Appendix of “The impact of overbooking on a pre-trial risk assessment tool”

Kristian Lum
kl@hrdag.org
Human Rights Data Analysis Group

Chesa Boudin
chesa.boudin@sfpd.gov
San Francisco Public Defender’s
Office

Megan Price
meganp@hrdag.org
Human Rights Data Analysis Group

ACM Reference Format:

Kristian Lum, Chesa Boudin, and Megan Price. 2019. Appendix of “The impact of overbooking on a pre-trial risk assessment tool”. In *Proceedings of FAT*2020: The ACM Conference on Fairness, Accountability, and Transparency (FAT*2020)*. ACM, New York, NY, USA, ?? pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 DE-DUPLICATION AND RECORD-LINKAGE

To operationalize the definition of “fair” booking charges we have adopted, we must track down the outcome of the case associated with each PSA. This requires linking the court records, which contain outcome status by charge, to the PSA data. In this process, we only consider records for which the full PSA was completed. We exclude any PSA records missing inputs for any of the PSA components, such as arrest date and any of the PSA predictions (FTA, NCA, NCVA). We also check that each complete record in the PSA data is unique in a process called de-duplication. We define two records as pertaining to the same incident if both records are for the same individual (have the same SFID), were recorded on the same day, and list the same charge(s). If there are multiple records of the same individual on the same day with the same charge(s), we retain only one copy.

To match the PSA data to the court data, for each PSA record, we first identify all court numbers (a unique identifier of a case—one individual can have multiple court numbers) that list the same SFID and an arrest date within one day before to two days after that listed in the PSA record. This determines the initial set of potential matches. For the vast majority of PSAs, there is only one case in this set, and it is declared the match. If there are multiple court numbers that meet these criteria, we select the court number(s) with the charges that most closely match those listed on the PSA using the following procedure.

We determine whether each potential match in the court records contains the charge code listed as the top charge on the PSA form. If only one of the potential matches does, this is declared the match. If multiple potential matches contain the charge code listed first on the PSA, we repeat this using the second listed charge code if one exists. In most cases, there is only one court number that matches on both first and second charges and it is declared the match. If

there are multiple, we declare all of them to be matches, and all charges associated with the matched court numbers are included in our calculations.

For those records for which this process found no match, we manually search through the court records to identify an appropriate match. There are several idiosyncratic reasons why matches were not found using our automated procedure. For example, in one case, the day and month of arrest had been transposed in the PSA data relative to the court data, i.e. they used different date formats. We consider this and other similar errors to be simple recording errors, and we manually determined a match on a case-by-case basis. In the end, despite our best efforts to find a match in the court records for every PSA, there were 31 PSA records for which we were unable to definitively find a match, and these were dropped from the analysis.

2 VALIDATION OF OUR REPRODUCTION OF THE PSA

Using the matched data, we confirm that our code is able to reproduce the results of the PSA based on the booking charges listed in the court data. We wrote a function that takes as input the PSA’s NCA and FTA predictions (which are not a function of booking charge so will not be sensitive to changes to this input), the factors used to calculate the NVCA flag (which is dependent on booking charges), and the list of booking charges as taken from the court records. Sometimes, this differs from the booking charges listed on the PSA form. As output, it returns our calculations for the NVCA flag, an indicator of a charge-based exclusion, an indicator of a charge-based bump-up, and the final recommendation.

2.1 NVCA flag

We follow the procedure to calculate the NVCA flag as outlined in the PSA Portal. One of the inputs to the NVCA flag is whether the booking charge is violent. For each PSA in our dataset, we compare the booking charges listed in the court records to the California PSA List of Violent Charges. If any of the booking charges are present on the list, we calculate the NVCA flag as though the current offense is violent. All other non-charge-based inputs to the NVCA flag are taken directly from the original PSA form. We find that our calculation of the NVCA flag agrees with that listed on the PSA form for 99.3% of the cases in our dataset.

2.2 Charge-based exclusions and bump-ups

In cases where there was a discrepancy between our code and what appears on the human-administered PSA, we did a manual review. In most cases, the discrepancy was due to a slightly different set of booking charges listed in the court records than were listed on the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

FAT*2020, January 27–30, Barcelona, Spain

© 2019 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

PSA form. In other cases, we believe it was due to inconsistency in how the PSA administrators handled ambiguous instructions. For example, recall that a charge-based bump-up is indicated if the booking charge “involved the Use of a weapon, caustic chemical, flammable substance, or explosive.” There were some cases in the PSA data where an individual received a bump-up based on a charge of wielding an imitation firearm (417.4 PC). Other cases that included this charge did not receive a bump-up, which we believe indicates some inconsistency in how bump-ups were applied with respect to this charge due to the ambiguity around whether an imitation firearm should be considered a firearm. Similar inconsistencies were noted around the interpretation of whether carrying a loaded firearm (25850(A) PC) counted as “Use of a weapon” for the purposes of a charge-based bump-ups.

2.3 Final recommendation

Finally, we compare our calculation of the final recommendation to the recommendation given in the PSA data. We find an agreement of 97.4%, which is slightly lower than for the other components. Further inspection reveals that the majority of the disagreements occur in the case when the FTA prediction was 5 and the NCA prediction was 4. This combination of NCA and FTA scores corresponds to the cell in the decision-making framework (see Figure ??) where additional decision rules are applied, presumably by a human. For cases that have this combination of NCA and FTA scores, if any of the booked offenses are a felony or a violent misdemeanor as indicated by the PSA Violent Offense List, the recommendation is “Release Not Recommended.” Otherwise, the recommendation is “SFPDP-ACM.”

We also manually reviewed several of these disagreements. In many cases of disagreement, there was a felony charge that did not appear on the violent offense list. Under the rule as we interpret it, any felony leads to the increase in level. We suspect that this rule was sometimes interpreted to mean that only violent felonies triggered the additional increase. In any case, for those cases that do not fall into this cell of the DMF, regarding the final recommendation, our reproduction is in agreement with the PSA data 99.5% of the time.

3 VIOLENT CHARGE LIST

The California PSA violent charge list is the list of charges used by the PSA to define violence. This list contains over 200 charge codes, so we do not reproduce the whole list here. Charge codes are separated on the list into several categories. For each category of charge that appears on the list, we give some example charges and their associated charge codes. We have tried to select charges that are representative of those that appear in each category without being redundant.

4 CONSISTENCY OF RACIAL CLASSIFICATION IN COURT DATA

In order to break down this analysis by the race of the arrested person, we first explore how consistently individuals were assigned to each race category. The court records contain a field called race which takes values B, C, F, H, I, J, U, W, and in some cases it is missing. We believe these stand for Black, Chinese, Filipino, Hispanic,

Indian, Japanese, Other, Unknown, and White, respectively. Some of these categories (F, I, J) were very rarely used.

To calculate our measure of consistency, we first restrict the analysis to all individuals who had more than one arrest record in the dataset. For each race category (shown in rows of Table ??), we filter the data to include only individuals who are listed as that race at least once in the data. Using this subset, for each individual we calculate the percent of their records that were classified as each possible race. We then take an average across all individuals. This serves as our measure of how consistently the individuals are categorized into each group. The i th entry of Table ?? is then the average percent of records that were classified as the j th race category among all individuals who were classified as the i th race category at least once. The rows do not sum to 100 as one might expect, due to the omission of rows and columns corresponding to rare categories and missing values. Looking at the row labeled as B , we see that for an individual who was categorized as B at least once, on average 98.4% of their records were labeled as B and 0.4% of their records were labeled as W . This shows a high level of consistency in designating people as B . However, for a person labeled as H at least once, 35.7% of their records were indicated as W , leading us to conclude there is a reasonably high degree of variability in how people who may be considered H are classified. By looking at the diagonal of this table, we see that the only category that is highly consistent under this measure is B with approximately 98% agreement, and so for the analysis presented here, the only race categories we consider are B and non- B , which we assume to be Black and non-Black.

Category	Example Charge	Example Code
Other offenses against public justice	Resist Police Officer: Cause Death/Serious Bodily Injury; Threatening witnesses, victims, or informants; Advocacy to kill or injure peace officer	148.10(A) PC F; 140(a) M; 151
Homicide	Murder first degree; Gross vehicular manslaughter while intoxicated; Solicit to commit murder	187(A) PC F 1; 191.5(A) PC F; 653F(B) PC F
Mayhem	Mayhem; Torture	203PC F; 206PC F
Kidnapping	Kidnapping; False imprisonment of a hostage	207(A) PC F; 210.5 PC F
Robbery	Robbery: first degree; Carjacking	211 PC F 1; 215(A) PC F
Attempts to kill	Assault on a public official; Attempted murder of public official; Obstructing railroad track, punishment; Throwing missile at common carrier with bodily harm	217.1(A) PC F; 217.1(B) PC F; 218.1 F; 219.1 F
Assaults with intent to commit felony, not murder	Assault with intent to commit a felony; Assault to commit a felony during the commission of first degree burglary	220 (A)(1) PC F; 220(B) PC F
False imprisonment and human trafficking	False imprisonment; Human trafficking	236 PC M; 236.1
Assault and battery	Assault; Assault on a peace officer of a school district; Assault on a highway worker; Battery on school employee; Battery against police, emergency personnel, etc.; Assault w/firearm on person; Shooting at an inhabited dwelling, vehicle, etc.	240 PC M; 241.4 M F; 241.5 M; 241.6 PC M; 243(B) PC M; 245(A)(2) PC; 246PC F
Rape, abduction, carnal abuse of children	Rape: victim incapable of consent; Sex intercourse w/a minor less than 18; Rape spouse by force/etc; Attempted lewd acts/w/child und 14yrs	261(A)(1) PC F; 261.5(A) PC F; 262(A)(1) PC F; 664/288 (A) PC F
Abandonment and neglect of children	Willful cruelty to child; Injuring a spouse, cohabitant, fiancé, boyfriend, girlfriend or child's parent; Inflict injury upon child	273 A(B) PC M; 273.5 (A) PC M; 273 D(A) PC M
Bigamy, incest, and crimes against nature	Incest; Sodomy: person under 18; Lewd/lascivious acts on dependent adult w/force; Oral copulation; Harmful material sent w/intent to seduce minor; Sexual penetration w/foreign object: victim drugged; Continuous sexual abuse of child	285 PC F; 286(B)(1) PC F; 288 (B)(2) PC F; 288 A(A) PC F; 288.2 (A) PC F; 289 (E) PC F; 288.5 (A) PC F
Other injuries to person	Poisoning	347(A) PC F
Crimes against elders, dependent adults, persons with disabilities	Cause harm/death elder dependent adult; Elder abuse: victim 70/or older	368(B)(1) PC M; 368(b)(2)(B) PC F
Crimes against the public peace	Rioting; Exhibit firearm. Drawing, exhibiting, or using a firearm	404(A) PC M; 417 (b) PC M
Crimes and penalties	Violate civil rights by force/threat	422.6(A) PC M
Arson	Arson causing great bodily injury; Aggravated arson; Causing fire of inhabited structure/property	451(A) PC F; 451.5 (A) PC F; 452 (B) PC F
Offenses by prisoners	Assault by a life prisoner; Escape from custody by force and violence	4500 F; 4530(A) PC F
Prevention and abatement of unlawful activity	Use of a weapon of mass destruction causing death	11418 (B)(2) PC F
Destructive devices and explosives generally	Use of destructive device and explosive to injure/destroy; Explosion causing death	18740 F; 18755 (a) F

Table 1: California PSA Violent Offenses List: representative charges by category

Race Designation	B	C	H	O	U	W
B	98.40	0.10	0.10	0.20	0.30	0.40
C	0.50	85.60	0.00	5.90	1.60	3.20
H	0.50	0.00	48.40	4.20	2.80	35.70
O	2.90	14.90	2.90	46.00	3.40	25.90
U	7.70	2.60	9.00	3.80	53.80	16.70
W	0.40	0.10	2.90	1.90	0.40	91.60

Table 2: The expected rate per 100 records at which individuals which each race designation and more than one record were classified as each possible designation (columns). For example, for individuals who were classified as U at least once and had more than one record available, on average those individuals had 8 % of their records listed as B.