Supplementary Information for

Police agencies on Facebook overreport on Black suspects.

Ben Grunwald, Julian Nyarko, and John Rappaport

Julian Nyarko
Email: jnyarko@law.stanford.edu

This PDF file includes:

  Supplementary text
  Figures S1 to S14
  Tables S1 to S4
  SI References
Supplementary Information Text

S1. Estimating Number of Exposures to Race-Crime Posts. To estimate the number of exposures to the 98,448 race-crime posts, we summed the number of followers of each agency’s page at the time each post was published and multiplied that sum by 10%, an approximation of the average “organic reach” of Facebook posts during our sample period (1-2). Organic reach measures the proportion of a page’s followers who are exposed to each of the page’s posts in their timeline or by visiting the page (3). The resulting figure totals 197 million. These followers, in turn, shared the posts more than 11 million times with their own Facebook friends, many of whom, presumably, did not follow the pages from which the posts originated.

S2. Equivalence of difference-in-means and regression analysis. In the main text, we define overexposure as the difference between (weighted) means. To obtain accurate confidence intervals, however, we needed to estimate cluster-robust standard errors that account for the correlation in the error structure at the post and arrest level. To that end, we fit a weighted linear regression of the form $I(black) - I(post) + FE_{agency}$, where $I(post)$ equals 1 if the observation is a post and 0 if it is an arrest; $I(black)$ equals 1 if at least 1 suspect is Black and 0 otherwise; and $FE_{agency}$ are agency-level fixed effects. Weights for post $i$ of agency $j$ are defined as

$$w_{dij} \left( \sum_{i=1}^{N_{rj}} w_{dij} \right)^{-1}$$

Importantly, the coefficient on $I(post)$ obtained from this regression is identical to our difference-in-means measure of overreporting $\theta$. But the regression allowed us to compute cluster-robust standard errors that account for the correlation of errors.

S3. Matching Facebook Pages with UCR Agencies. As described in Materials and Methods, we compiled our list of U.S. law enforcement Facebook pages in two stages. In the first stage, we used CrowdTangle to obtain an initial list of 12,071 pages.

We began the second stage by matching these Facebook pages with their respective UCR agencies. Our matching strategy primarily leveraged page and agency names and state locations. The UCR provides state information on all agencies but obtaining state information for pages was more involved. We extracted state information from the page name (e.g., “chester county sheriff’s office, pa”) or from the address, area code, or “More Info” segments in the page’s “About” section. Where otherwise not available, we retrieved state data by manually reviewing the page and any website linked in the “Additional Contact Info” segment of the “About” section. In total, we obtained state information on 99.6% of pages.

To make page and UCR agency names as consistent as possible, we removed some punctuation, expanded abbreviations (e.g., “Dept.” to “Department”), made common terms more consistent (e.g., “sheriffs’ office” to “sheriff’s office”), and removed some uninformative words (e.g., “the”, “official”, “detectives”). We then matched pages and agencies if their names were identical and they were located in the same state. In total, 10,282 pages and 9,694 agencies were matched via name and state. We then used address information to match pages and agencies that, though not identical in name, were identical in address and state. Through this method, we matched an additional 264 pages and 260 agencies.

We then sought to match the remaining 1,525 pages manually. We reviewed each page and any corresponding website in the page’s “Additional Contact Info” section to locate useful identifiers for the agency. We then changed the page name in our dataset to match the agency’s name in the UCR and matched by name and state. In rare cases, the UCR contained multiple agencies with the same name in the same state. We matched these cases manually—relying on name,
state, and county—to assure match quality. Ultimately, we matched an additional 491 pages and 441 agencies through this process. In total, we matched 11,037 Facebook pages to 10,395 UCR agencies, 9,738 of which reported arrest information to the UCR during the study period.

As noted in the Materials and Methods, after matching those pages to their corresponding UCR agencies, we identified all the remaining UCR agencies that remained unmatched, report arrest information, and have a UCR population estimate greater than zero. We then manually searched on Facebook for pages associated with each of these agencies. We found 1,919 additional pages through this process, which brought our final list to 13,990 pages, of which 12,971 were matched to 12,324 agencies.

**S4. UCR arrest data.** We obtained agency-month-level data on the number of Part I UCR arrests reported by law enforcement agencies from 2010 to 2019 from (4). The UCR reports arrests using four racial categories: “American Indian,” “Asian,” “Black or African American,” and “White.” It does not reliably collect information on arrests of Hispanic individuals. If an arrest involves multiple suspects, each arrestee is counted separately.

We took several steps to clean the data. First, a very small number of agency-months report 10,000, 20,000, 30,000, 40,000, 50,000 or 60,000 arrests for a given crime type. Because these figures are improbably large, we assumed they actually reflect missing values. Second, in 0.7% of rows, the total number of arrests for at least one crime type is smaller than the sum of race-specific arrests for that crime. In these cases, we replaced the total number of arrests with the sum of race-specific arrests. Third, some agencies do not report arrests in every month and the UCR does not provide a variable to identify nonreporting agency-months. We generally assumed an agency reported no arrest information in a given month if it was missing a row for that month or if it reported zero arrests for all crimes—including “Part II” UCR offenses for minor crimes, which are extremely common—in that month. For a few agencies, however, we applied a different rule based on our review of the data. The Chicago Police Department regularly reports an annual arrest count in December and no other months. Similarly, in 2010, most agencies in Alabama report arrest data only in December. For these agencies, we assumed that the number of arrests reported in December covered the entire year. Finally, we disaggregated the agency-month-year level dataset so that each row represents an arrest.

One of our robustness checks reweights agencies based on population size (Fig. S4). As noted in the main text, measuring agency population is challenging because many law enforcement agencies serve overlapping jurisdictions. As a plausible but imperfect solution, we use the UCR’s agency-level population estimates, which assign a population estimate to municipal police departments based on their cities’ respective 2010 Census population counts. For county departments, the UCR subtracts the total population assigned to each municipality in the county from the county’s total population estimate. For a small number of other departments—like university, school, and parks departments—the UCR assigns a population of zero.

**S5. NIBRS offense data.** We extracted two datasets from the National Incident-Based Reporting System (NIBRS) for 2010 to 2019 from (5). The first is an offense-level dataset. To parallel the hierarchical offense structure of the UCR, which we built into our algorithms to detect offenses in Facebook posts, we aggregated this data up to the incident level by identifying the most serious UCR offense in each incident. We then dropped incidents in which the most serious offense was not a UCR Part I offense. The second dataset is at the offender level and has a row with demographic information for every reported suspect associated with each incident. Incidents for which no suspects were reported have a row in this dataset but all columns are missing information. We merged the information about the most serious offense into the offender-level dataset based on incident number, Originating Agency Identifiers (ORI), and year.

A substantial number of incidents lack suspect information and a substantial number of reported suspects lack race information. Table S4 shows the proportion of rows in the offender dataset, by
most serious offense, that lack race information. The number is high for property offenses, ranging from 55% to 66%, and relatively low for violent offenses, ranging from 11 to 17%. We drop rows for suspects who lack race information and rows associated with incidents in which no suspect information was reported at all. We then compute agency-level measures of the proportion of reported offenders who are Black based on the remaining rows.
Fig. S1. Overexposure regression coefficients, alternative weighting schemes. Estimates for average overexposure using alternative measures of exposure. (Top-Left) No exposure weights; equivalent to overreporting. (Top-Right), (Bottom-Left), (Bottom-Right) Weighted on followers, interactions, and shares, respectively. To calculate accurate standard errors that preserve the uncertainty of agency-level estimates for overexposure, we fit a weighted linear regression (see S2). Error bars show 95% confidence intervals with standard errors clustered at the agency level.
Fig. S2. Decay functions for distance weighting. Relationship between distance and assigned weight for three alternative decay parameters. At a distance of 300 miles, the weights are infinitesimal. Because posts so heavily discounted do not impact our analysis, and for computational efficiency, we set the weight for posts beyond a 300-mile radius to 0.
Fig. S3. Overexposure regression coefficients by offense type with different decay functions. Estimates for average overexposure for different decay parameters. To calculate accurate standard errors that preserve the uncertainty of agency-level estimates for overexposure, we fit a weighted linear regression (see S2). Standard errors are clustered at the post-level as well as at the level of the agency publishing the post. Left reflects our default decay parameter. Middle and Right reflect alternative decay parameters.
Fig. S4. Overexposure regression coefficients by offense type with population weights. Estimates for average overexposure, where each agency is weighted by the size of the population in the agency’s jurisdiction.
Fig. S5. Overexposure regression coefficients by offense type (NIBRS). Estimates for average overexposure using offense data from NIBRS instead of arrests.
Fig. S6. Overexposure regression coefficients by offense type with different decay functions, arrest posts only. Estimates for average overexposure for alternative decay parameters. The method of analysis was identical to that for Fig. S3 but only posts describing arrests were used.
Fig. S7. Map of overexposure across the United States. Agency-level, distance-weighted overexposure to Facebook reports on Black suspects. Each dot represents an agency that reports arrest data to the UCR.
**Fig. S8. Histogram of agency-level overexposure.** Dotted navy line indicates the median overexposure and orange lines show the 20th and 80th percentiles.
Fig. S9. Overreporting by county-level characteristics. This figure is analogous to Fig. 5 in the main text but depicts local overreporting as opposed to local overexposure. (Top) Scatterplot of county-level overreporting by the average of county-level Republican vote share in the 2012 and 2016 presidential elections. (Bottom) Scatterplot of county-level overreporting by the county-level proportion of Black residents, based on Census counts. We use Census data from 2010, the most recent available figures that do not rely on intercensal estimates. Election data is from (6). Loess curves with a smoothing parameter of 0.75 indicate weighted average overexposure rates, separately for majority Republican- and Democrat-voting counties in Bottom. The grey-shaded areas indicate the 95% confidence intervals of the loess curves.
Fig. S10. Overreporting by the share of Black sworn officers in an agency. Scatterplot of local overreporting by the share of Black sworn officers in an agency. Data were gathered from the five most recent Law Enforcement Management and Administrative Statistics (LEMAS) surveys (7), conducted in 2000, 2003, 2007, 2013, and 2017. For agencies that reported to LEMAS in both 2013 and 2017, we compute the average share of Black sworn officers from those two years. For agencies that reported in only one of those years, we used data from the year they reported. Among the remaining agencies that did not report in 2013 or 2017 but reported to at least one survey from 2000 to 2007, we compute the average across all surveys to which they reported. Loess curve with a smoothing parameter of 0.75 indicates weighted average overreporting rates. The grey-shaded area indicates the 95% confidence interval of the loess curve.
Fig. S11. Overreporting by Republican vote share and agency type. Scatterplot of overreporting by the average of county-level Republican vote share in the 2012 and 2016 presidential elections. Data are plotted at the county-agency-type level, i.e., one observation for each police department in each county and one for each sheriff’s office in each county. Loess curves with a smoothing parameter of 0.75 indicate weighted average overreporting scores, separately for police departments and sheriffs’ offices. The grey-shaded areas indicate the 95% confidence intervals of the loess curves.
Fig. S12. Cumulative sum of race-crime posts and posting agencies, 2010-2019. The cumulative sum over time of (Left) all race-crime posts and (Right) agencies publishing at least one race-crime post.
Fig. S13. Overexposure regression coefficients by offense type (agencies with populations > 10k). Estimates for average overexposure including only agencies in cities with 10,000 residents or more.
Fig. S14. Overexposure regression coefficients by offense type (always-reporting agencies). Estimates for average overexposure including only agencies that reported in every month from 2010 to 2019.
Table S1. Distance-weighted regression analysis of overexposure.

|                | (1)               | (2)               |
|----------------|-------------------|-------------------|
| Post           | 0.248*** (0.003)  | 0.180*** (0.003)  |
| Crime Controls |                   | ✓                 |
| Distance Weights| ✓                 | ✓                 |
| Agency FEs     | ✓                 | ✓                 |
| Observations   | 216,631,316       | 216,631,316       |
| R²             | 0.202             | 0.217             |

*p<0.05; **p<0.01; ***p<0.001

Model (1) is equivalent to our baseline model reported in Fig. S3, Left. Model (2) adds a dummy variable for each type of offense (e.g., murder, robbery). Standard errors are clustered at the post-level as well as at the level of the agency publishing the post. If overexposure were driven entirely by agencies reporting more frequently on offense types for which Black people represent a larger proportion of arrestees relative to other offenses, adding dummies for offense types would dramatically shrink the coefficient on the post variable. Model (2) shows that the coefficient shrinks by only 27%.
Table S2. The best-performing classifier for each Part I offense in the UCR.

| Label               | Classifier          | Accuracy | AUC    | N     |
|---------------------|---------------------|----------|--------|-------|
| Crime Report        | BERT                | 0.96     | 0.99   | 990   |
| Arrest              | BERT                | 0.94     | 0.97   | 764   |
| Murder              | BERT                | 0.96     | 0.96   | 764   |
| Rape                | Keywords            | 1        | 1      | 715   |
| Robbery             | BERT                | 0.96     | 0.98   | 712   |
| Aggravated Assault  | BERT                | 0.94     | 0.97   | 477   |
| Burglary            | BERT                | 0.87     | 0.9    | 397   |
| Theft               | Multinomial Naive Bayes | 0.8     | 0.81   | 346   |
| Auto Theft          | GradientBoosting    | 0.94     | 0.74   | 209   |

Classifiers were trained as binary classifiers, where a positive label indicates that the post contains a description of the relevant crime and a negative label indicates that the post contains no description of the relevant crime. Posts containing descriptions of crimes higher in the UCR offense hierarchy were discarded at each training step. The resulting number of labeled observations is listed in column N. The algorithms in each comparison included a Bernoulli Naive Bayes classifier, a Multinomial Naive Bayes classifier, an ADA Boosting classifier, and a Gradient Boosting classifier, using a bag-of-words representation of the text as inputs. In addition, a BERT model was fine-tuned to classify each label. Performance was measured as average accuracy across all folds. Accuracy is the fraction of the sum of true positive and true negative labels divided by the sum of false positive and false negative labels. Due to class imbalance, however, accuracy can be difficult to interpret. We thus additionally measured performance using the area under the receiver operating characteristic curve (AUC). The AUC indicates how likely a classifier is able to identify the positive label, given one example of a positive label and one example of a negative label. It thus accounts for class imbalance. The AUC is bounded between 0 and 1.
Table S3. Number of agencies and posts.

| Panel A: Number of UCR Agencies | Panel B: Number of Posts |
|----------------------------------|--------------------------|
| **Condition** | **# UCR Agencies** | **Condition** | **# Posts** |
| Any Agency with Arrest Data | 15,851 | Any Facebook Post | 11,058,289 |
| + Agency Matched to Page | 11,196 | +Page Matched to Agency with Arrest Data | 8,215,173 |
| + Race Post | 7,130 | +Race Post | 108,782 |
| + Race-Crime Post | 6,673 | +Race-Crime Post | 86,699 |
| + Race-Part I Crime Post | 5,925 | +Race-Part I Crime Post | 70,310 |

(A) Number of agencies. Our full sample consists of 15,851 UCR agencies with arrest data. That number falls to 5,925 in our overreporting analysis (but not our distance-weighted overexposure analysis) when we restrict the sample to agencies matched to a Facebook page, agencies with a race post, agencies with a race-crime post, and agencies with a race-crime post for a Part I offense. (B) Number of Facebook posts. Our initial sample consisted of 11,058,289 posts from all agency Facebook pages. That number decreased to 70,310 after we restricted the sample to posts from pages matched to an agency, to race posts, to race-crime posts, and to race-crime posts for a Part I offense.
Table S4. Proportion of rows in NIBRS offender dataset that are missing race information.

| Crime            | n     | Missing Race |
|------------------|-------|--------------|
| Murder           | 63,489| 17.0%        |
| Rape             | 382,514| 15.2       |
| Robbery          | 1,225,175| 16.6    |
| Aggravated Assault| 2,313,906| 11.2     |
| Burglary         | 5,702,789| 65.5      |
| Theft            | 19,669,205| 54.8     |
| Auto Theft       | 2,108,134| 66.4      |

The NIBRS offender-level dataset contains a row for every suspect reported to NIBRS. Criminal incidents in which no suspects were reported receive one row but all columns are set to missing. This table reports, by most serious Part I offense, the proportion of rows in the offender dataset that are missing race information either because a suspect was reported with no race information or because no suspect was reported for an incident.
References

1. S. Kemp, *Digital around the world in April 2020*. We Are Social (2020). Available at https://wearesocial.com/us/blog/2020/04/digital-around-the-world-in-april-2020/ (accessed 4 August 2022).

2. K. A. Quesenberry, M. K. Coolsen, What makes Facebook brand posts engaging? A content analysis of Facebook brand post text that increases shares, likes, and comments to influence organic viral reach. *J. Curr. Issues & Res. Advertising* 40, 229-244 (2019).

3. H. A. M. Voorveld et al., How advertising in offline media drives reach of and engagement with brands on Facebook. *Int. J. Advertising* 37, 785-805 (2018).

4. J. Kaplan, Jacob Kaplan's Concatenated Files: Uniform Crime Reporting (UCR) Program Data: Arrests by Age, Sex, and Race, 1974-2019. Inter-university Consortium for Social and Political Research. Available at https://doi.org/10.3886/E102263V14. Deposited 27 September 2021.

5. J. Kaplan, Jacob Kaplan's Concatenated Files: National Incident-Based Reporting System (NIBRS) Data, 1991-2020. Inter-university Consortium for Social and Political Research. Available at https://www.openicpsr.org/openicpsr/project/118281/version/V5/view. Deposited 9 March 2022.

6. MIT Election Data & Science Lab, “County Presidential Election Returns 2000-2020.” Harvard Dataverse. Available at https://doi.org/10.7910/DVN/VOQCHQ. Deposited 10 June 2021.

7. Law Enforcement Management and Administrative Statistics (LEMAS) Series. Inter-university Consortium for Social and Political Research. Available at https://www.icpsr.umich.edu/web/NACJD/series/92. Deposited 20 August 2020.