

Preliminary Analysis on AFS and Justice Metrics Data

Hanlong Fu, Ph.D.

As part of the validation study of LS/CMI, a risk and needs assessment tool currently used by Los Angeles County Probation Department, the Research Section (RS) of Systems Accountability Bureau (SAB) performed a preliminary assessment of the sample data shared by Information Services Bureau (ISB) on March 15, 2021. There are two sets of data: one is a monthly sample of data prepared for the Adult Field Services (AFS), containing selected data elements from the Adult Probation System (APS). The other is a group of four datasets called the Justice Metrics (JM); it is a CEO data initiative with collaboration among various County departments.

AFS

The provided AFS dataset is a monthly snapshot of cases that are under Probation's active supervision. Due to its cross-sectional nature, it does not have probationers' case history. However, it does contain most of the data fields that the LS/CMI validation study is likely going to need, including the LS/CMI total score, case-related information (e.g., case grant date and supervision expiration date), biographic and criminal background, and treatment-related information. It does not have LS/CMI subdomain scores, precluding any analysis on specific risk factors that LS/CMI assesses. If recidivism is defined as return to probation, a longitudinal dataset (not the monthly sample data currently provided to the RS) containing information over time may be adequate for a limited assessment of the relationship between LS/CMI total score and recidivism.

Justice Metrics (JM)

Different from the AFS dataset, the JM datasets provide 10-year worth of longitudinal information (2011-2021) on 1) assessments, 2) dispositions, 3) cases, and 4) court events. Together, the JM datasets are more complete than the AFS dataset because they provide case history information, subdomain scores, and other related information. As a preliminary assessment, the RS only analyzed the case and assessment data as they are most relevant to the proposed study. Until Feb 23, 2021, there were 143,798 unique individuals¹ and 183,546 cases, of which 109,927 (76.44%) individuals had no recidivism², and 33,871 (23.55%) had at least two or more recidivism, or probation case records and 9,762 (6.79%) individuals had three or more recidivism, or probation case records. An individual could be identified by the x-number while a case is denoted by the court case number. In addition, each case has an associated case granting date and supervision closing date. The assessment dataset contains x-numbers, assessment dates, and LS/CMI total and sub-domain scores. Until March 11, 2021, there were 141,525 LS/CMI assessments in total, of which 136,713 were completed after Jan 1, 2011.

To establish a relationship between LS/CMI and recidivism necessitates a data merge between the case and assessment datasets. The complexity lies in the fact that an individual could have multiple cases associated with different charges on the same case granting date. With each case,

¹ Unique individuals are defined as unique x-numbers.

² Recidivism is defined as having at least two probation contacts documented in the Justice Metrics Sample data.

there could be multiple LS/CMI assessments depending the duration of the supervision. On top of that, LS/CMI assessment dates are not in sync with case granting dates as there is typically a delay between client induction and LS/CMI assessment. The following tables illustrate the complexity of the data structures.

Case Data		
Person 1	Case 1	Case granting date 1
Person 1	Case 2	Case granting date 1
Person 2	Case 1	Case granting date 2
Person 2	Case 2	Case granting date 2
Person 2	Case 3	Case granting date 3

Assessment Data		
Person 1	Assmt 1	Assmt date 1
Person 1	Assmt 2	Assmt date 2
Person 1	Assmt 3	Assmt date 3
Person 2	Assmt 1	Assmt date 4
Person 2	Assmt 2	Assmt date 5

For the preliminary analysis, case and assessment data were joined through this logic in R³: An assessment was matched with a corresponding case if their x-numbers matched, and the assessment date was within the case granting and supervision expiration dates. This meant that each case could have multiple assessments associated with it as a supervisory period could be long enough to accommodate many assessments. However, a potential issue was that supervision starting and closing dates were not mutually exclusive from case to case, meaning that one assessment could be matched with multiple cases. This could result in undesirable redundancies.

Single-level Analysis

With above limitations in mind, the RS tried to simplify the data structures to enable a preliminary analysis, with the understanding that a more complete and in-depth analysis is forthcoming once the datasets are satisfactorily joined through an agreed-upon mechanism. As such, the joined data from above was collapsed into a cross-sectional dataset in the following way. Within an individual, the number of months passed since previous case was calculated from case granting dates and recidivism was indexed. Binary variables for recidivism within 6 months, 1 year, 18 months, 2 years, 3 years, and 3 years and more were subsequently created from months passed. Then the binary variables, along with x-numbers were separated from the main dataset to allow for collapsing. For this subset, the maximum was selected within an individual for each binary column. In other words, if there was a 1 for an individual, only 1 would be left for that column. If there were only 0s, then 0 would be left. The remaining case dataset was then collapsed by selecting the first case per individual. In the end, the subset containing the binary outcomes was re-joined with the remaining collapsed columns through x-number.

Binary logistic regressions were conducted on the resulting cross-sectional dataset to assess the relationship between LS/CMI total score and recidivism within certain timeframes. The results indicated that there is a small but significant positive relationship between LS/CMI total score and recidivism⁴. In particular, LS/CMI seems to be able to predict recidivism that occurs three years later ($\beta=.038^5$, $p<.001$) better than recidivism that occurs within six months ($\beta=.025$, $p<.001$). In

³ R is a popular, open-source statistical software.

⁴ Beta coefficient for recidivism within 6,12, 18, 24, 36, and more months was .025, .032, .030, .037, .038, .037 respectively

⁵ A beta coefficient of .038 ($e^{.038} = 1.04$) means that one point increase in LS/CMI score is associated with 4% increase in the odds of reoffending.

other words, the longer the period between offenses, the better the prediction. This could be because clients are less likely to have their probation terminated or revoked due to violation of probation terms and conditions or return to criminal behavior while they are under supervision, and/or because LS/CMI is a relatively stable measure of risk levels. A preliminary analysis such as this precludes a definitive answer. But this analysis does provide baseline values against which later analyses could benchmark.

Multilevel Analysis

While the single-level analysis could provide baseline values against which later analysis could compare, it is very limited because it fails to capture within-individual variation. As mentioned above, each individual could have multiple cases and associated LS/CMI assessments. A multilevel analysis can accommodate this complexity by factoring in the hierarchical nature of the data when computing the estimates. As such, results from a multilevel analysis should in theory be more accurate than single-level analysis.

Using the same data-joining logic explained earlier, a multilevel dataset was created by matching LS/CMI assessments with corresponding cases and dispositions within individuals. In the resulting dataset, an individual could have multiple cases with varying starting and ending supervision dates (often overlapping dates between cases). Within an individual, a LS/CMI assessment was matched with a case so long as the assessment date was within a case's starting and ending supervision dates, also known as exposure period.

Recidivism

In the single-level analysis, recidivism was defined as return to probation for simplicity and convenience. In the multilevel analysis, recidivism was defined according to the Countywide Criminal Justice Coordination Committee (CCJCC)'s return to custody framework⁶. Based on this framework, there are five qualifying events for recidivism: 1) convictions, 2) felony arrests, 3) misdemeanor arrests, 4) revocations of community supervision, and 5) flash incarcerations. Conviction data are captured by the Trial Court Information System (TCIS) used by the Los Angeles Superior Court. Arrests are captured by the Sheriff's Department's Automated Justice Information System (AJIS). Flash incarcerations and revocations are retained by Probation's APS data system. Due to data availability, this multilevel analysis could only address flash incarcerations and revocations as measures of recidivism. Specifically, this study defines recidivism as any of above qualifying events within 36 months after the exposure period.

The JM datasets include three types of probationers: post-release community supervision (PRCS), mandatory supervision (split sentence), and formal probation. Although flash incarcerations can be used on individuals under mandatory supervision, a deputy probation officer would need a waiver of a hearing from the probationer prior to the initiation of flash incarceration. For that reason, flash incarcerations are rarely used in practice in this probationer subgroup. We do not include them in our analysis. Additionally, revocations are defined by a combination of disposition codes in APS. Due to the lack of a complete list of disposition codes related to revocations, we were not able to replicate the exact figures of probationers brought back to custody for revocations as reported by the 2020 Public Safety Realignment Evaluation Study.⁷ For that reason, we do not

⁶ <https://ceo.lacounty.gov/wp-content/uploads/2021/02/2021-2-17-Justice-Metrics-Framework-Baseline-Report.pdf>

⁷ <http://ccjcc.lacounty.gov/LinkClick.aspx?fileticket=st1N9mx0nBs%3D&portalid=11>

include revocations in this preliminary analysis. We do plan to re-examine revocations as a recidivism measure in a follow-up analysis.

Figure 1: Flash Incarcerations among PRCS Cohorts

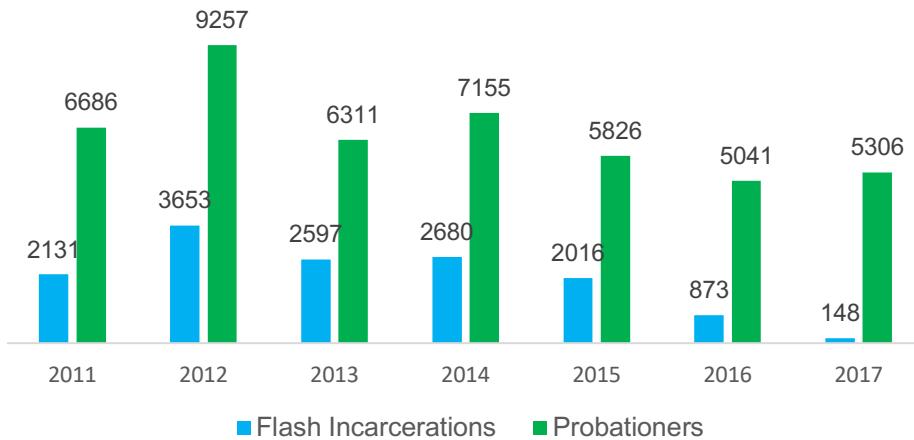


Figure 1 shows flash incarcerations as a percentage of probationers under PRCS within each yearly cohort. Each cohort is tracked for three years after the exposure period. Any flash incarcerations that happened within this time frame were counted. Overall, the data shows a sharp decline over time in using flash incarcerations among PRCS populations. This is consistent with Probation's policy to reduce the use of flash incarcerations to address non-compliance with supervision terms, warrants or violations to hold individuals in custody for court hearings.

Multilevel logistic regressions were conducted to examine the relationship between LS/CMI total score and recidivism as measured by flash incarcerations. Compared to the single level analysis, beta coefficients from the multilevel models were attenuated by roughly 50%, which was understandable given that single level models are known to overestimate parameters. Despite their size differences, similar conclusions could be drawn from the results of both types of analysis. Multilevel analysis shows that LS/CMI total score could modestly predict recidivism (flash incarcerations). From 2011 to 2017, the average beta coefficient was .016, which meant that one point increase in LS/CMI total score was associated with 1.6% increase in the odds of being brought back into custody through flash incarceration. The relationship was the strongest in the 2015 cohort when one point increase in LS/CMI total score was associated with 2.1% increase in recidivism, and the weakest in 2017 cohort when one point increase in LS/CMI total score was associated with 1% increase in recidivism.

Table 1: LS/CMI Predicting Recidivism by Cohort

	2011	2012	2013	2014	2015	2016	2017
LS/CMI beta coefficient	.018***	.017***	.011**	.017***	.021**	.018***	.009**
Odds of Recidivism	1.8%	1.7%	1.1%	1.7%	2.1%	1.8%	1%

** p<.01 *** p<.001

Conclusion

In conclusion, a preliminary analysis of the AFS and Justice Metrics datasets shows that (1) AFS data, in its longitudinal form, may allow for a limited validation study on the relationship between LS/CMI total score and recidivism defined as probation supervision; (2) Justice Metrics datasets are more complete because they contain 10-year worth of case history and LS/CMI data; (3) single-level analysis shows that there's a very small but positive relationship between LS/CMI and recidivism defined as probation supervision/contact; (4) multilevel analysis shows that there's a small but positive relationship between LS/CMI and recidivism defined as flash incarcerations.

In this preliminary analysis, only flash incarceration was used as a measure of recidivism due to data availability. As revocations with remand to custody is readily available within APS, a follow-up analysis will include both flash incarceration and revocation as measures of recidivism. A more complete analysis of recidivism and LS/CMI would need to include convictions and arrests data, which are not currently available within APS.

Appendix A: AFS Data Elements

1. Case number (X-NMBR)
2. Case grant and expiration dates
3. DPO/SDPO name and number
4. Bio and demographic info: Name, sex, race, DOB, gang affiliation, PDJ, SSN, address, drug use, employment status
5. Risk classification
6. LS/CMI date and total score (no sub-domain scores)
7. Supervision type (mandatory, PRC, formal probation)
8. Primary charge level and code (O, I, M, F)
9. Treatment program date and type
10. Release from custody date
11. Hub name and orientation date
12. Sex registration, arson, weapon, narcotic, and obey condition flag