi
acquitted	Remanded
dismissed	Resolved
no true bill	Sent/Probation Revoke
prosecution declined	No Indictment Presented
unknown favorable disposition	
removed to family court	
covered by/consolidated	
other	
interim disposition -- info only	
no disposition information	

Table 8. State-level Availability of Misdemeanor Conviction & Sentencing Data by State and Quality Level

	Quality				
State	**High**	**Medium**	**Low**	**Unusable**	**Total**
AZ	0	0	2	0	***2***
ID	0	1	1	0	***2***
GA	0	0	1	1	***2***
MN	0	2	0	1	***3***
NY	1	0	1	0	***2***
PA	0	2	0	0	***2***
TX	0	0	1	1	***2***
VA	0	1	0	2	***3***
WA	0	1	0	2	***3***
Total	***1***	***7***	***6***	***7***	***21***

Table 9. Agency-level availability and quality of conviction and sentencing items for misdemeanors

			Quality Dimensions		
State	**Agency**	**Type of misdemeanors**	**Coverage**	**Complete-ness**	**Standar-dization**
High Quality					
NY	Division of Criminal Justice Services	Fingerprintable	+	+	+
Medium Quality					
ID	Idaho State Judiciary	all	+	+	+
MN	Admin Office of Courts	all	+	+	+
MN	Bureau of Criminal Apprehension	Gross & targeted	+	+	+
PA	Sentencing Commission	all	+	+	+
PA	Admin Office of Courts	all	+	+	+
VA	Supreme Court	all	+	+	+
WA	Admin Office of Courts	all	+	+	+
Low Quality					
AZ	Admin Office of the Courts	all	o	o	o
AZ	Dept of Public Safety	hi	o	+	
ID	Bureau of Criminal Identification	all	o	o	o
GA	Bureau of Investigation	Fingerprintable	+	+	+
NY	Office of Court Admin	all	+	o	o
TX	Bureau of Investigation	Class A & B	o	+	+
Incomplete: Conviction But No Sentencing Data					
VA	State Police	1 & 2	+	+	+
No Systematic Data on Misdemeanors					
MN	Sentencing Commission	n/a	n/a	n/a	n/a
VA	Sentencing Commission	n/a	n/a	n/a	n/a
WA	State Police	n/a	n/a	n/a	n/a
WA	Sentencing Commission	n/a	n/a	n/a	n/a
No Case-Level Data					
GA	Admin Office of Courts	n/a	n/a	n/a	n/a
TX	Admin Office of Courts	n/a	n/a	n/a	n/a
*** key to table cell entries: + = high; + = medium; o = low; n/a = unavailable**					

Several of the agencies with promising NJRP felonies data have more limited misdemeanor data. In most of these cases, this is because the structure of the court system combined with state reporting regulations do not ensure the universal gathering of conviction and sentencing information for misdemeanors. In Georgia, misdemeanor cases are tried in two different kinds of courts. Only Superior Courts cases must be reported to the Georgia Bureau of Investigation. This may exclude misdemeanor cases heard in more than 150 courts statewide. Similarly, only county and district courts in Texas report to the criminal records depository. As a result, all the misdemeanors that might be tried in the municipal and justice courts (over 1700 locally funded courts statewide) are excluded from their data. Even New York's Court Administration Office which maintains high quality data on "fingerprintable" misdemeanors is subject to these coverage issues. The state's town and village justice courts are under no obligation to report "non-fingerprintable" misdemeanors. As a result, data on these cases are very spotty. Codes are not standardized and data are often missing altogether.

Taken as a whole, the types of misdemeanors, associated coverage and quality ratings shown in Table 9 demonstrate that statewide misdemeanor data is quite heterogeneous. Adding the expressed concerns with data quality lead us to conclude that there is a low likelihood of capturing usable statewide misdemeanor data for NJRP.

4. Conclusions

This pilot study was conducted to explore two types of inquiry:

- ***methodological*** – how to best gather information from states in order to ascertain the existence, availability and quality characteristics of statewide judicial data
- ***substantive*** – what insights can be gleaned from a limited selection of states about the statewide NJRP and related judicial information.

Our subjective sample of states covered what we believed to be the diversity of state judicial system environments across the U.S., including states with sentencing commissions, populous and rural states, and states with previous BJS data collection experience (via SCPS). The pilot study provided many insights into how various state agencies and sentencing commissions maintain judicial information.

4.1 Methodological Conclusions

With respect to our *methodological objectives*, the pilot study served its purpose well. The Appendix includes a semi-structured interview protocol that was revised and enhanced in light of our interaction with multiple state agencies and conversations with agency staff. We affirmed that the strategy of commencing exploratory discussions with state sentencing commissions provides an efficient portal to state government for the purpose of identifying statewide judicial data systems.

We discovered that the process of soliciting technical details about statewide information systems is often cognitively challenging to state agency respondents. Their knowledge about the data system is user-based and experiential, rather than that of a designer or data manager. This leads us to rely more heavily on hardcopy or electronic documentation supplied by state staff. Unfortunately, because state employees have limited knowledge about what constitutes adequate technical documentation, we were typically provided piecemeal glimpses into the state judicial data systems. Nonetheless, we managed to assemble sufficient information from each state to make general assessments of each state's judicial data systems.

Our methodology for collecting information was necessarily restricted in scope due to time and resources limitations. We were only able to determine the potential for states to provide NJRP and other related judicial data. We accomplished this through "quality" ratings across three dimensions, and combined these in an overall quality rating. A more rigorous assessment of the data systems would necessarily involve acquisition and processing of prototype judicial data. This would allow discernment of item missing data rates, wild codes, standardization of coding, and levels of effort for processing needed to transform state records into NJRP-suitable data.

Our methodological findings lead the following recommendation:

> ***Recommendation:*** Given the limited resources available to this project, we believe that NJRP would benefit most from a continued effort to canvas the remaining states in order to (a) identify and document the agencies and contact persons for each state in preparation for a re-designed NJRP that seeks statewide judicial data; and (b) assess and assign quality ratings on the three dimensions we developed that reflect the potential for acquiring NJRP data from those agencies.

An alternative strategy is to stay with these nine states and conduct a more in-depth assessment of judicial data systems so that state-specific inventories of NJRP items can be developed. We do not believe that there are sufficient resources and time in the project to adopt the alternative strategy.

4.2 Substantive Conclusions

4.2.1 NJRP Data

We found that statewide NJRP sentencing and conviction data appeared to be available for at least four of the nine states in our pilot. While it is reasonable to expect that statewide data availability will expand over time, an "all-statewide" NJRP will most likely not be feasible for at least 10 years.

Limited time and resources did not permit us to draw definitive conclusions about the availability and quality of NJRP data items for our pilot states. But we were able to acquire and review sufficient documentation to establish that some states most likely have NJRP data available and it is readily accessible. In conjunction with our previous finding we propose a transition strategy for consideration in the NJRP re-design:

> ***Recommendation:*** Until such time that all states maintain accessible statewide judicial data systems, a re-designed NJRP should plan for and exploit the availability of statewide data for those that do. In effect, this is tantamount to the adoption of a phase-in strategy towards statewide data collection over time (from the current county-based strategy). It will allow for *testing* on a relatively small sample of states in the initial years of hybrid county-statewide NJRP data collection. This should prove highly useful for developing "best practice" and efficiency as statewide judicial data systems increasingly become available.

Undoubtedly, there will be obstacles posed by the acquisition process itself. The pilot study revealed multiple reports of anticipated turnaround delays in obtaining approvals and in processing requests for extractions from state agency data systems. Under a statewide NJRP data collection paradigm, we believe that this could be effectively addressed though a strategic planning process, thus leading us to recommend:

> ***Recommendation:*** We recommend that state agencies be identified and contacted a year or more in advance of NJRP data collection to identify and overcome access barriers and more generally to establish a solid working relationship. This will facilitate the data acquisition process once the NJRP is launched.

We also found barriers associated with the lack of standardization and documentation of data elements among state judicial data systems. This leads to the following recommendation:

> ***Recommendation:*** We recommend that NJRP program consider an outreach strategy to engage state agencies and their staff, possibly by providing technical assistance, with the objective of promoting standardization, full geographic coverage, completeness of data, etc. in state agency judicial data systems for subsequent NJRP-type data extract requests. Moreover, BJS could instill and promote the value of the NJRP to states so that state agencies become stakeholders rather than entities upon which the burden of NJRP data collection is placed. Venues for such outreach activities would include professional conferences, workshops, etc. The objective would be to conduct outreach on *a continuous, ongoing basis* as part of the NJRP research program.

4.2.2 Criminal Histories

The pilot study revealed that criminal histories are available from state sentencing commissions and criminal records depositories. The data would require considerable processing to standardize and transform into NJPR items. A concern for the NJRP redesign effort would be the level of effort for such processing.

We conclude that "the jury is still out" regarding the statewide collection of criminal histories for the NJRP. We offer no recommendation other than to continue researching the extent of processing that would be necessary to develop a standardized set of criminal history data suitable for NJRP research goals. The cost well may outweigh the benefit.

4.2.3 Pre-sentencing Data

Our exploration of the availability of statewide pre-sentencing data resulted in a pessimistic outlook for NJRP. Only a few states out of the nine reported having such data. The states with such data reported severe data quality issues. Based on these limited reports we developed the following recommendation:

> ***Recommendation:*** We recommend further exploration of the statewide availability of pre-sentencing for NJRP be ***dropped*** from further consideration. In all likelihood only a few states in the nation will have pre-sentencing data available statewide, and those that do will not vouch for its integrity. At this time it does not seem feasible to collect statewide pre-sentencing data.

4.2.4 Other Dispositions and Misdemeanors

Other dispositions. We conducted a limited exploration into the statewide availability of *other dispositions* for NJRP. Our limited findings suggest data for *other dispositions* are sketchy at best. Standardized data are not being captured routinely by the states in our pilot sample. That portends a low likelihood that such data is available for other states. As such, our recommendation is:

> ***Recommendation:*** Given the variety of items that fall into the 'residual' category of data we call "other dispositions," we see little promise that NJRP could benefit from additional exploration of availability for these types of items. Thus, we recommend that this aspect of the NJPR redesign effort be suspended.

Misdemeanors. Our pilot study findings on misdemeanors suggest sizeable heterogeneity in the number and types of items stored, as well as coverage and quality. We conclude that there is a low likelihood of capturing usable statewide misdemeanor data for NJRP. As such, we recommend:

> ***Recommendation:*** We recommend that efforts to further explore the availability of misdemeanor for NJRP be suspended.

APPENDIX A: INTERVIEW PROTOCOL

Screener

Hi, I'm calling you from the Urban Institute in Washington, D.C. We're currently working with the Bureau of Justice Statistics to think about strategies for making data collection more efficient.

Currently, NJRP (National Judicial Reporting Program) collects data on felony convictions and case histories. Data is gathered from selected counties within states, but BJS wants to know how much of these data states like yours might be collecting already and what other data might be available to expand the scope of the NJRP.

- **What is your role at _____ (name of agency)?**

- **How long have you been in this position?**

We really want to talk to both someone who can give talk to us about:
  - The sources of court case data available at your office
  - Its structure and scope
  - Data quality and processing
  - Data requests and access

- **Would you be the right person to speak to? Is there someone else who might want to join us?**

- **When would be a convenient time to schedule an interview?**

Overview

To start out, it would be wonderful if you could give us a brief overview of the data your agency has access to, records, or maintains on judicial proceedings?

- **What data bases does your agency access or maintain?**
  [probe: name if have one]

Sources of data

Thanks for that introduction. Now I'd like to ask you a few more questions to better understand the data that you have.

- **Where does your data come from?**

  - [If they get their data from courts]
    - **Which type of courts? (i.e. traffic, local, county, appellate)**

  - [If they get their data from law enforcement]

        - **Which type of agencies? (i.e. local police, sheriffs, state police, probation officers, etc.)**

- **Do all of these courts/agencies report their data to you?**
    - [If not] **Which ones do not?**
        - **How would one go about getting data from <u>THOSE</u> agencies?**

Scope of data

Now I'd like to get a sense of the kinds of data that are contained in the database(s) your office uses.

- **Which types of cases do these agencies report to you?**
  Probe: juvenile vs. adult
  Civil vs. criminal
  Level of offense (traffic, summary, misdemeanor, felony)

- **At what points in the judicial process do they gather these data? (i.e. arrest, filing of charges, case disposition, sentencing)**

Data Quality

- **How often do agencies transmit the data to you?**

- **What percentage of their cases do they report to you?**
    - How does this vary by agency?

- **What method do they use to transmit the data to you? (i.e. shared database, hardcopy, etc.)**

- **How standardized is the data?**
  [Probe: Same fields, some coding across agencies]

- **How complete is the data that appears on the forms?**
    - Are there particular fields that have high missing rates?

- **How much data cleaning do you do once you receive the data?**
  [Probe: back and forth with agencies for missing items, incorrect coding]

Data Requests and Access

- **What are the steps needed to access this data?**
  [Probe: request form, review process, fees if applicable]

- **How long does this process typically take?**

- **At what level is the data available? (offense, offender, aggregate only)**

- **In what format is this data available?**

- **How big is the database?**

I know that you get your data from [AGENCY NAME/TYPE].

  - **Through them, do you have access (data use agreements) to:**
    - **Other types of cases?**
    - **Data from other parts of the judicial process?**

  - If so, can you refer me to someone at that agency?
- **Is your data linkable to other databases that the state maintains?**
  - How easy or difficult is this process. Please describe.

- **Which items are included in the database?**
  - Can you send me a list of fields and codes or a codebook?

APPENDIX: STATE SPECIFIC PROFILES

State Profiles

In this section we provide synopses of our findings from each state included in this study.

Washington

Courts. Washington is unique because it represents a state that operates a *decentralized* court system yet maintains a *centralized* court repository run by the Administration Office of the Courts (AOC)[3]. Most counties enter their court case management information into databases that upload in real time to the data warehouse. Pierce county uses its own system and does double-data entry (i.e., once in their system, and again in the state's central database). Seattle Municipal Court uses an FTP site to upload its data to the data repository nightly.

All types of courts (i.e. appellate, superior, municipal, and district courts report) share data on all of their cases. This includes both adult and juvenile criminal and civil cases from the time charges are filed to closing out of the case after sentencing is performed and fines paid, etc. The only exception to this is Seattle Municipal Court, which only sends records for criminal cases to the data warehouse. The standardization of Washington's data varies substantially depending on the type of courts. For sentencing, superior courts only have a check box for sentences to prison and then a catch-all text field for all other sentences. In contrast, sentencing is all code driven for courts of limited jurisdiction and much easier to extract.

Criminal records. The Washington State Police (WSP) manages the state criminal record depository, which links arrest to disposition data. Law enforcement agencies are only required to report arrests for adult felony charges, but many voluntarily report arrests for misdemeanors, gross misdemeanors, and juvenile cases. The procedure for submitting arrest data is highly automated; around 98% of it is sent using Live Scan machines. The remainder, mostly from smaller rural agencies, are submitted on fingerprint cards that sent to WSP within 15 days of arrest are hand-entered into the database. As per their annual audit, WSP estimates that law enforcement reports nearly 100% of felony arrests. The arrest data is highly standardized and very complete. Live Scan sends back error messages automatically if data is missing and WSP staff follow-up immediately with individual law enforcement agencies if items are not filled in or are unintelligible. Arrest data includes demographic information, date of arrest, and initial charges.

Arrest data for all cases with a single charge (53%) are linked to disposition data through the AOC's data warehouse; the rest of the dispositions for arrests with multiple charges are sent in hard copy from the AOC to WSP where they are entered by hand. Since only felony dispositions are required to be sent to the criminal records depository, disposition

[3] In the state of Washington, most but not all courts in WA report data to the AOC. Only 13 very small city courts fail to report to the central data warehouse because of low case volume (100 cases/year).

data only comes from the courts with jurisdiction over these offenses-- superior and appellate courts. As a result, dispositions for voluntarily reported misdemeanor arrests will be missing for all cases that were tried in lower municipal and district courts. Disposition data consists of simple information on conviction and non-conviction. No data on subsequent sentencing is available. If applicable, there is information on current custody status that is transmitted through Live Scan when an offender enters or leaves a correctional facility.

Sentencing Commission. The state of Washington maintains a database for all adults convicted of felony offenses to monitor state compliance with sentencing guidelines.[4] All data is submitted to the Sentencing Commission by individual superior or appellate courts using hard copy sentencing worksheets about every other week. Each county has a different form; they vary in length from 5 to 20 pages. The staff that hand enter the data at the Sentencing Commission take care to clarify all missing and unintelligible fields with the individual courts as needed. As a result, the data submitted are very complete. The Sentencing Commission also regularly cross-checks its case counts with the Administrative Office of the Courts' own statistics and follows-up with courts that have inconsistent counts to ensure all mandated felony cases are being captured. Records for offenders sentenced to state prison are matched once a year to corrections data in order to calculate actual time served. NJRP data items included on the record include demographics, the conviction charges, and the type of sentence imposed. There is also information on criminal history, though it is tailored to the state's own sentencing guidelines and scoring systems.

Arizona

Courts. Arizona has a decentralized court system and, as such, maintains fairly limited case-level data. However, its Administrative Office of the Courts (AOC) has made some moves towards a more centralized case management system. In total, about 50 percent of cases statewide are reported using "Aztec databases" that allow individual courts to enter data into standardized fields and upload them nightly to a central data warehouse. However, because the primary purpose of judicial statistics is to ensure the proper payment of judges and monitor caseloads, there are few required fields. The "Aztec databases" most accurately capture case filings and whether cases have been disposed or pended. Most other fields are optional and have little or no code standardization. For the remaining 50 percent of cases, the AOC requests aggregate counts of cases filed, cases pended, and cases disposed. Two of the state's largest counties-- Pima and Maricopa—only submit these summary statistics.

Criminal records. All law enforcement agencies are required to submit arrest information for felony, domestic violence, and several other key offenses set out by statute to the Department of Public Safety (DPS). However, law enforcement routinely submits arrests for other offenses as well. The state is in the process of automating this process and has recently installed Live Scan technology in many agencies, though a

[4] They also maintain a parallel database on juvenile felony and misdemeanor convictions. However, access to these data is highly restricted. Only aggregate data can be requested.

significant amount of arrest data still comes in hard copy. The data fields used are highly standardized and very complete. Error messages are generated automatically when there are missing Live Scan data and DPS staff follow-up with individual agencies when there are issues with hard copy fingerprint cards. DPS depends on law enforcement agencies' relationships with their local courts to get the corresponding disposition data. The respondent was unsure of how much of that data tends to be missing. Sentencing data is in theory available on the records, but is often missing.

Georgia

Courts. In Georgia, the court system is decentralized and courts are under no legal obligation to report to the Administrative Office of the Courts (AOC). As a result, the AOC is not only limited to aggregate statistics on case filings, but it also has difficulty getting courts to provide even these data. The respondent shared that they have to ask for everything with "please and thank you" and then still find some courts completely unresponsive. The only courts that consistently report their statistics are superior courts which, by nature of being the exclusive courts of jurisdiction for felony cases, have to report case level disposition data to the state's criminal records depository. Even courts that do respond to AOC's requests for data often send back unwieldy .pdfs of all the case filings for year. AOC staff then have to go through the document line-by-line and tally the numbers. To further complicate matters, because the state agency does not "own" the data, the AOC has to send all numbers back to the courts for them to review and edit at will.

Criminal records. The Georgia Bureau of Investigation (GBI) serves as the criminal records depository for the state. It receives data on all arrests for felony and "fingerprintable" misdemeanors. Participation of law enforcement agencies is near universal, with only a few smaller agencies not reporting. Most of the arrest data (98%) is received electronically from Live Scan machines, making it highly standardized and generally very complete. GBI staff enter the remainder of the arrest data which comes in on standardized fingerprint cards by hand; they confer with individual agencies as needed to clarify issues with the data. Like other arrest data, they generally include demographics, arrest date, and charges. Georgia's superior courts, which have exclusive jurisdiction for felony cases, are required by state statute to report their case dispositions and corresponding sentences to GBI. These data are updated using an automated electronic system for about 92% of GBI arrests. The rest of the disposition data is sent to GBI in hard copy for data entry. Notably, the GBI criminal records database links to corrections and also provides detailed information on the custody status of offenders.

Texas

Courts. The Texas court system is decentralized and structured in a very complex way. Consequently, there is no common data kept on a case level. However, courts do submit some monthly aggregate statistics on their criminal caseload; compliance is nearly 100% for all types of courts.

Criminal records. By law, all law enforcement agencies must report data on all arrests for felony and class A and B misdemeanor offenses to the central criminal records depository, managed by the Texas Dept of Public Safety (TXDPS). Compliance is very high, with an estimated 99% of these arrests documented. The process by which they are submitted is mostly automated. Approximately 90% of the arrests come in directly from Live Scan machines; the remainder are submitted on fingerprint cards by individual agencies. Case disposition and sentencing data are subsequently added on to the record. About 80% of these data are uploaded to an FTP site by county and district courts, which hear all felony cases and many of the misdemeanors; the rest arrives in hard copy and is entered by hand into the database. At any given time, about 25% of dispositions that are missing, but it's hard determine if this is a matching issue or simply an indicator that the case is still in process. The court data is highly standardized.

Pennsylvania

Courts. Pennsylvania has a unified court system and maintains detailed case-level documentation of all cases in the state, with the exception of those heard in Philadelphia Traffic Court. Unlike the other states which main one central data system, Pennsylvania actually maintains separate databases for appellate courts, courts of common pleas, and magisterial courts.[5] The fields and the types of offenses in each are unique to the type of court entering the data. For example, the magisterial court database includes traffic offenses and preliminary hearings for misdemeanors and felonies, while the court of common pleas database will be the exclusive domain for all juvenile cases as well as the main trial court for all criminal offenses. Data within each of these three databases is very complete and highly standardized. Offense tracking numbers (OTNs) and state ID (SIDs) are linkable to the State Police's criminal record system; though they have not automated this linkage.

Criminal records.

Sentencing Commission. To fulfill its responsibility of ensuring statewide compliance with sentencing guidelines, the Sentencing Commission maintains a database containing pertaining to the felony or misdemeanor conviction of adults and juveniles tried as adults. They download their data monthly from the statewide Justice Network[6] interface where counties enter their data on an on-going basis. Because only common police courts of record are required to report sentencing worksheet data, Philadelphia Municipal Court data are excluded. Data are regularly cleaned, de-duplicated, and checked for errors and the Sentencing Commission sets out a new analytic dataset annually for research purposes.

[5] Because of idiosyncrasy in the system, the Philadelphia Municipal Court, which is a counterpart of other low level Magisterial Courts, appears with the Courts of Common Pleas database.

[6] Many agencies use and upload data to this common database, including: Administrative Office of Pennsylvania Courts (AOPC), Pennsylvania State Police (PSP), Juvenile Court Judges' Commission (JCJC), the Pennsylvania Commission on Sentencing (PCS), and the Pennsylvania Chiefs of Police Association. As a result, common charge codes and protocols for data exchange have been established that could facilitate linkages.

New York

Courts. New York has a unified court system and maintains detailed case-level data. However, because its court structure is so complex, it maintains many different databases (e.g., for appellate courts, general criminal courts, civil courts, family courts, and surrogates' courts). To further complicate matters, there are a total of 6 different kinds of courts that try criminal cases: supreme courts, county courts, district courts, city courts, town and village justice courts, and the criminal court of the city of New York. Some of these courts fall under the direct jurisdiction of the state of New York and, consequently, use the same real time comprehensive case management database to document all of their cases. However, the rest—mostly town and village justice courts— use a completely different case management system and are record reliable data in certain fields for "fingerprintable offenses" because of their obligation to do so for the statewide criminal records depository. All other data fields they submit to the Office of Court Administration (OCA) are usually either completely missing or use codes unique to each individual court that are difficult to interpret or standardize across the state.

Criminal records. All law enforcement agencies in the state submit arrest data for all "fingerprintable" offenses to the Division of Criminal Justice Services (DCJS); most of this is done electronically. The respondent estimated that about 90% of applicable arrests make it into the database. New York is unique because DCJS and OCAactively work together to manage and improve their state's data. In most other states, the criminal records agency gets disposition data from the courts, but no back and forth relationship exists. In New York, the relationship is bi-directional. The courts get data from criminal records and vice versa; and both agencies are involved in improving data quality. In addition, there are units dedicated to following-up with jurisdictions about missing or inaccurate data. As a result, the New York criminal records are of particularly highly quality. They are complete, standardized, and relatively free of duplicates. Not only does DCJS use the data for standard criminal records searches for law enforcement and employers, it produces analytical datasets quarterly now and will be producing them monthly once their data is migrated to a new system they are putting in place.

Idaho

Courts. Idaho has a unified court system and a simplified court structure. As a result, all of their courts use a common case management database for all of their cases, including all civil and criminal (from traffic to felony) cases for both adults and juveniles. Cases are tracked from the time charges are filed to the time the cases are closed. This includes detailed information on conviction and sentencing, as well as demographics and all court transactions. Individual courts input their data directly into their county's database and all the counties (44) upload these data to a central state data warehouse. While the fields themselves are all standardized, there has been quite a lot of variation by county in the codes used. For this reason, the Idaho State Judiciary (ISJ) is in the process of standardizing all codes. Their staff run reports periodically to evaluate the accuracy and

completeness of the data and then do some work with courts and counties to correct problems.

Criminal records. Idaho's Bureau of Criminal Identification (BCI) is the central depository for criminal records. Law enforcement agencies report all non-traffic offenses; half of these submissions are done using LiveScan and the other come in hard copy. Reporting compliance is relatively low. The respondent estimated that their database captures about 50-70% of arrests "at best." Most of the corresponding disposition data is downloaded daily from the centralized ISJ database; the rest is received from individual courts and entered by hand. There is a lot of missing data because there is no common agreement on required fields. Similarly, coding varies widely. There is little or no standardization across agencies. In addition to disposition data, probation and parole status are also included on the arrest records when applicable.

Minnesota

Courts. Minnesota's unified court system and maintains a common case management system for all district courts in the state. These courts hear both civil and criminal cases, from traffic to felony offenses, for both adults and juveniles. Appellate and municipal courts, which deal mostly with local code violations, are not included in the state's database. While all courts use the same fields and codes, the respondent pointed out that "there are a million data entry errors – disposition incorrectly, sentencing incorrect, hearing, trials recorded incorrectly, just about everything in there could be incorrect." When they run reports, staff at the Administrative Office of the Courts (AOC) find that every court or county has their own way of doing things. There are some statewide court practice policies, but not for everything. AOC dedicates two staff exclusively to data quality and working with courts to correct errors, but there are still many issues.

Criminal records. The Bureau of Criminal Apprehension serves as the central depository for all criminal records. All law enforcement agencies are required to submit arrest information for all felonies, gross misdemeanors, and "targeted" misdemeanors that would be considered grounds for sentencing enhancements for repeat offenders. About 98% of these data comes in electronically through Live Scan machines; the rest come in hard copy and are entered by hand. The respondent was fairly confident that most arrests are being captured, but about 10% of incoming dispositions tend to go unmatched because law enforcement agencies have not reported the arrest. All disposition data is uploaded directly from the central AOC database. The Department of Corrections also submits electronic Live Scan records for all offenders sentenced to incarceration when they come into custody. The Department of Probation then manually submits discharge information once the offender is released. Data fields and codes are highly standardized. The biggest problem is getting all the dispositions to match. In an effort to improve the match rate, they've put in place a system called "auto notify" that sends a message to law enforcement when a court disposition comes in that is not matchable on 4 different indicators. Agencies can then go in using a direct interface to update or correct their own data to fix the problem.

Sentencing Commission. This agency maintains analytic datasets for all adults and juveniles tried as adults who have been convicted of felonies in the state of Minnesota. There are two main sources for this data. The first is the Department of Probation, which electronically submits required data on demographics, the conviction offense, and criminal history from mandatory sentencing worksheets. The remainder of the data, mostly further detail on sentencing, is electronically downloaded from the central AOC database. There is very little missing data and staff at the Sentencing Commission do a lot of data cleaning and following up with agencies when there are issues.

Virginia

Courts. The Virginia Supreme Court maintains a central database for all district courts in the state, and all circuit courts, excluding Alexandria, Fairfax, and Virginia Beach, which maintain their own separate data systems. All types of cases are included: civil, criminal (from traffic to felony), and adult as well as juvenile. All fields in the database are the same but there is quite a lot of disagreement across courts in terms of coding practice. Some cities are even unresponsive when the Supreme Court attempts to address the coding challenges. Charge codes are particularly problematic. They are to be entered in a free text field and are very messy.

Criminal records. The Virginia State Police (VSP) is the main depository for criminal records. Law enforcement must report all adult felony and misdemeanor 1 & 2 arrests to VSP and all juvenile felonies and offenses "adjudicated delinquent" for juvenile; but some law enforcement submit data on all their arrestees because they fingerprint everyone. The respondent estimated that about 5% of the required felony and misdemeanor offenses are missing because they are based on indictments, direct indictments, or failed appeal warrants which don't have an initial arrest record. About 94% of the arrest data comes in through Live Scan and is highly standardized and reliable. Almost all (98%) of disposition data is updated automatically through the Virginia Supreme Courts central database. However, three of Virginia's largest jurisdictions (Fairfax, Alexandria, and Virginia Beach) have to have their data entered by hand because they haven't migrated to the centralized system yet. About 15% of disposition data is missing at any given time, but it is difficult to determine when it's actually missing and when the case has simply not been disposed yet. VSP has a missing dispositions section that works with law enforcement and the courts to resolve these issues. There is no sentencing data explicitly on the records, but offenders who are sentenced to probation or incarceration have their custody status posted to their criminal record.

Sentencing Commission. The Virginia Sentencing Commission actually maintains 2 separate data sets. The first is its Sentencing Guidelines database which contains conviction, sentencing, and criminal history data for about 95% of all adult felonies. Five percent are still not covered by the sentencing guidelines and for this reason do not appear. All of the data comes from hard copy forms filled out by judges at the time of sentencing in each of the individual courts where felonies are tried and sent in daily. All of this data is entered by hand by Commission staff. Because reporting to the Sentencing

Commission is mandated by law, the data is very complete and highly standardized. Its other database, the Automated PSI (pre-sentencing), houses data from 80% of felons—those sentenced to probation or serving time in state prison; those serving time in local jails are missing. The Virginia Department of Corrections transmits the data electronically to the Sentencing Commission. They include more detailed data on arrest, demographics, criminal history, conviction, and sentencing. However, these data can only be released with a court order.