A Appendix

Table of Contents

Figures A19, A20, and A21 show the adverse impact observed under hypothetical threshold policies, analogously to Figure 5.

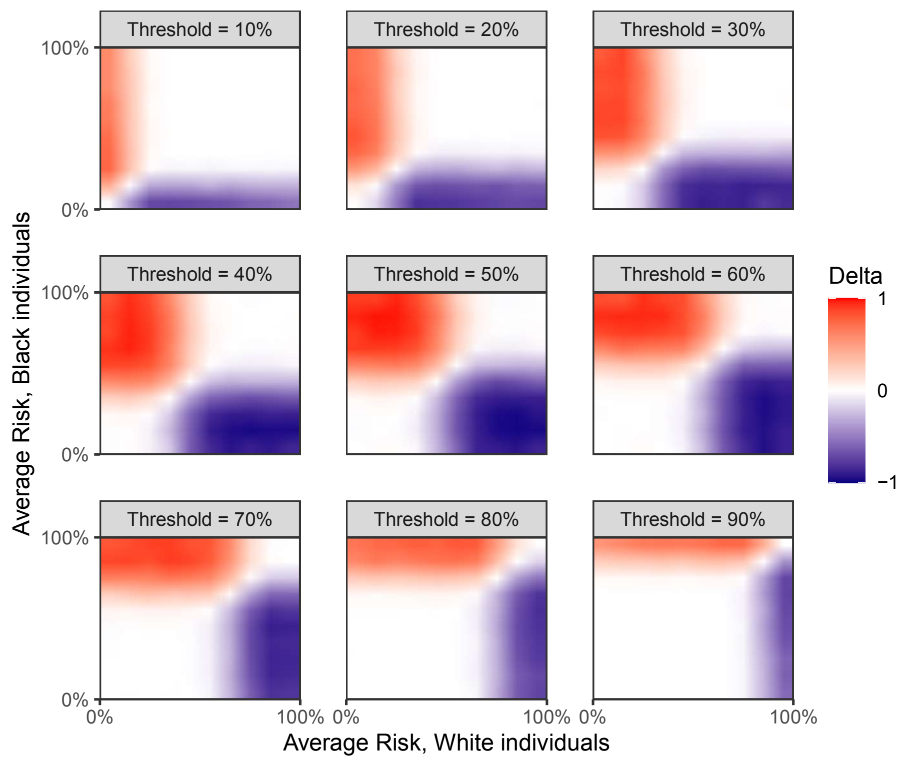


Figure A1: *Extension of the scenario in Figure 1 to continuous risk. In this example, risk follows a beta distribution with a mean between 0 and 1 and a fixed variance of 0.01. Each panel shows values of Δ (Delta) calculated under a fixed search threshold for synthetic groups of Black and white stopped individuals, with risk values randomly generated from a beta distribution with the mean specified on the corresponding axis. As in Figure 1, each panel has a white diagonal line, indicating that the Δ measure is correctly 0 when risk distributions are identical. When the mean of both risk distributions is far from the search threshold, Δ is also close to 0. However, when the mean of either risk distribution is close to the search threshold, the calculated value of Δ is more sensitive to small changes in the shape of the risk distribution(s) whose mean is close to the search threshold.*

REASON FOR STOP: *(Select the primary reason for stop)*

☐ Traffic Violation: ☐ **Moving** ☐ **Equipment** ☐ **Non-moving**

Code section related to violation: ______________________

☐ **Reasonable suspicion that the person was engaged in criminal activity**

Select all that apply to describe the basis of suspicion:

- ☐ Officer witnessed commission of a crime
- ☐ Matched suspect description
- ☐ Witness or victim identification of suspect at the scene
- ☐ Carrying suspicious object
- ☐ Actions indicative of casing a victim or location
- ☐ Suspected of acting as a lookout
- ☐ Actions indicative of a drug transaction
- ☐ Actions indicative of engaging in a violent crime
- ☐ Other reasonable suspicion of a crime

If known, Code for suspected violation: ______________________

☐ **Known to be on parole/probation/PRCS/mandatory supervision**

☐ **Knowledge of outstanding arrest warrant/wanted person**

☐ **Investigation to determine whether the person is truant**

☐ **Consensual encounter resulting in a search**

☐ **** Possible conduct warranting discipline under Education Code (EC) 48900, et al***

Code Section: ☐ 48900 ☐ 48900.2 ☐ 48900.3 ☐ 48900.4 ☐ 48900.7

When EC 48900 is selected, specify the subdivision: ______________________

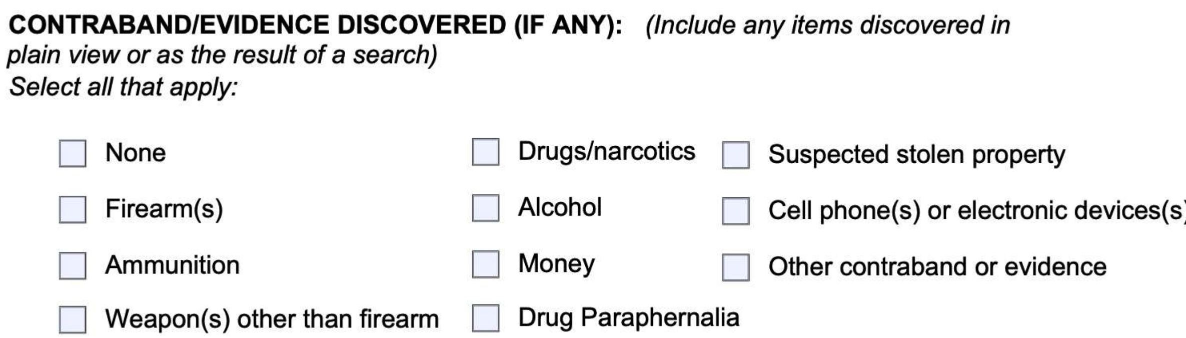
☐ **** Determine whether the student violated school policy***

Figure A2: *Stop reasons listed on the RIPA data collection form. Text position is altered slightly for readability.*

BASIS FOR SEARCH: *(Only applicable when the Actions Taken include "Search of person was conducted" and/or "Search of property was conducted. Select all that apply)*

- [ ] Consent given
- [ ] Officer safety/safety of others
- [ ] Search warrant
- [ ] Condition of parole/probation/PRCS/mandatory supervision
- [ ] Suspected weapons
- [ ] Visible contraband
- [ ] Odor of contraband
- [ ] Canine detection
- [ ] Evidence of crime
- [ ] Incident to arrest
- [ ] Exigent circumstances/emergency
- [ ] Vehicle inventory **(for search of property only)**
- [ ] **Suspected violation of school policy*

Figure A3: *Search bases listed on the RIPA data collection form. Text position is altered slightly for readability.*



CONTRABAND/EVIDENCE DISCOVERED (IF ANY): *(Include any items discovered in plain view or as the result of a search)*
Select all that apply:

- ☐ None
- ☐ Firearm(s)
- ☐ Ammunition
- ☐ Weapon(s) other than firearm
- ☐ Drugs/narcotics
- ☐ Alcohol
- ☐ Money
- ☐ Drug Paraphernalia
- ☐ Suspected stolen property
- ☐ Cell phone(s) or electronic devices(s)
- ☐ Other contraband or evidence

Figure A4: *Types of contraband listed on the RIPA data collection form. Text position is altered slightly for readability.*

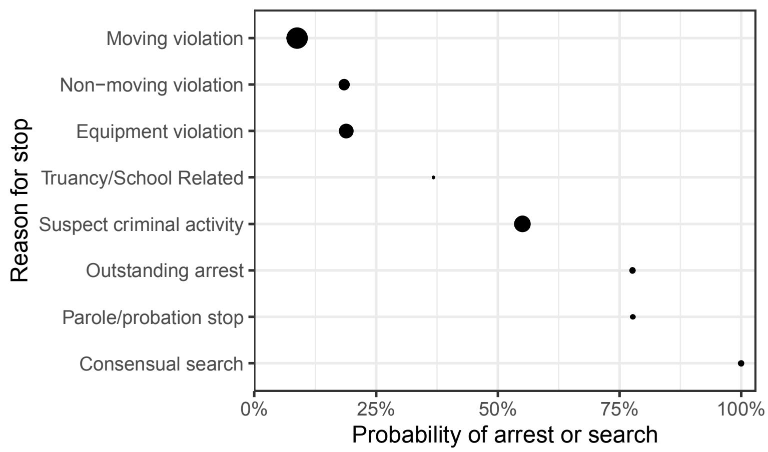


Figure A5: *For all stops recorded by the 50 law enforcement agencies, the proportion of stops resulting in a search or arrest, disaggregated by stop reason. The size of the points corresponds to the number of stops prompted by each reason. Stops initiated due to an outstanding arrest, parole, or probation are excluded from the main analysis, as are consensual searches and stops initiated due to school policy.*

Agency	Stops	B (pop)	H (pop)	B (stop)	H (stop)
Los Angeles PD	320,702	11.8	3.8	2.7	2.1
Los Angeles Shf	273,000	6.5	2.7	2.1	1.8
San Diego PD	92,898	6.8	1.8	1.8	1.4
San Bern. Shf	81,548	1.7	0.6	1.1	0.9
Riverside Shf	68,946	2.8	1.5	1.2	1.2
San Jose PD	39,776	7.4	4.3	1.9	1.8
Orange Shf	38,916	4.2	1.8	2.0	2.2
Sacramento PD	38,113	7.4	1.7	2.1	1.7
Sacramento Shf	24,902	4.2	1.0	1.3	1.2
Anaheim PD	21,349	2.7	1.1	1.4	1.3
Irvine PD	20,335	4.3	3.3	1.3	1.1
Ventura PD	20,121	4.1	1.5	1.1	1.2
Burbank PD	19,930	7.7	3.3	1.8	1.7
San Diego Shf	18,482	1.9	1.4	1.7	1.9
Fontana PD	18,255	1.5	0.9	0.9	0.9
Santa Rosa PD	18,189	6.7	2.1	1.3	1.3
Escondido PD	16,111	4.4	1.1	1.3	1.1
Riverside PD	16,090	4.7	1.1	2.0	1.2
San Mateo PD	15,136	7.6	3.3	2.2	1.7
Placer Shf	14,853	8.8	1.6	1.4	1.2
Redding PD	14,789	2.6	0.5	0.9	0.9
Oxnard PD	14,360	3.1	1.4	1.5	1.5
San Mateo Shf	14,315	5.1	2.0	2.5	1.6
San Fran. PD	14,273	8.8	2.2	1.7	1.3
Oceanside PD	14,255	2.7	1.7	1.1	1.5
Oakland PD	13,941	6.9	3.0	1.2	1.0
Rialto PD	13,421	1.5	0.6	1.3	0.9
Costa Mesa PD	13,103	5.6	1.0	1.6	1.1
Coronado PD	12,825	9.3	6.7	1.7	2.3
Chino PD	12,575	1.0	1.2	0.7	0.9
Fresno Shf	12,140	2.5	1.3	0.9	1.0
Pasadena PD	11,968	9.2	2.5	3.6	2.2
Alameda Shf	11,595	4.7	2.1	1.7	1.2
Clovis PD	11,336	3.8	1.1	1.1	1.1
Santa Ana PD	11,078	3.3	1.0	1.3	0.9
Hanford PD	10,680	3.1	1.8	1.4	1.1
Orange PD	10,244	5.2	2.1	1.6	1.5
BART PD	10,032	11.2	1.3	1.1	1.1
Folsom PD	10,011	3.8	1.2	2.3	1.5
Livermore PD	9,730	11.9	2.5	1.4	1.2
Htg. Bch. PD	9,634	4.6	1.4	1.1	1.1
Vacaville PD	9,428	6.6	1.5	1.8	1.3
Visalia PD	9,424	5.0	1.9	2.0	1.3
Roseville PD	9,274	6.8	1.1	1.2	1.1
San Joaquin Shf	9,218	1.6	0.9	1.0	1.1
Petaluma PD	9,213	7.5	2.4	1.7	1.3
Fairfield PD	9,105	3.1	1.0	1.3	1.0
Glendale PD	8,985	9.9	3.5	2.2	2.1
Carlsbad PD	8,946	6.9	2.2	1.2	1.3
Pacifica PD	8,766	4.6	1.9	1.4	1.2

Table A1: *For the 50 agencies with the most stops recorded in 2022, adverse impact of search decisions for Black (B) and Hispanic (H) individuals, relative to white individuals. Adverse impact is defined at the population-level ("pop") by the ratio of race-specific, per capita search rates. Per capita search rates are computed as the number searched over the population of the jurisdiction served by the agency. For example, a value of 2 for "B (pop)" implies that the per capita search rate for Black individuals was twice as high as that of white individuals. Adverse impact is defined at the stop-level ("stop") by the ratio of stop-level search rates. Stop-level search rates are computed as the number searched over the number stopped. CHP: California Highway Patrol; PD: Police Department; Shf: County Sheriff; Bern: Bernardino; Htg Bch: Huntington Beach; BART: Bay Area Rapid Transit.*

Variable	Description	Possible values
Race/ethnicity	Race or ethnicity of the driver, as perceived by the officer. Approximately 99% of individuals in the data are classified into a single race or ethnicity. The remaining 1% are considered 'Hispanic' if they are classified as Hispanic along with any other race or ethnicity, 'Black' if they are classified as Black and any other race or ethnicity other than Hispanic, 'White' if they are classified as 'White' and no other race or ethnicity, and 'Other' otherwise.	White | Black | Hispanic | Other
Gender	Gender of the driver, as perceived by the officer. Approximately 99.7% of individuals in the data are perceived as either male or female. Due to small sample size, the remaining 0.3% of individuals are excluded from the analysis.	Male | Female
Reason for stop	Reason for the stop, as recorded by the officer. Only a single reason can be recorded.	Suspected criminal activity | Equipment violation | Moving violation | Non-moving violation
RAS factors	Reasonable articulable suspicion (RAS) factors recorded by the officer if criminal activity is suspected. Multiple factors may be recorded.	Officer witnessed a crime | Suspect matched description | Witness or victim identification of suspect | Carrying suspicious object | Casing a victim or location | Acting as a lookout | Actions indicative of drug transaction | Actions indicative of violent crime | Other
Search basis	If a search is carried out by the officer, the recorded basis for the search. Multiple bases may be recorded.	Contraband in plain view | Odor of contraband | Officer safety or safety of others | Suspected weapon | Evidence of crime | Emergency | Canine detection of contraband | School policy
Traffic offense by ped.	Whether the traffic offense that led to the stop was likely carried out by a pedestrian (e.g., jaywalking).	Yes | No
RAS offense by driver	Whether the reasonable suspicion offense that led to the stop was likely carried out by a driver. For example, driving under the influence.	Yes | No
Is call for service	Whether the stop was in response to a call for service.	Yes | No
Multi-person encounter	Whether more than one individual was recorded as stopped. The main analysis considers only the first person listed.	Yes | No
Motivating offense FEs	Fixed effects for the motivating offense that led to the stop.	Robbery | Assault | Burglary | Theft | DUI | Speeding | Etc.
City FEs	Fixed effects for the city in which the stop took place.	Los Angeles | San Diego | San Jose | San Francisco | Etc.
Month FEs	Fixed effects for the month of the stop.	0-11
Weekday FEs	Fixed effects for the day of the week on which the stop took place.	0-6
Hour FEs	Fixed effects for the hour of the day in which the stop took place.	0-23

Table A2: *Covariates included in each agency's random forest risk model used to estimate the likelihood of recovering contraband from a discretionary search.*

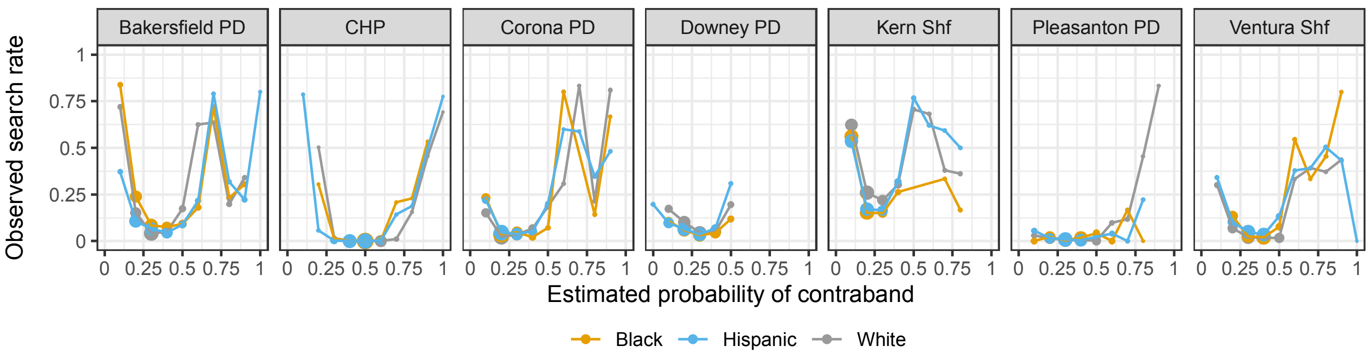


Figure A6: *Large agencies with invalid risk models. These agencies are larger or as large as the 50 agencies included in the expanded analysis. If the estimated risk of recovering contraband is indeed the main motivation for carrying out a discretionary search, then discretionary search rates should increase monotonically as a function of estimated risk. If not, the risk measure may not be an appropriate measure of qualification for a discretionary search. There may be, for example, an agency-specific requirement for conducting a search that is not recorded in the RIPA data. For these agencies, search rates do not increase monotonically with estimated risk, so these agencies are excluded from the main analysis.*

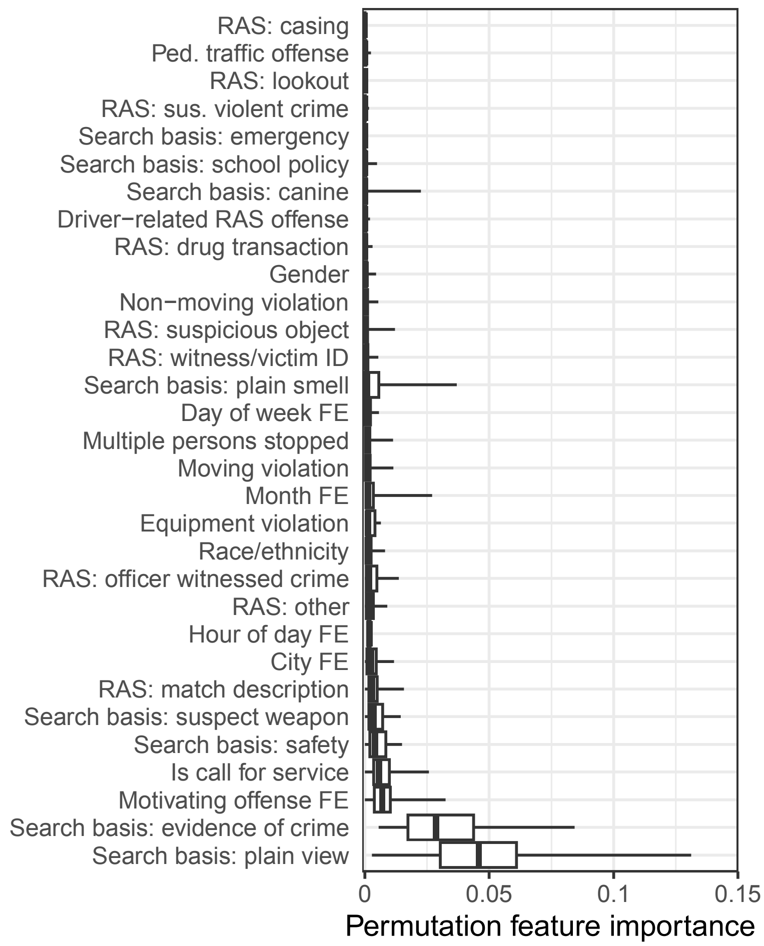


Figure A7: *Boxplots of the permutation feature importance values for each covariate included in the agency-specific random forest risk models. The boxplots show the distribution of feature importance values across all 50 agencies. In most jurisdictions, plain view contraband and evidence of a crime are by far the most influential features for predicting contraband recovery from a discretionary search of a stopped individual.*

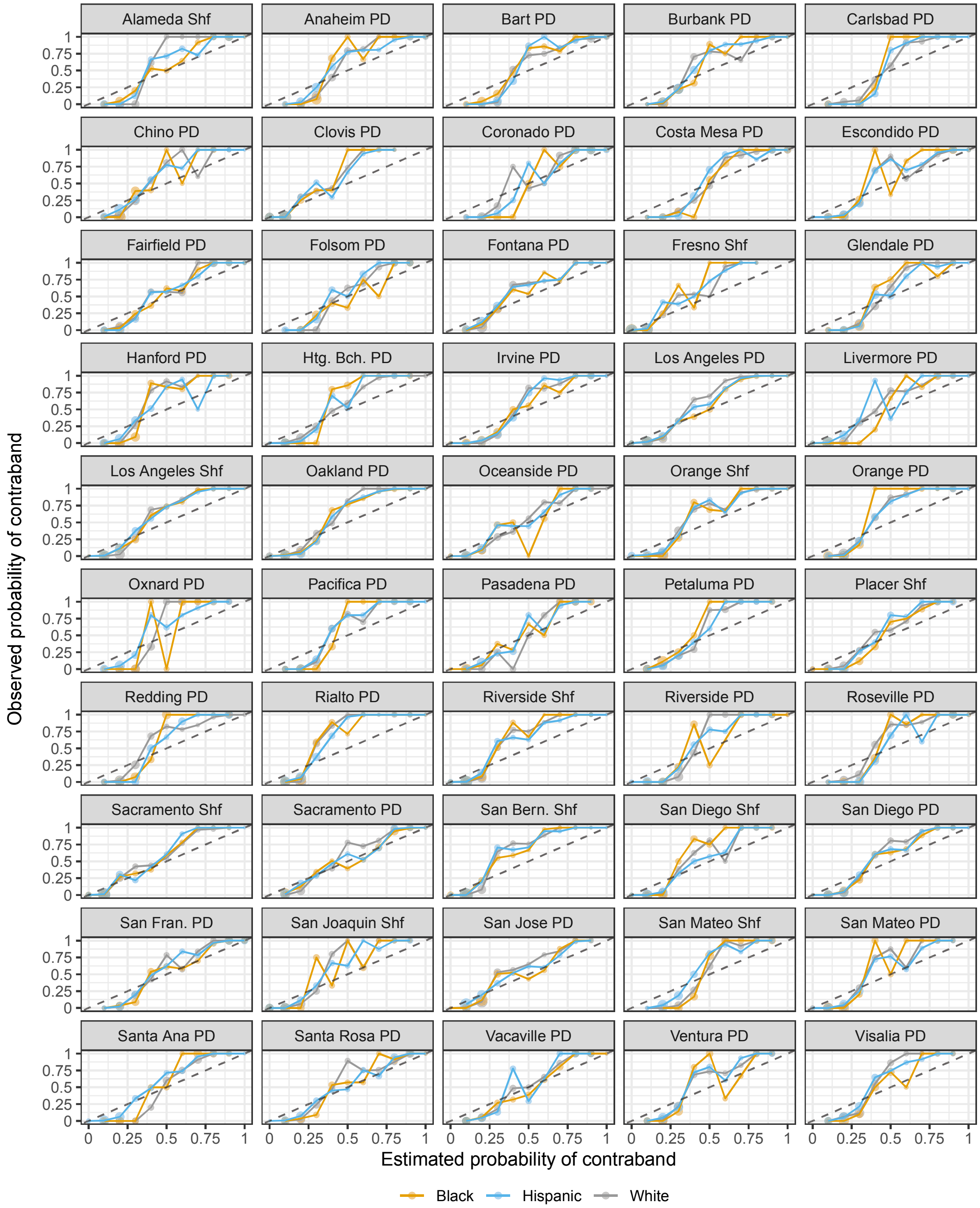


Figure A8: *Calibration plots for the contraband-carrying risk models fit to each of the 50 largest agencies. Larger point sizes correspond to more observations. For a well-fitted risk model, the observed probability of carrying contraband should be monotonic as a function of the estimated probability of carrying contraband. Monotonicity ensures that thresholded search policies will return the same results regardless of whether the model is recalibrated to fall along the parity line. The monotonicity requirement is approximately met by most agencies and race/ethnicity groups.*

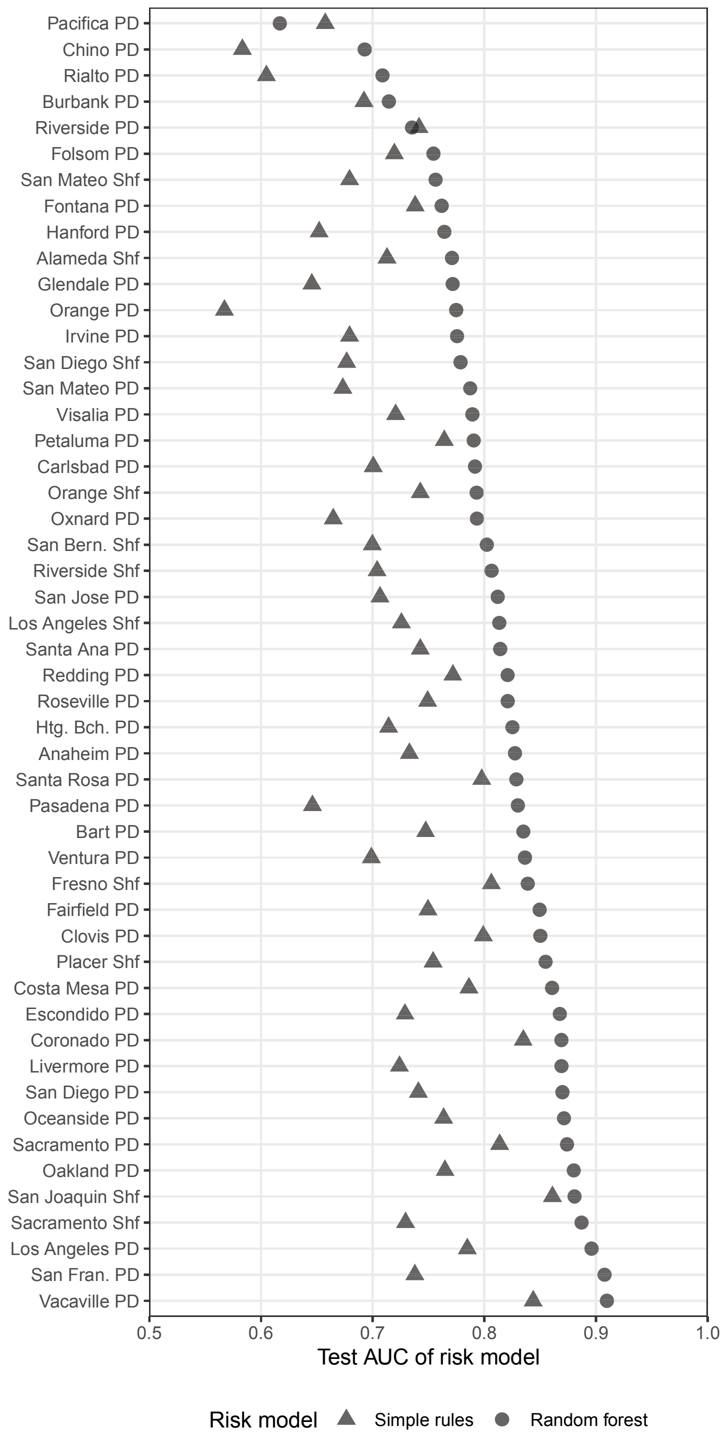


Figure A9: *Estimated out-of-sample AUC for the random forest and simple rules risk models for the 50 largest agencies. AUC is calculated with an 80/20 train/test split. Performance of the random forest risk models is moderate-to-strong across agencies, with almost all agencies having an AUC above 0.7, and the majority having an AUC above 0.8. For most agencies, the simple rules model performs worse than the random forest model, though the performance of the simple rules models exceed 0.65 in almost all jurisdictions, with the majority exceeding 0.7.*

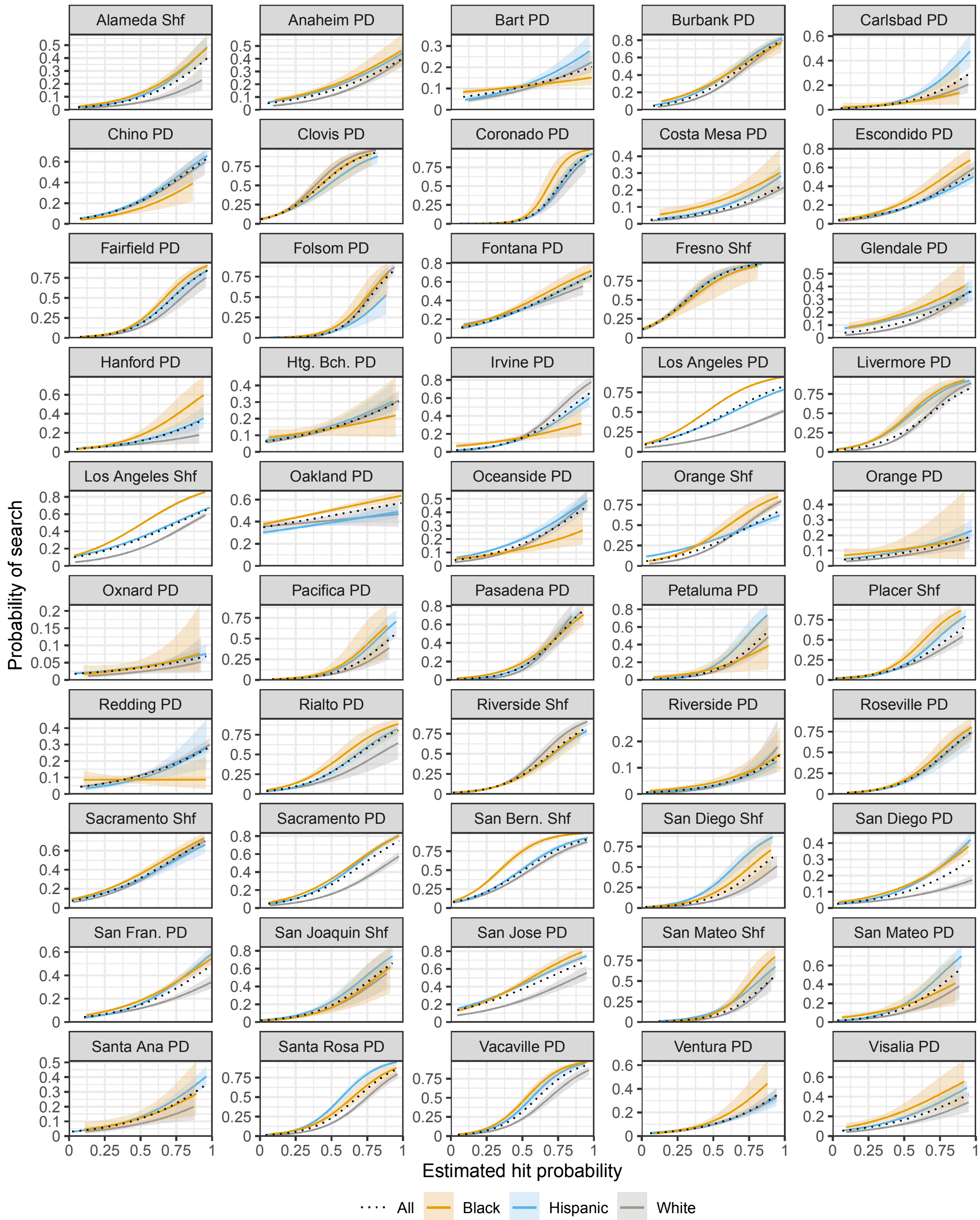


Figure A10: *For each of the 50 largest agencies, observed probability of search as a function of estimated probability of recovering contraband. Lines are fit via logistic regression, with 95% confidence bands. For agencies where the error bands do not overlap, such as the Los Angeles Police Department, the observed difference in discretionary search rates (i.e., adverse impact) cannot be fully explained by the likelihood of recovering contraband from a discretionary search.*

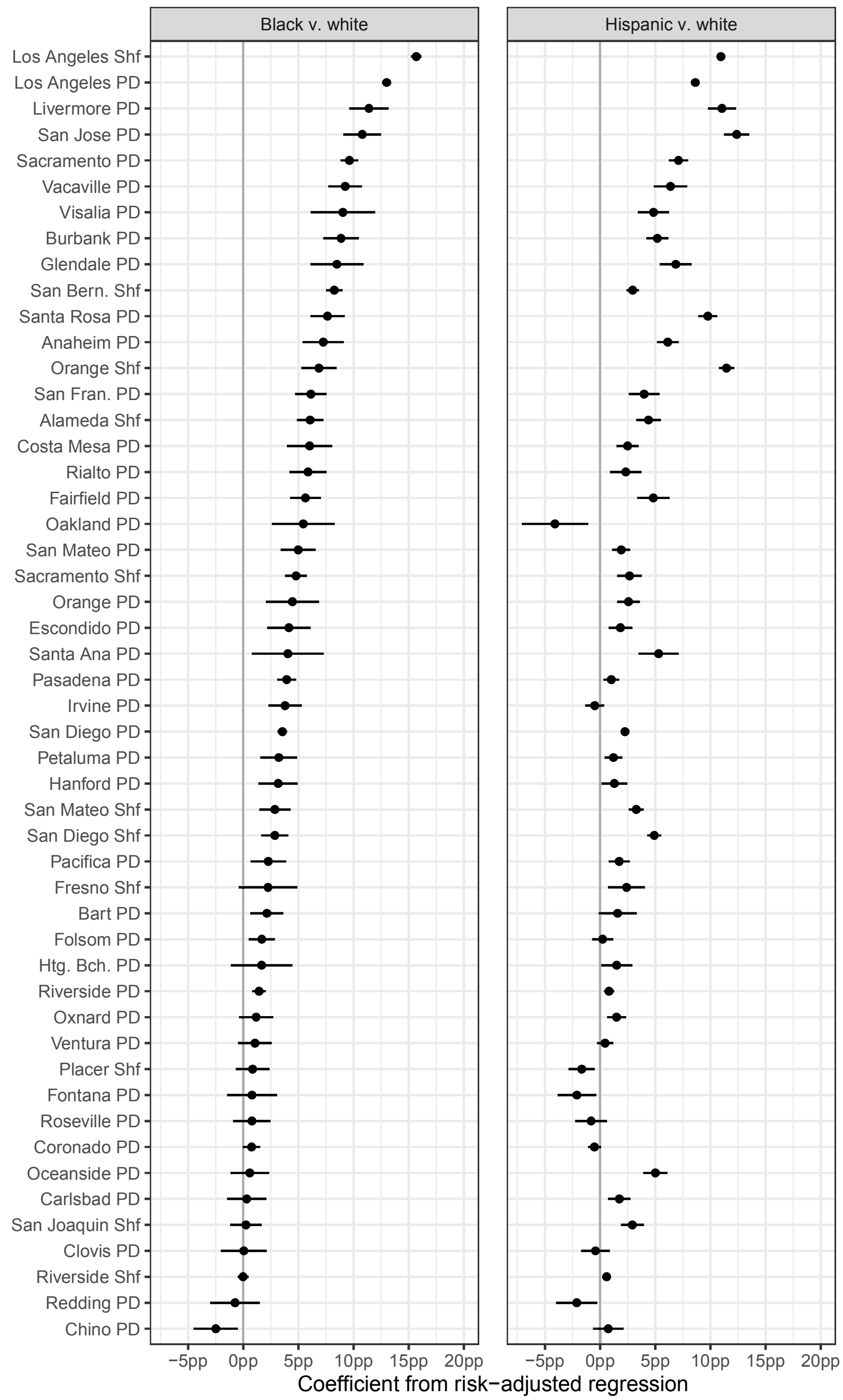


Figure A11: *For each of the 50 largest agencies, race and ethnicity coefficients from a risk-adjusted regression model fit separately to each agency, with 95% confidence intervals. When confidence intervals do not overlap with the 0pp line, the observed differences in search rates across race/ethnicity cannot be fully explained by the estimated risk of recovering contraband.*

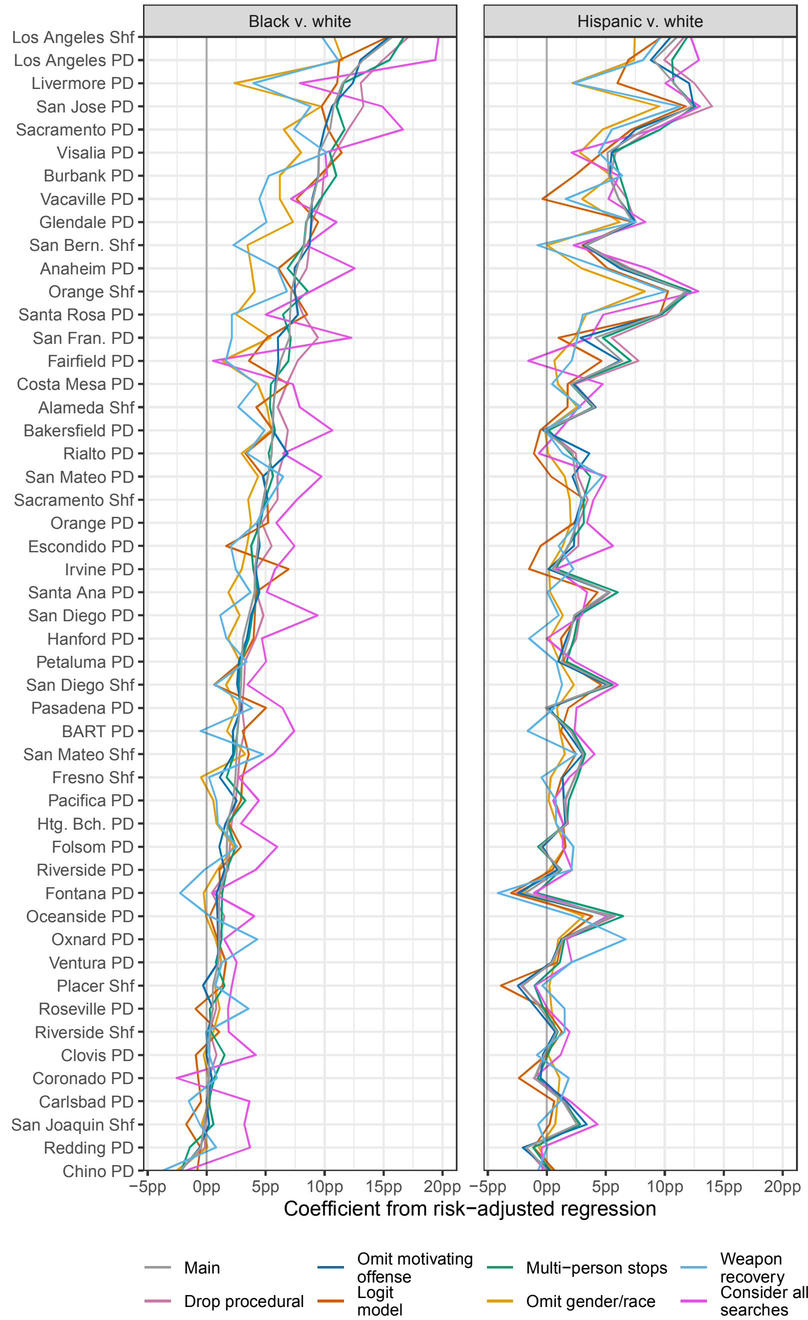


Figure A12: *Robustness checks of the risk-adjusted regression models. "Main" indicates the race and ethnicity coefficients from the unaltered models in Figure A11. "Drop procedural" removes stops resulting in a non-discretionary search. "Omit motivating offense" uses risk models that do not account for the motivating traffic violation or suspected offense that prompted each stop. This model simulates what might happen under moderate omitted variable bias, as this covariate is the third most predictive feature in the random forest models. "Logit model" uses a logistic regression with no interaction terms to fit the risk models, instead of random forests. "Multi-person stops" includes all individuals stopped in multi-person encounters, instead of just the individual recorded first. "Omit gender/race" excludes gender and race from each risk model. "Weapon recovery" uses weapon recovery as the outcome of the risk models, as opposed to using any contraband recovery. "Consider all searches" uses the unaltered search label for non-discretionary searches. Results are qualitatively similar across specifications, though coefficients are somewhat attenuated under the "Omit gender/race" and "Weapon recovery" specifications, and coefficients for Black individuals tend to be larger when non-discretionary search labels are unaltered.*

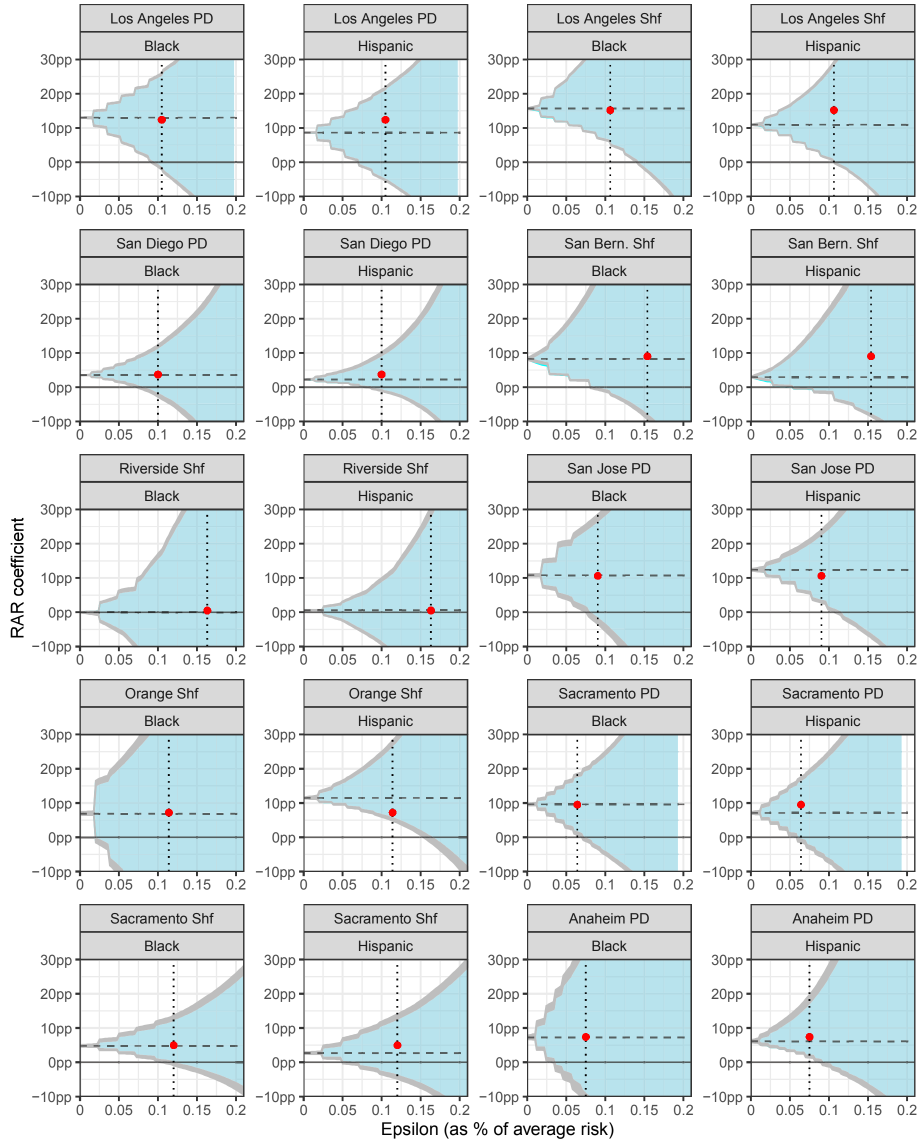


Figure A13: *Sensitivity analysis for the risk-adjusted regression results in Figure 4. The horizontal dashed line shows the risk-adjusted regression coefficient for each race and ethnicity. The x-axis indicates the mean absolute deviation (MAD) between true risk and estimate risk, divided by the overall contraband carrying for the given agency. Jung et al. [2023] refer to this value as ϵ (Epsilon). The blue shaded area indicates the most conservative bounds in the risk-adjusted regression coefficient, conditional on the value of the MAD between true and estimate risk indicated on the x-axis. The gray shaded area shows 95% confidence intervals from a bootstrapping procedure for generating the bounds. The red dot indicates the risk-adjusted regression coefficient from the "Omit motivating offense" specification in Figure A12, which is intended to simulate a degree of moderate omitted variable bias by blinding the model to the traffic violation or suspected offense that prompted the stop. For all agencies, the actual and blinded estimates are quite similar. As a benchmark, the vertical dotted line that intersects the red dot is the observed MAD between risk estimated under the actual model and risk estimated under the blinded model. If the bounds at this benchmark MAD exceed 0pp, then the results are robust to any type of confounding that results in the same MAD. The results for stopped Black individuals are approximately robust to moderate confounding for the Los Angeles PD, the Los Angeles County Sheriff, the Sacramento PD, and the Sacramento County Sheriff. For Hispanic individuals, the results are approximately robust for the Los Angeles County Sheriff, the San Diego PD, the San Jose PD, the Orange County Sheriff, the Sacramento PD, and the Anaheim PD.*

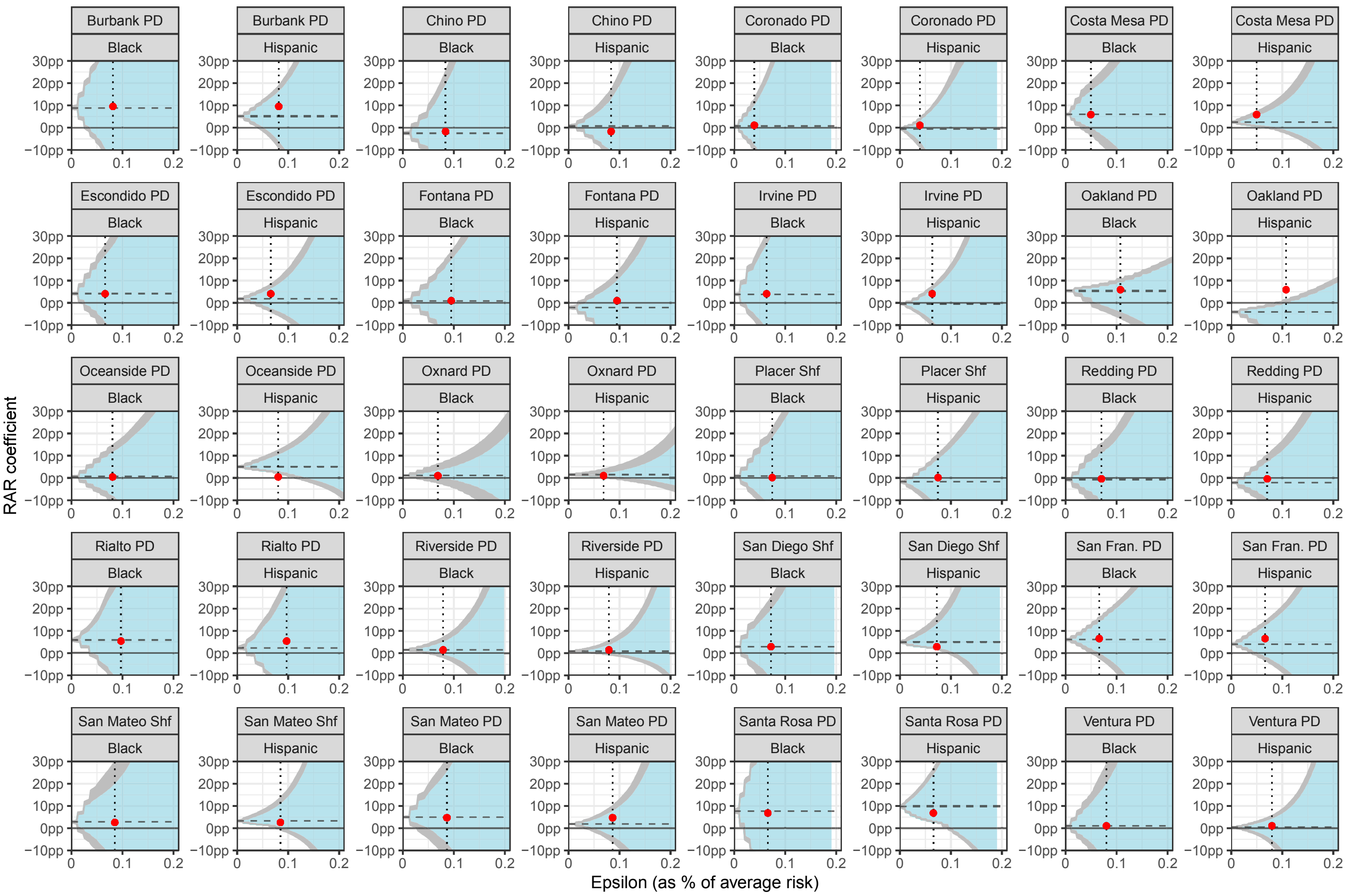


Figure A14: *Sensitivity analysis for the 11th through 30th largest agencies, using the same methods as Figure A13.*

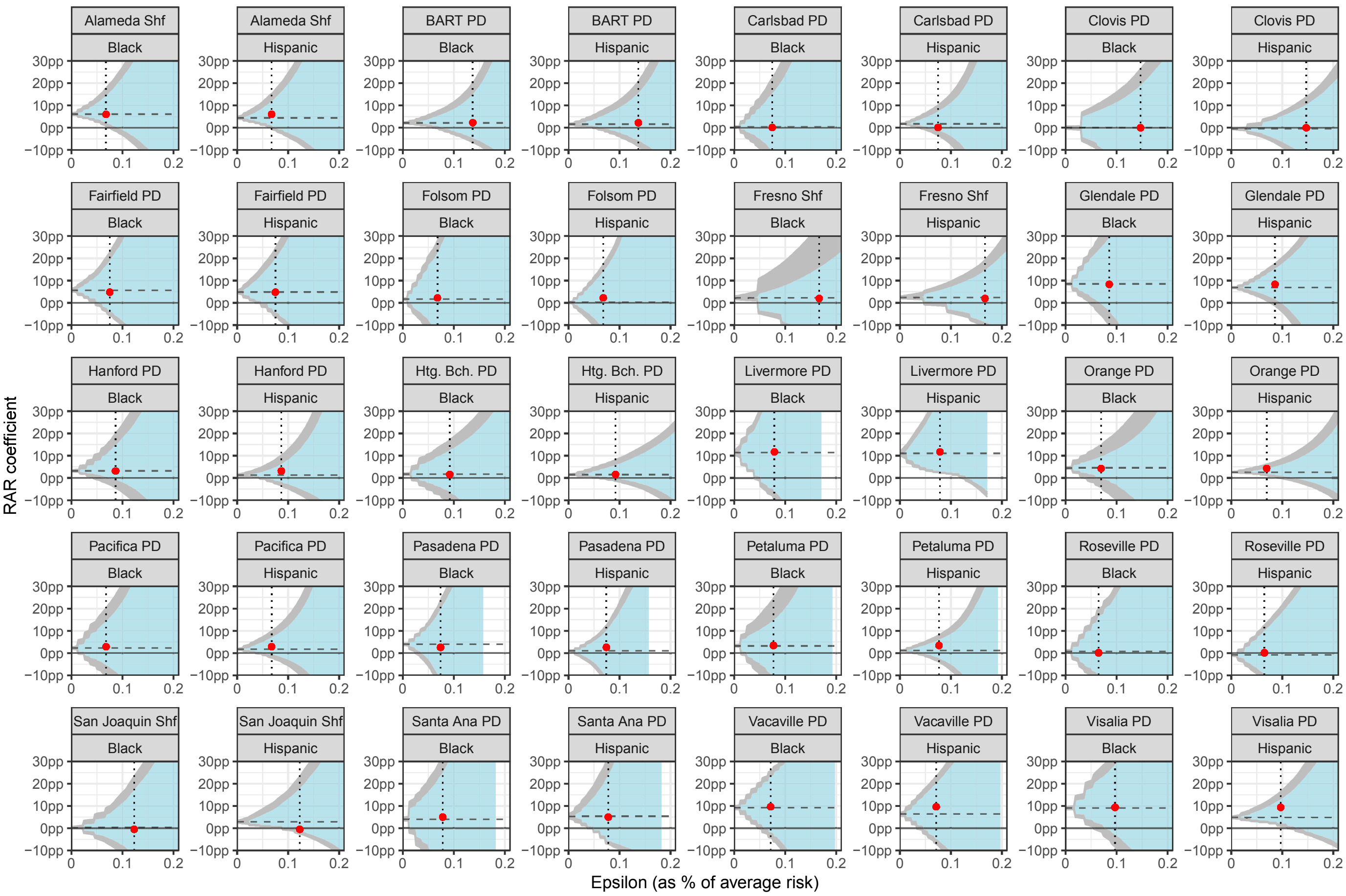


Figure A15: *Sensitivity analysis for the 31st through 50th largest agencies, using the same methods as Figure A13.*

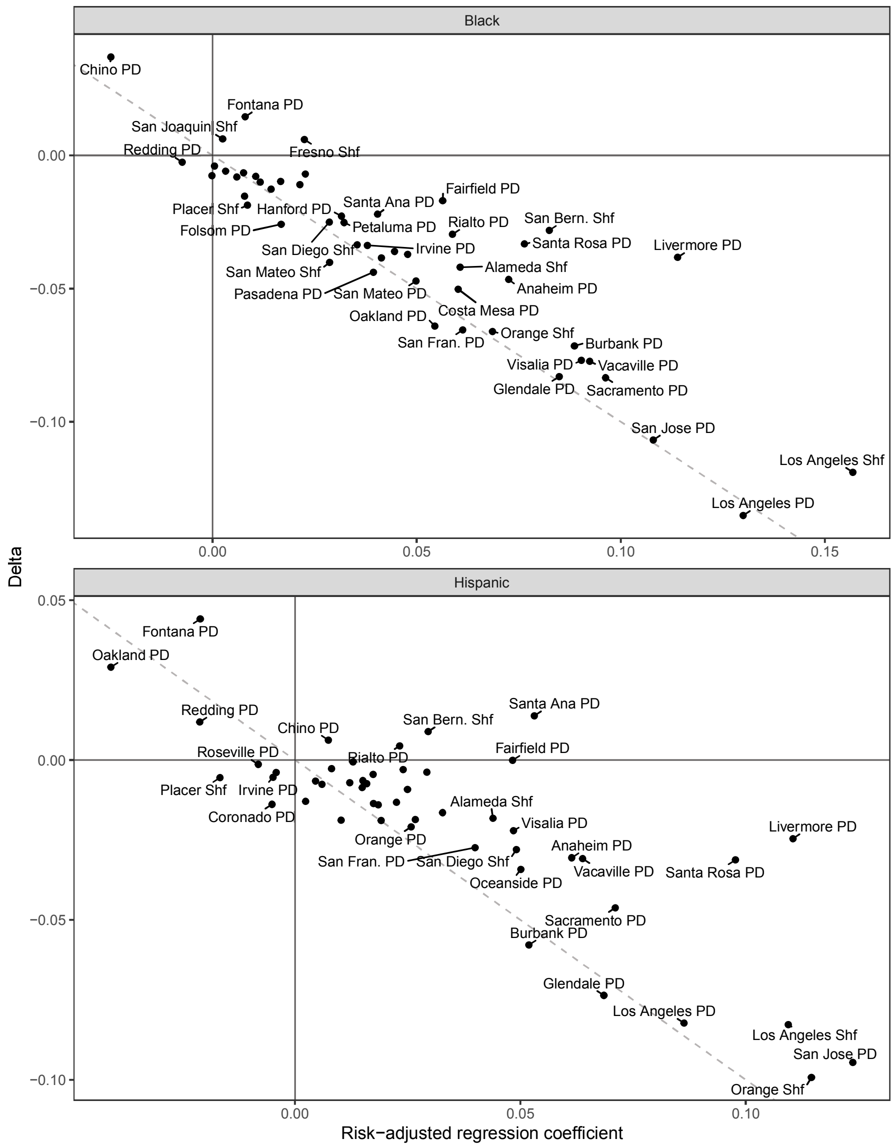


Figure A16: *For each of the 50 largest agencies, comparison of the Δ (Delta) measure of discrimination proposed by Arnold et al. [2021] to the corresponding estimate from risk-adjusted regression. Across jurisdictions, the estimates tend to be similar in magnitude, though there are notable deviations. For example, the Δ measure is close to zero for Hispanic drivers stopped by the Fairfield Police Department, yet risk-adjusted regression suggests a significant difference in risk-adjusted search rates. This discrepancy is possibly a result of the inframarginality concerns raised in the main text. Indeed, the distributions of estimated risk among Hispanic and white drivers stopped by the Fairfield Police Department are quite different (See Figure A17). Figure A18 is a zoomed-in version of this plot with labels for points closer to the origin.*

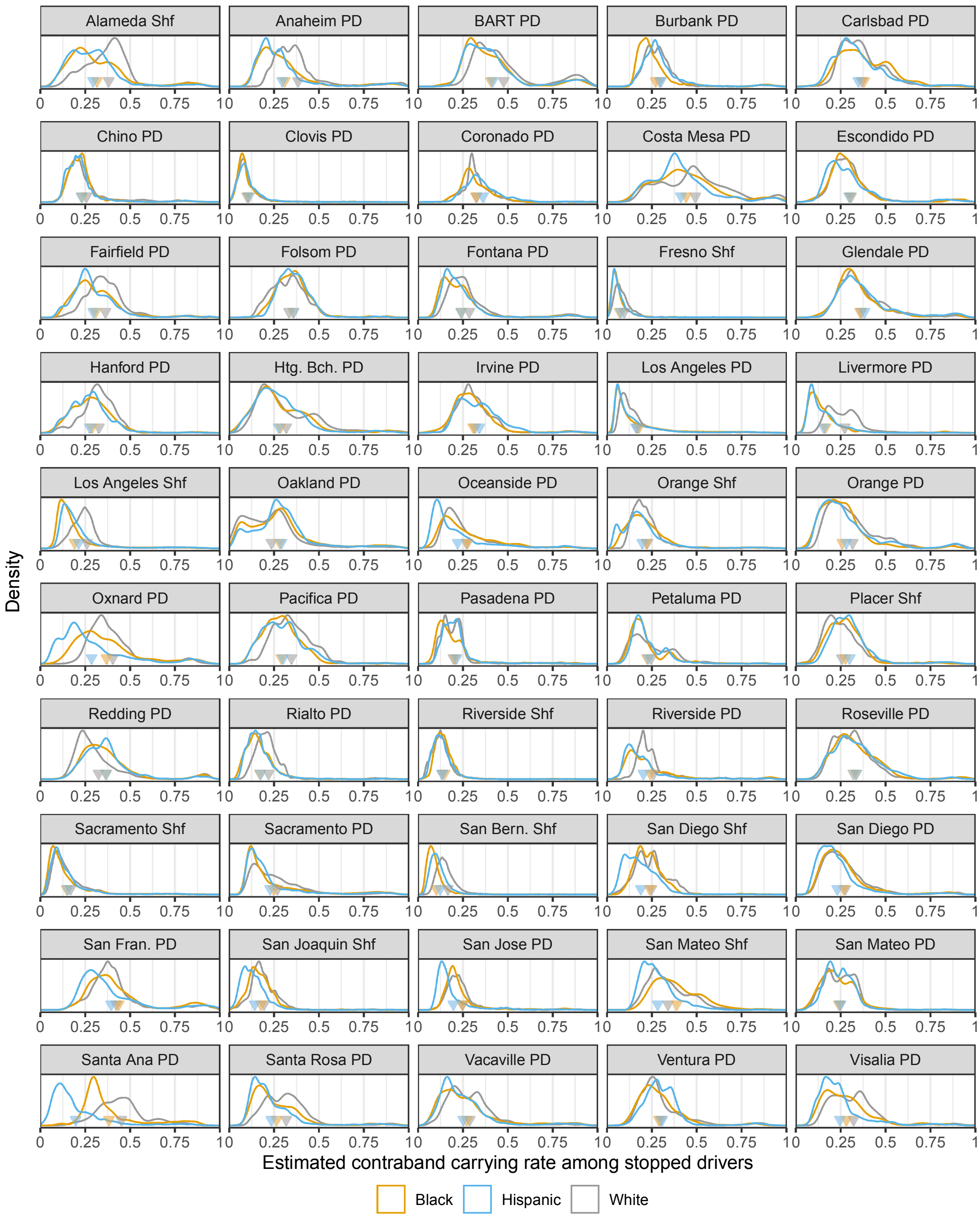


Figure A17: *Distribution of estimated risk by race and ethnicity across the 50 largest agencies. While the distributions are quite similar for certain agencies (e.g., the Sacramento Police Department and County Sheriff), they are markedly different for other agencies (e.g., the Santa Ana Police Department). When underlying distributions of risk differ across groups, inframarginal statistics, such as differences in error rates, may falsely suggest or refute discrimination.*

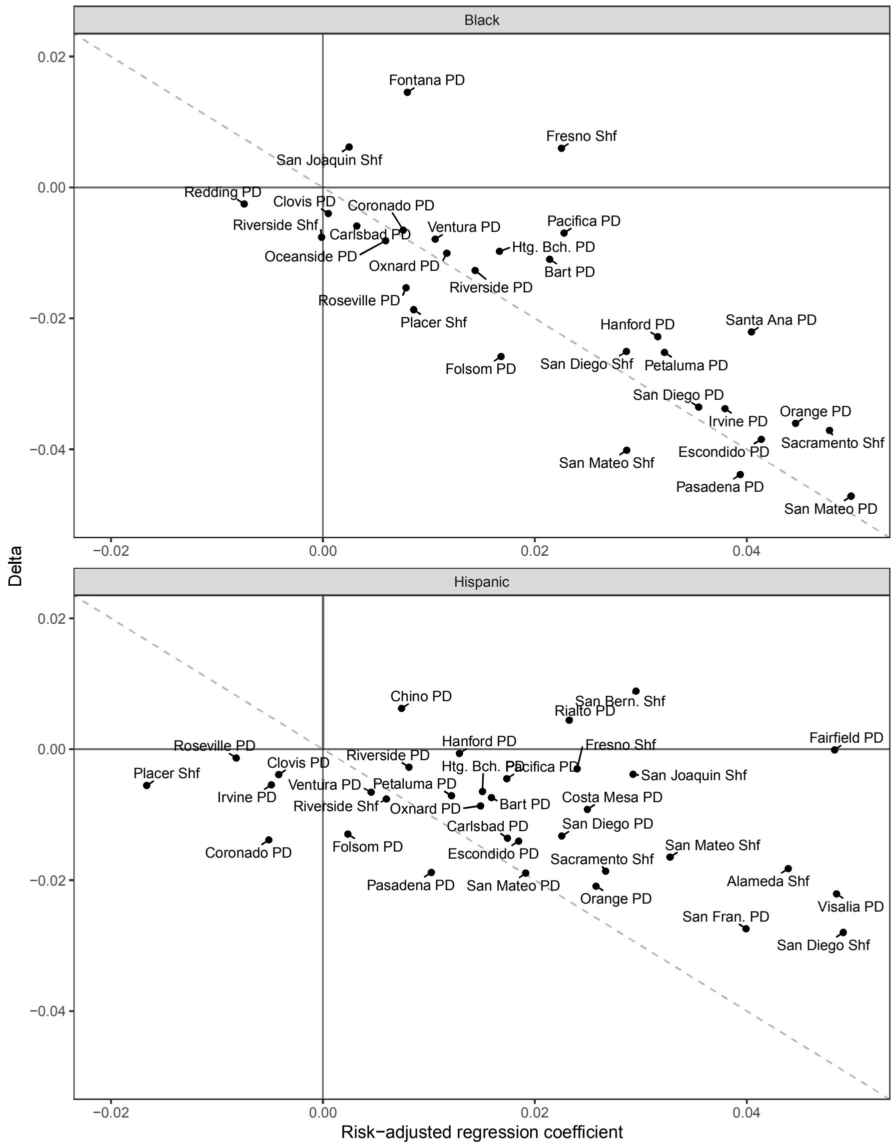


Figure A18: *Zoomed in version of Figure A16 for identifying agencies close to the origin.*

Constructing the simple rule

To construct the simple rule risk model, we begin with the same set of covariates as those listed in Table A2. Figure A7 shows that, across jurisdictions, plain view contraband and evidence of a crime are the most predictive factors of contraband recovery from a search, with predictive feature importance dropping off quickly. Risk varies substantially with agency. As such, we fit agency-specific logistic regression models to estimate the likelihood of recovering contraband using the city, the traffic violation or suspected offense than prompted the stop, and the two most predictive factors: whether the search was prompted by contraband in plain view, and whether the search was prompted by evidence of a crime. We fit this model only on individuals who were searched at the discretion of the officer, as the contraband recovery outcome is unknown for individuals who are not searched.

With this fitted model in hand, we multiply the fitted coefficients of the two key factors by 10 to put the coefficients on an approximate integer scale, and then round the two coefficients to the nearest integer. Using just these two rounded coefficients, we calculate a risk score for each searched individual. For example, if the rounded coefficients are 10 for plain view contraband and 15 for evidence of a crime, a stopped individual with plain view contraband but no evidence of a crime would receive a score of 10. Finally, we fit an additional logistic regression model that predicts contraband recovery using this score, the city, and the motivating offense that motivated the stop. This model is again fit just to searched individuals. This final fit provides the optimal coefficients for each city and motivating offense.

To operationalize the simple rule, each agency could select a risk threshold above which officers are obligated to conduct a discretionary search (e.g., search if there is at least a 20% chance of recovering contraband). Using the final fitted model, each agency would determine, for each combination of city and motivating offense, the minimum score needed for the estimated risk of contraband recovery to exceed the desired threshold (e.g., 20% risk might correspond to a score of 12 if the stop occurs in Santa Clara and the stop was prompted by a speeding driver). Before initiating a stop, officers could report the city and motivating offense to obtain a risk threshold. Once the stop begins, officers could quickly calculate the individual's risk score using the two factors, and then conduct a search if the score exceeds the known threshold.

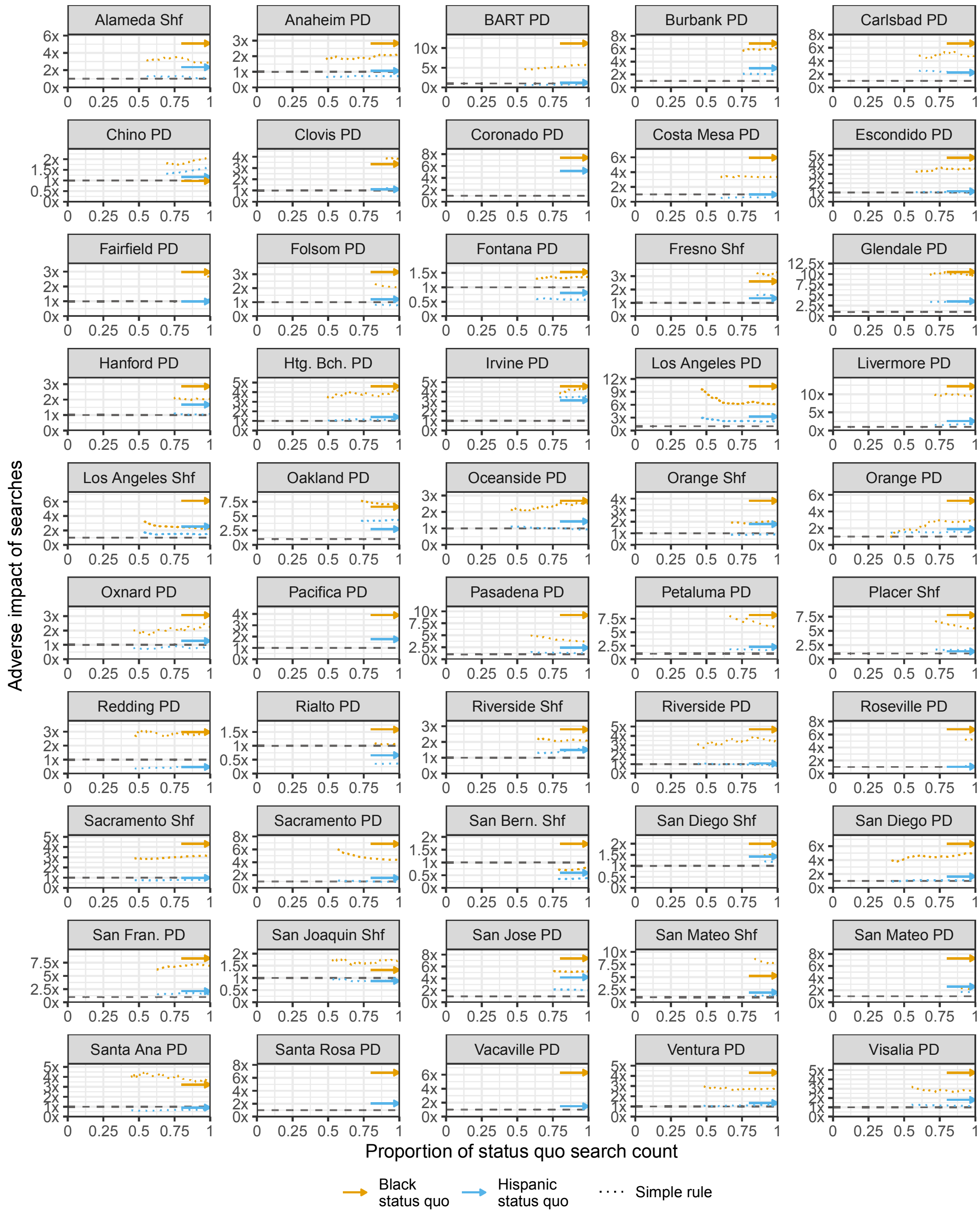


Figure A19: *Expanded version of Figure 5 for all 50 agencies.*

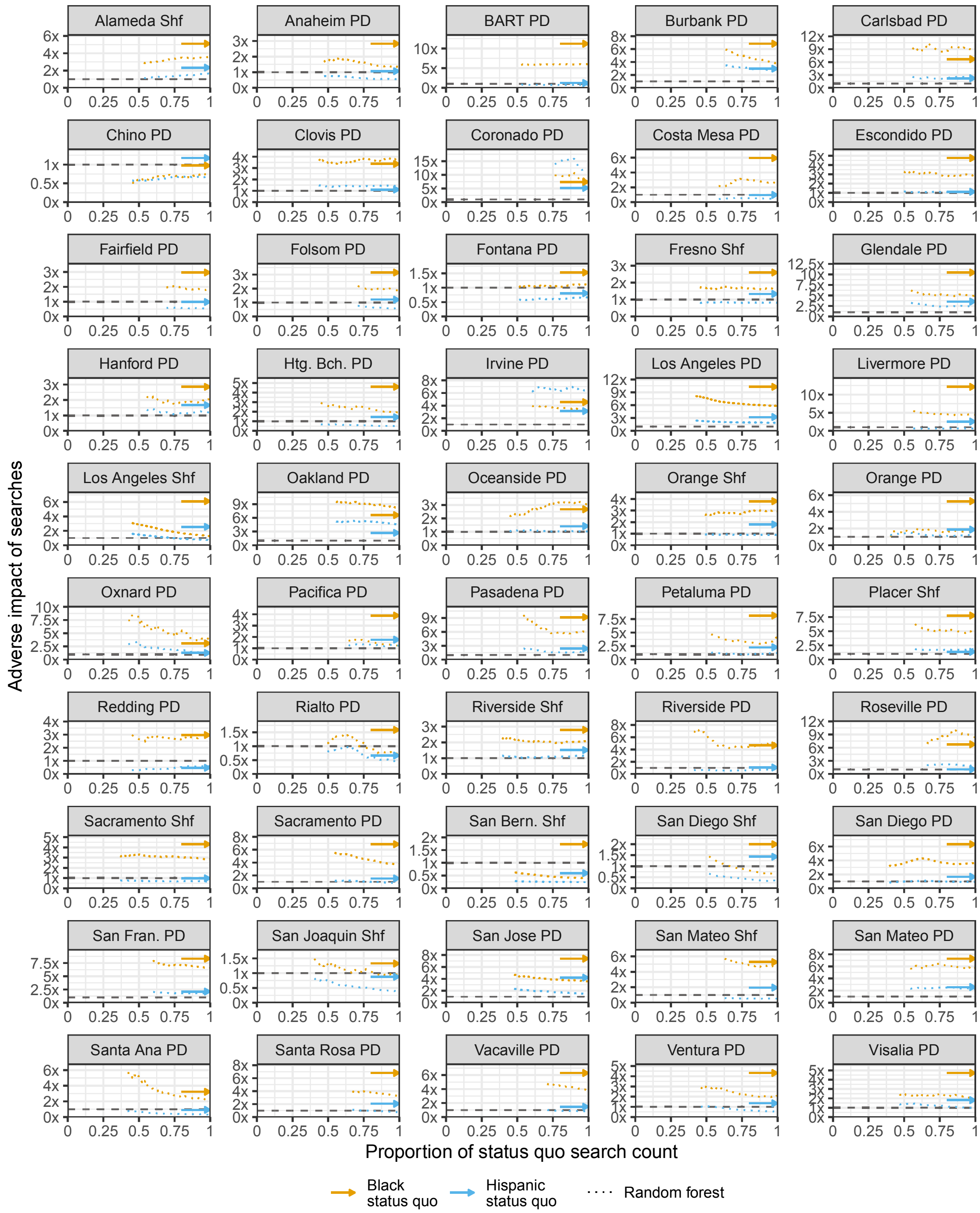


Figure A20: *Expanded version of Figure 5 for all 50 agencies. Instead of using a simple rule risk model, these plots use a random forest risk model to generate risk estimates.*

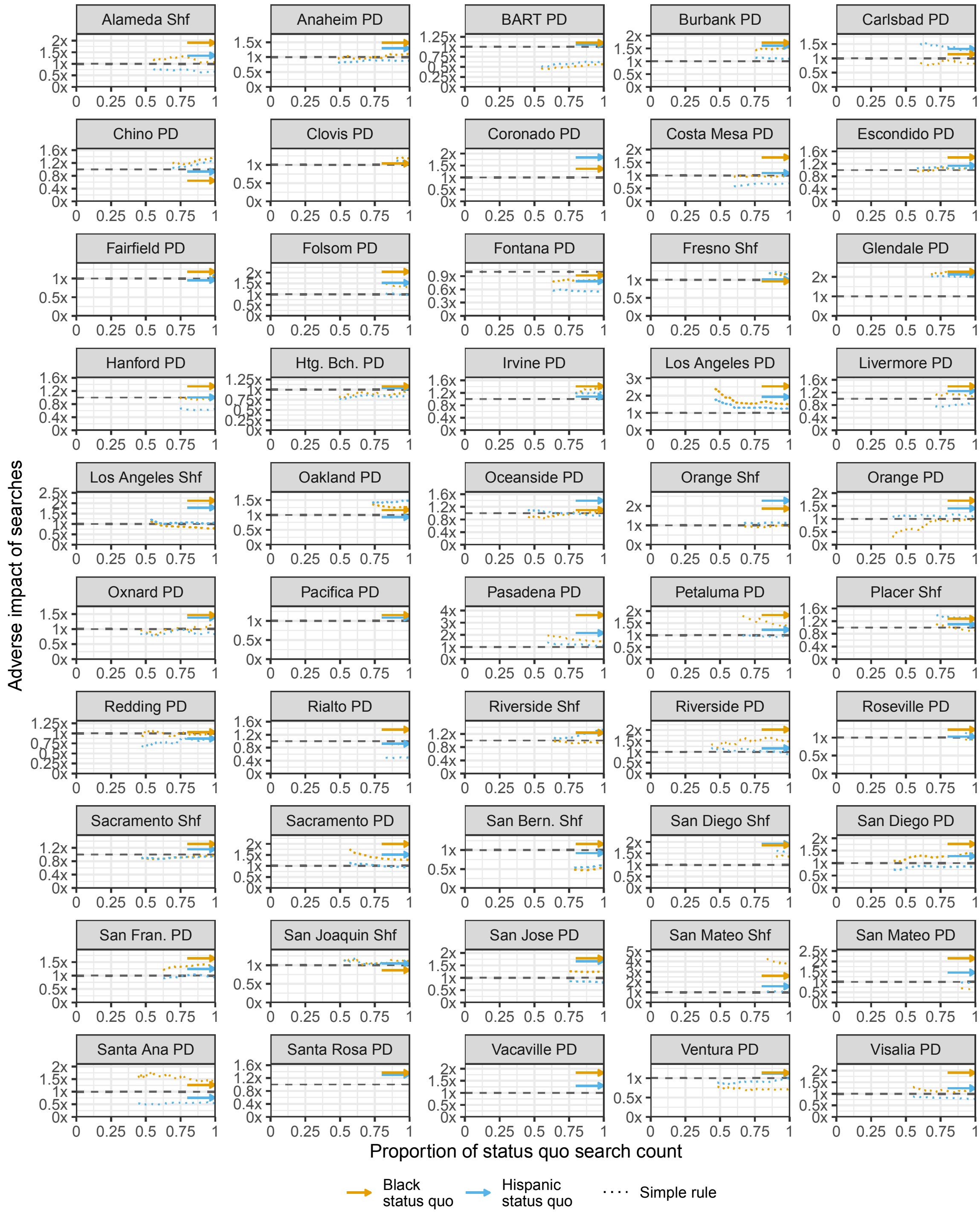


Figure A21: *Expanded version of Figure 5 for all 50 agencies. Instead of measuring adverse impact with a population-level benchmark, these plots use a stop-level benchmark (see Figure A1).*