Officer bias in stop and search is exacerbated by deployment decisions

Lara Vomfell[1,*] and Neil Stewart[1]

[1]Warwick Business School, University of Warwick, Coventry, UK
[*]Corresponding author: (l.vomfell@warwick.ac.uk; 0000-0001-5264-9897)

June 22, 2020

Abstract

Black and Asian people are more likely to be stopped and searched by police than white people. Following a panel of 36,000 searches by 1,100 police officers at a major English police force, we provide officer-specific measures of over-searching relative to two baselines: the ethnic composition of crime suspects officers interact with and the ethnic composition of the areas they patrol. We show that the vast majority of officers over-search ethnic minorities against both baselines. But we also find that the over-searching by individual officers cannot account for all of over-representation of ethnic minorities in stop and search: Over-patrolling of minority areas is also a key factor. Decomposing the overall search bias we find that the over-representation of Asian people is primarily due to over-patrolling while the over-representation of Black people is a combination of officer and patrol effects, with the larger contribution coming from officers' biases.

Introduction

Ethnic minorities are over-represented in police searches compared to white people. In England, Black and Asian people make up 11% of the population, yet they account for 30% of all English police searches, called stop and search[1].

Search decisions come with considerable consequences: searches can create feedback loops where an individual is repeatedly searched because they were searched in the past[2,3] which increases their likelihood of being arrested, thereby creating further feedback loops in the criminal justice system[4,5]. High levels of searches further result in diminished citizen engagement with police, diminished political engagement, reduced perceptions of police legitimacy and trust in police[6,7,8,9,10,11]. In addition, invasive search encounters can result in psychological harm to searched individuals, leading to increased symptoms of stress, anxiety and trauma[12,13,14].

It is therefore crucial to understand the reasons for the over-searching of ethnic minorities. Here we explore ethnic bias in search decisions at the officer level by focusing on individual officers' levels of bias and factors shaping these biases. Our approach is two-fold. First, we investigate officers' search biases against an ethnic group relative to two officer-specific baselines: the ethnic composition of crime suspects and of the areas they patrol. Second, we then examine the contributions of officers' search biases and of biases in deployment decisions to the over-representation of ethnic minorities in stop and search.

We demonstrate that the majority of officers over-search Asian and Black people, whichever baseline we compare their searches against. Our results show that officers perform more searches of ethnic minorities than can be explained by the ethnic composition of officers' interactions with crime suspects or with the population they patrol. However, over-searching by individual officers cannot account for all of the over-representation of ethnic minorities in stop and search. Over-patrolling is part of it: The median officer in our sample patrols areas which are 1.16 times more Asian and 1.37 times more Black than the West Midlands police force area. In other words, police officers are deployed to more ethnically diverse areas. Such deployment decisions contribute to the over-searching of ethnic minorities[15,16]. We find that these deployment decisions exacerbate the effects of individual officers' biases. Both together are responsible for the overall bias against ethnic minorities in stop and search.

Our approach and results connect to a rich literature on stop and search, and on ethnic bias in policing more generally. Stop and search in the United Kingdom is a widely used policing power characterised by police forces as a crucial tool to prevent and investigate crime[17,18,19]. If achieving these aims justifies persistent ethnic disparities is a long-standing debate—for example in the landmark Scarman and MacPherson reports—especially as the evidence on the effectiveness of stop and search is mixed[14,20,21,22,23,24].

Police frequently attribute the over-representation of ethnic minorities in stop and search to their over-representation in crime, implying that ethnic minorities perpetrate more crime than white people[17,18,19,25,26,27,28]. This argument can result in a self-fulfilling prophecy because the process of observing and recording crime already depends on the wider social context of policing. In this context, deployment decisions[15,29,30,31,32], arrest probabilities[5,33] and the accurate recording of crime[34,35] are not independent of ethnic group. As a consequence, crime data are not an objective benchmark of true criminal behaviour. In our work we do not take this into account for a simple reason: We are interested whether police officers' actions match the benchmarks they assemble themselves. Thus we compare officers' stop and search decisions to their own encounters with crime suspects.

In light of police's potential to criminalise minorities, the role of ethnic bias in police decision-making deserves further inquiry. Police officers operate within the tension between their roles as individual decision-makers and agents of the institution of the police, influenced by the organizational protocols and structures[17,36,37]. At the individual level, there is ample evidence of biased attitudes held by police officers[38,39,40,41,42,43] as well as of racially or ethnically motivated behaviour[44,45,46,47]. Ethnic bias at the institutional level is equally important. The Stephen Lawrence Inquiry in 1999 with its emphasis on institutional racism has sparked a varied discussion on the role of police forces in creating ethnic disparities in the United Kingdom[48,49,50]. Two factors have been highlighted in particular: First, structures within the police force perpetuate and broadcast biased beliefs through various hierarchies[17,23,48,50]. Second, deployment decisions by the police force—that is, decisions about which areas to prioritise and deploy officers to—are under scrutiny, given that these decisions can create disparities at the population-level, independently of how individual officers behave[22,23].

Officer teams are intermediaries between officers and the police force, often with their own norms and cultures[36,49]. A recent study noted remarkable differences between different teams within the same police force in England: Teams tasked with proactive policing not only performed the highest number of searches within the force but were also over-searching Black people at higher rates than other teams[51]. In our analysis we explore the relevance of officer teams by accounting for differences between teams and by including the ethnic composition of officers' teams into our model.

The tension between individual and institutional behaviour also applies to other policing activities such as drug enforcement[5,16], arrests[52] and use of force[53]. The literature on use of force in particular is currently debating an important consequence of this tension: What is the appropriate level of analysis of use of force data? We will briefly outline this debate since the analysis of stop and search data is characterised by the same tension and because our results can directly speak to an ongoing discussion within the use of force literature. In the United States, Black people are subject to higher rates of police use of force, particularly lethal use of force, than white people relative to their shares in the population[53,54]. Some studies have argued that the general population in an area is not the appropriate comparison: Instead one should compare rates of use of force to how often Black and white people come into contact with police[55,56,57]. After conditioning on the rate with which police encounter Black individuals, Fryer[57] finds a reversal of ethnic disparities: Police are apparently less likely to employ lethal force on Black people than white people.

An issue with this approach is pooling: Fryer[57]'s analyses are at the police department level, pooling all officers together. In response, Ross, Winterhalder and McElreath[58] develop a generative model in which officers are biased against Black people but differ in how often they encounter them. Small groups of officers which encounter Black people at high rates are sufficient to confound the pooled analysis and point toward a reversal of disparities in use of force. In other words, pooled analyses of use of force can fail to detect ethnic bias under heterogeneity[58]. Their finding calls back to Simpson's paradox where the pooling of heterogeneous groups, differing in their anti-Black bias for example, at the aggregate police force-level can reverse group-level patterns[58,59,60].

The pooling problem directly applies to pooled analyses of police searches. If officers differ in how often they encounter criminals of different ethnicities, then a police department level analysis of searches conditioned on crime can be confounded and fail to identify the direction of the disparity. Generally, analyses of searches tend to find over-representation of ethnic minorities even after conditioning on crime. For example, Gelman, Fagan and Kiss[61] find over-representation of Black and Hispanic people in pedestrians stops-and-frisks in New York City after adjusting for race-specific representations in crime, a pattern substantiated in other analyses[31,62]. In addition to pedestrian searches, traffic stops (where similar ethnic biases persist[63]) are often compared to benchmarks of criminal behaviour[64,65,66,67,68]. All of these analyses are performed at the police department level meaning that they could be potentially confounded.

Internal benchmarking is an officer-specific approach which matches each police officer to similarly-situated officers[69]. The officer's behaviour is possibly problematic if it deviates substantially from their peers'. A drawback of this method is that it can only reveal individual officers' biases relative to their peers. For example, only 15 out of 2,756 officers of the New York Police Department are flagged as potentially biased[69], far too few to explain the overall level of over-representation of ethnic minorities in stop and search.

In our study, we explore stop and search behaviour at the level of the individual officer, following a panel of officers over time. We compare an officer's searches of an ethnic group to the officer's direct experiences of the crime involvement of this group. By not pooling our data we thereby circumvent the issue of Simpson's paradox.

Our analysis is focused on so-called suspicion searches. In the United Kingdom, police officers routinely stop and question members of the public. During these unrecorded conversations, officers can ask individuals to account for themselves. If at any point the officers believe that the person has committed a crime, is about to commit a crime or is carrying illegal items such as weapons, drugs or burglary tools, the officers can initiate a search of the person's clothing and belongings. At this point, the encounter must be recorded in the form of a stop and search record detailing information about the searched person and the officer's justification for the search. At the end of the search encounter, the searched person has to be supplied with a reference number to the record. A search may be initiated only under powers requiring "reasonable grounds for suspicion" as detailed or with prior authorisation. In our analysis we restrict our attention to suspicion searches, which account for 99.4% of all searches, because only these searches are initiated at the discretion of the searching officer.

Our data consists of records of searches between 01/04/2014 and 30/09/2018 provided by West Midlands Police in England as well as all recorded crime in the same period. In our analysis we use self-defined ethnicity, which is someone's response to the question "What is your ethnic group?". We focus our analysis on Asian, Black and white people because sample sizes are too small for the remaining ethnic groups.

In our analysis we rely on two officer-specific baselines: the crime suspects an officer encounters and the residents in the officer's patrolling area. We obtain the crime suspect information by linking officers to the reported crime cases they responded to and then counting the person(s) suspected by police of having committed the offense. For the patrolling information, we calculate how often an officer visits a given geographical census unit using additional data and obtain an officer-specific patrol intensity share for the area. We use the smallest geographical unit provided by the 2011 ONS census, 2011 Output Areas[70]. We then multiply the number of residents in each census unit with the intensity share and sum them to obtain patrolling intensity-weighted counts of the residents in an officer's patrolling area.

Together our data form a panel of search counts, crime suspect counts and patrol counts for each officer over 9 half-year (6 month) intervals. Using a Bayesian Softmax regression model (sometimes also called Multinomial logit model) we then infer search shares p, crime suspect shares ζ and patrol population shares ρ for each officer i in time period t. They represent the share of each ethnic group e in the officer's searches, crime suspect encounters and patrol counts, respectively.

These shares form the basis for our two measures of search disparities:

1. D^S, the disparity of search relative to crime suspects. For each officer we obtain D^S_{ite} by dividing the officer's search share p of ethnic group e in time period t by the officer's suspect share ζ of e in t. If D^S is larger than 1 then the officer over-searches an ethnicity relative to how often they encounter the ethnic group as crime suspects. If D^S is smaller than 1 then the officer under-searches an ethnic group relative to suspects and if D^S is exactly 1 then the officer searches that ethnicity at the same rate as they appear in the officer's crime suspects.

2. D^P, the disparity of search relative to patrol. For each officer we obtain D^P_{ite} by dividing the officer's search share p of ethnic group e in time period t by the officer's patrol share ρ of e in t. D^P has the same interpretation as D^S: If D^P is larger (smaller) than 1 then the officer over-searches (under-searches) that ethnic group relative to the ethnic composition of the area they patrol.

For example, for the median officer Asian people make up 23% of their searches, 15% of the crime suspects they interact with and 23% of the areas the officer patrols. The disparity of search relative to crime for Asian people by this officer is $D^S = 0.23/0.15 \approx 1.53$ which means that the officer over-searches Asian people relative to crime suspects by a factor of 1.53. The disparity of search relative to patrol by this officer is $D^P = 0.23/0.23 = 1$, meaning that this officer searches Asian people about as much as they encounter Asian people on patrol.

Such disparities or biases are not equivalent to discrimination. Conclusively attributing empirical patterns of disparities to ethnic or racial discrimination is challenging[60, 71, 72]. We believe it is nonetheless important to uncover, document and dissect ethnic disparities because differential rates of contact with police entail far-reaching consequences, not least the legitimacy of the institution of police. In our study we make two important contributions to the literature on ethnic bias in policing: First, we provide officer-specific measures of search bias relative to the crimes suspects an officer encounters and relative to the population in the area the officer patrols. Second, we find that officers' search biases are smaller than search bias on the police-force level, suggesting that deployment decisions contribute to the overall search bias against ethnic minorities in stop and search.

Results

We perform Bayesian inference. Before seeing the data, we have prior information about likely values of the parameters which are updated with the likelihood of the data to obtain the posterior distribution. A sample from the posterior is a plausible parameter value consistent with the prior information and observed data. We provide 90% uncertainty intervals for the parameters, sometimes also called credible intervals[73]. 90% of our posterior distribution over the parameter lies within the 90% uncertainty interval.

We present our results in three parts: (i) estimation of search shares, (ii) measures of disparity D^S and D^P and (iii) the discrepancy between officer-level and force-level search bias.

Inference of search shares

We infer p_{ite}, the share of each ethnic group e in officer i's searches in time period t, as a function of the officer's suspect shares and patrol shares in time period t, and their gender, age, experience, ethnic group and the share of white officers in their team. The model is described more formally in the Data & Methods section.

Figure 1 shows the posteriors of p_{ite} for each ethnic group over all officers and time periods based on the full model. Due to the aggregation over officer-specific posteriors

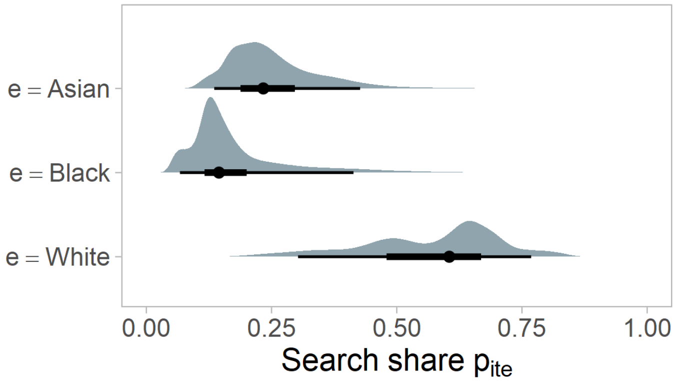

Figure 1: Posterior densities of search shares p_{ite} over all 1,194 officers and 9 time periods from the full model. Search shares are the proportion of each ethnic group in the officer's searches. The black dot represents the medians of the distributions aggregated over e and t and represent search shares for the median officer. Black lines show 50% and 90% uncertainty intervals which represent the spread of beahviour by 50% and 90% of the police officer workforce. There are multiple modes in the posteriors which simply means that there are different clusters of officers with similar search shares. (For a version disaggregated by time see Figure A.1.)

they represent the (posterior) behaviour of the entire workforce of searching officers and show that searches by the median police officer are 23% Asian, 15% Black and 60% white.

As explained above, we infer the search shares as a function of officer and team characteristics and the officer's suspect and patrol shares. To do this, we infer each officer's propensity to search Asian and Black people, called θ_{Asian} and θ_{Black}, and then transform these propensities into search shares. In Figure 2, we show the posteriors of these coefficients. We find no credible evidence that officer age and ethnicity have an effect on search shares. There are minor effects of officer gender and experience where female or experienced officers search fewer ethnic minorities. Relative to the other effects, they are scarcely meaningful.

Instead, we find associations of search shares with officer-level suspect and patrol shares. The effect of Asian suspect shares is positive, meaning that officers with a higher share of Asian crime suspects also have a higher Asian search share. Interestingly, the effect of Black suspect shares is negative, meaning that officers who encounter more Black crime suspects search fewer Black people. The effect of patrol shares is more intuitive: the ethnic composition of searches reflects that of the areas officers patrol. In principle, this admits two competing hypotheses: Either, officers are searching at random or they explicitly adjust for the population in their patrol areas. As we demonstrate in the next section however, officers over-search ethnic minorities relative to their patrolling areas which suggests that officers do not search at random.

Last, we comment on our team-level results. In predominantly white teams, Asian and

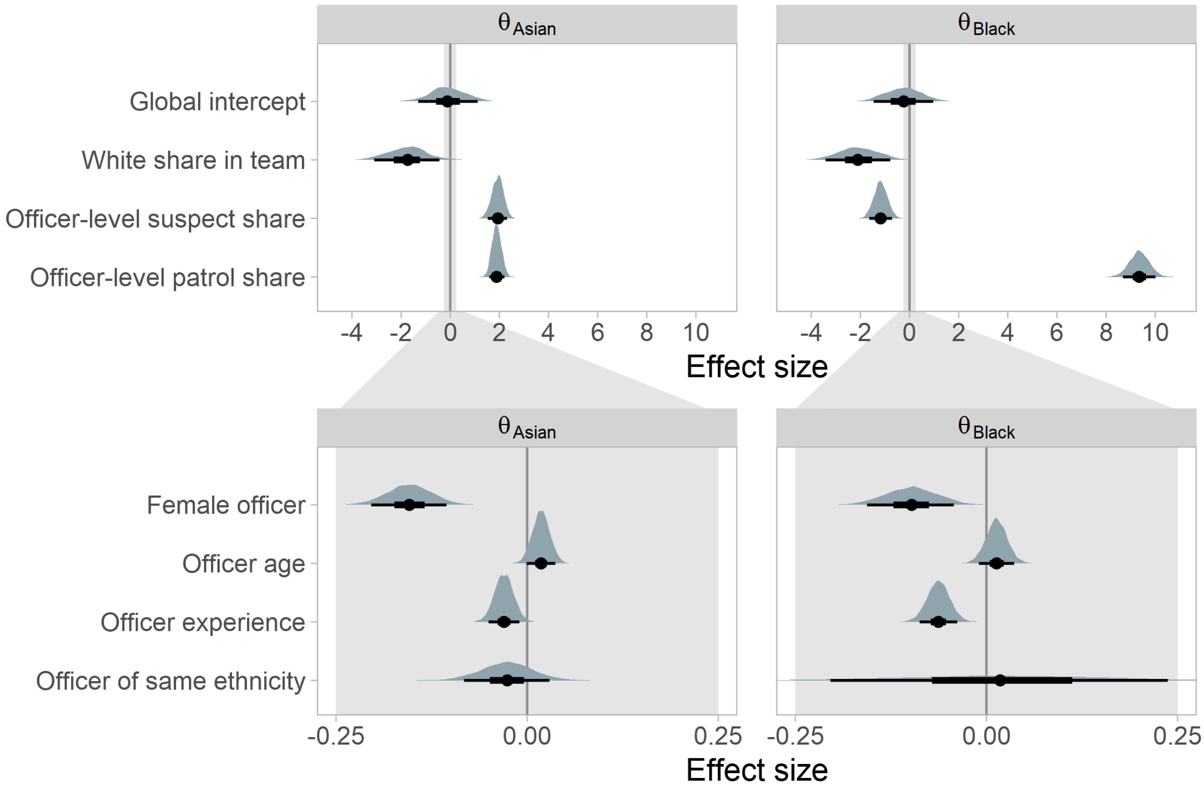

Figure 2: Posterior densities of the coefficients used to infer search shares p. A positive effect on θ_{Asian} means a larger Asian search share. Similarly, a positive effect on θ_{Black} implies a larger Black search share. Black dots show the median of the posteriors while black lines show 50% and 90% uncertainty intervals which contain 50% and 90% of the posterior distribution. For visual clarity we show the effects of officer gender, age, experience and ethnicity on a zoomed in scale of [-0.25, 0.25] compared to the other effects. A table of the medians and 90% uncertainty intervals is available (Table A.1).

Black people make up a lower share of searches than in more ethnically diverse teams. While it would be preferable to differentiate between Black and Asian officers, we have to treat them as a single group in the analysis as there are at most 4 Black police officers in a team. Our Bayesian model includes team-specific intercepts to account for differences in search shares between teams. The results show that the ethnic composition of searches varies considerably between teams as evidenced by the intercepts' standard deviations. Specifically, they are 0.35 (90% UI [0.27, 0.46]) for Asian searches and 0.43 (90% UI [0.34, 0.57]) for Black searches. Presumably, these differences are due to team specialisation, as officers' routines are determined by their responsibilities.

Measures of Disparity

Next, we discuss disparity of search relative to crime suspects (D^S) and disparity of search relative to patrol (D^P). We first show the posteriors of D^S and D^P in Figure 3. Again, the distributions represent the aggregate over officer-specific posteriors and, as such, the behaviour of the entire workforce of searching officers in our sample. The median officer over-searches Asian people by a factor of 1.56 (90% UI [0.80, 6.92]), Black people by a factor of 1.41 (90% UI [0.41, 12.37]) and under-searches white people by a factor of 0.84 (90% UI [0.51, 1.23]) relative to suspects. The uncertainty intervals for Asian and Black disparities are wide on the aggregate because they also are wide on the officer-level. The interpretation is that we are uncertain about the precise level of officers' search bias against ethnic minorities relative to suspects, but officers are more likely to over- than under-search Asian and Black people. In contrast, the results for white disparities are clear: More than half of officers under-search white people relative to suspects.

We can be more confident about the actual levels of disparity of search relative to patrol. The right-hand side of Figure 3 shows that the median officer over-searches Asian people by a factor of 1.03 (90% UI [0.58, 2.24]), Black people by a factor of 1.78 (90% UI [1.06, 4.40]) and under-searches white people by a factor of 0.88 (90% UI [0.64, 1.25]) relative to patrol.

The summaries of D^S and D^P presented so far are coarse: They only allow us to make statements about the aggregate of all officers. To refine the resolution, we compute the posterior probability that an individual officer over-searches a particular ethnic group, both relative to suspects and patrol from the posteriors of the officer-specific disparities. For example, if this probability is 1, then the officer always over-searches. Similarly, if this probability is 0.5, the officer's search shares perfectly match the suspect or patrol baselines.

Figure 4 shows histograms of these probabilities for all officers. The left-hand side is in line with what we have already seen on the aggregate in Figure 3: Most officers over-search Asian and Black people while virtually all officers under-search white people

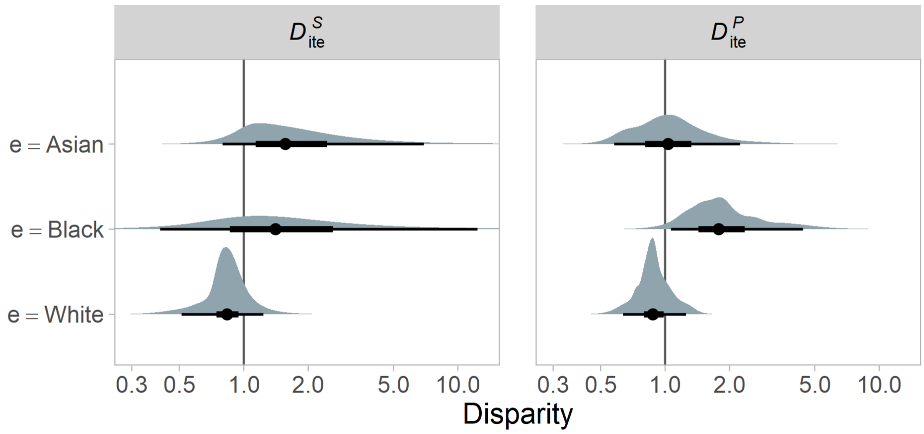

Figure 3: Posterior densities of D^S_{ite} and D^P_{ite} aggregated over all officers and time periods. The distributions represent the behaviour of the entire workforce of searching officers. The black dots represent the median officers and the black lines represent 50% and 90% uncertainty intervals. For visual clarity, we only show values between $[0.3, 13]$. 1.65% of all posterior probability is excluded by this choice. (For a version disaggregated by time see Figure A.2.)

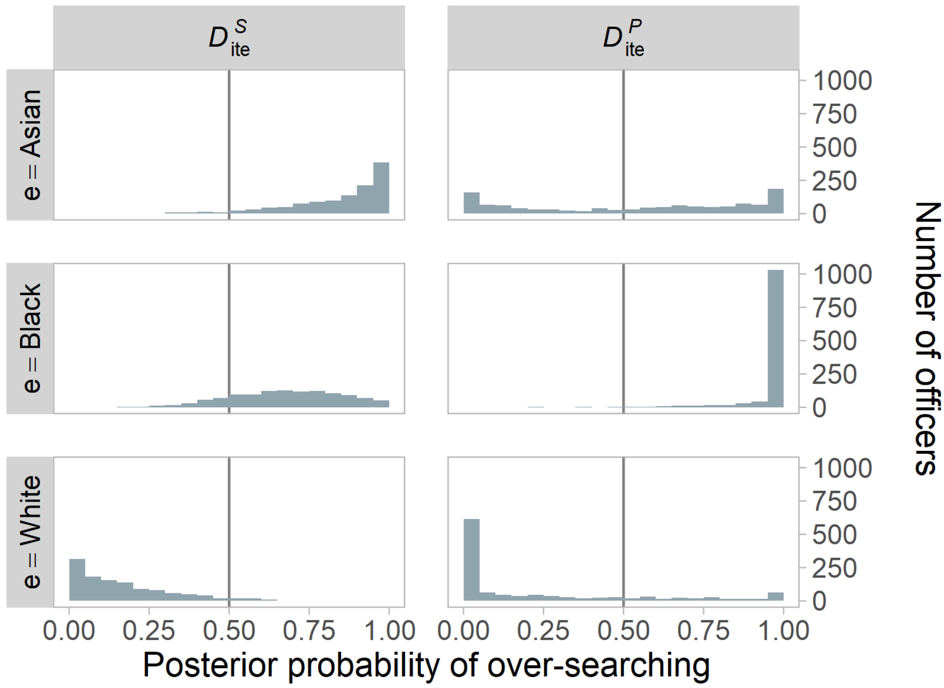

Figure 4: Histograms of the posterior probabilities of D^S_{ite} and D^P_{ite} above 1 for each officer. If the posterior probability above 1 for an officer is 1, the the officer always over-searches an ethnic group. If the posterior probability above 1 for an officer is 0.5 then the officer's search shares perfectly match the suspect or patrol baselines.

relative to suspects. However, the right-hand side of Figure 4 reveals a pattern that would be left obscured by only studying the aggregate. Particularly, we observe a split between officers: Some officers under-search Asian people, while others consistently over-search them relative to patrol. Since these officer groups are of roughly the same size, the aggregate incorrectly suggests that officers do not over-search Asian people. In contrast, the officer-level results for Black and white people match the aggregate: Virtually all officers over-search Black people relative to patrol. In fact, 86% of the officers have a posterior probability of over-searching Black people that exceeds 0.95. Similarly, the vast majority of officers under-search white people relative to patrol. There is no change and no discernible dependence in D^S and D^P over time, a point we explore in more detail in the appendix.

Officer- compared to force-level bias

Last, we discuss the implications of our officer-level results on the overall over-representation of ethnic minorities in stop and search. The median officer patrols more ethnically diverse areas than are representative for the police force's area of operation. For the remainder of the analysis, we only consider over- and under-searching relative to the patrolling baseline. This is because while police forces have direct control over patrolling decisions, the same cannot be said for the ethnic composition of suspects they encounter. Thus, analysing patrolling decisions allows us to decompose over-searching into officer- and force-level decision making.

Our analysis so far treats the officer patrolling areas as given. However, patrolling areas are not allocated at random. Rather, police departments' deployment decisions are the consequence of prioritising certain areas. Similarly to how we constructed an officer-specific measure of over-searching relative to patrol, we can construct a force-level measure of over-searching relative to population share. This allows us to multiplicatively decompose force-level over-searching into three factors: officer bias, over-patrolling and the aggregation discrepancy between officer- and force-level.

$$\begin{array}{ccccccc} \text{Over-searching} & = & \text{Officer bias} & \times & \text{Over-patrolling} & \times & \text{Aggregation discrepancy} \\ \dfrac{\text{Force search share}}{\text{Population share}} & = & \dfrac{\text{Officer search share}}{\text{Officer patrol share}} & \times & \dfrac{\text{Officer patrol share}}{\text{Population share}} & \times & \dfrac{\text{Force search share}}{\text{Officer search share}} \end{array}$$

Officer bias is just D^P—our measure of officer over-searching by an officer relative to patrol. Over-patrolling is the disparity between the individual officer's patrol share and the population share in the police force area. Last, the aggregation discrepancy

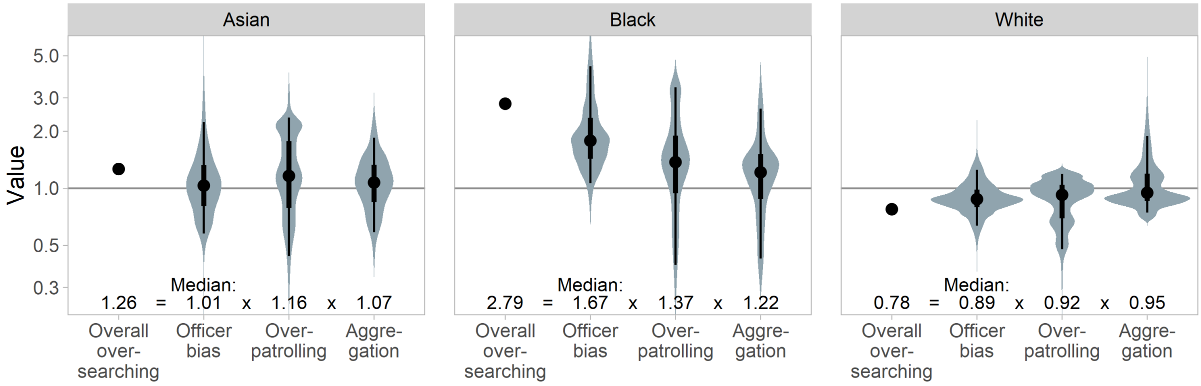

Figure 5: Decomposition of over-searching into officer over-searching, over-patrolling and aggregation discrepancy. Grey areas show posterior densities of terms calculated based on officers' entire posterior distributions. The text at the bottom are results for the median officer.

is the disparity between the force-level search share and the officer's individual search share. The aggregation discrepancy just gives how different this officer is from the force-level search share of Asian people due to how force-level searches are the aggregation of officers' searches.

For example, we can decompose the over-searching of Asian people based on the medians of these three terms. Relative to population, Asian people are over-searched by a factor of $0.2506/0.1982 \approx 1.26$, which is their share in all searches by the police force divided by their population share. Median officer over-searching is $0.2335/0.2304 \approx 1.01$, median over-patrolling is $0.2304/0.1982 \approx 1.16$ and the aggregation discrepancy is $0.2506/0.2335 \approx 1.07$.

Of course, any summary based on medians alone would be unsatisfactory. We therefore study the distributions over these three terms as induced by the officer-specific posteriors. On a practical level, this entails calculating them for every draw from each officer-specific posterior, the result of which is shown in Figure 5. Note that the distributions of officer over-searching shown in Figure 5 are the same as in Figure 3. At this point, it is important to recall that aggregated officer over-searching of Asian people obscures that some officers over- and some officers under-search Asian people relative to patrol which "cancels out" on the aggregate, resulting in a median of 1.01. This is only a concern for Asian over-searching since only there did the officer-level patterns differ from the aggregate. Taken together, over-searching of Asian people is driven by a combination of officers over-searching and over-patrolling Asian people but over-patrolling has a larger overall impact.

Black over-searching decomposes differently: Relative to the population Black people are over-searched by a factor of 2.79 which is primarily due to officer over-searching. Still, over-patrolling also contributes to the overall over-searching of Black people relative to population. Last, we find that white people are under-searched relative to population. This is primarily due to officers under-searching white people but also due to under-patrolling of white areas.

Discussion

Ethnic minorities are over-represented in stop and search compared to both their representation in the population and in crime. Our analysis exploits a panel of officers' searches from a major police force in England. We investigate the role of individual officers and police structures in the over-searching of ethnic minorities in stop and search.

For each officer, we first infer officer-specific search shares—the share of an ethnic group in an officer's searches. The ethnic composition of officers' searches is not meaningfully explained by officer characteristics. For example, the ethnicity of the officer has almost no effect on their searches, which matches some the mixed literature on the effect of officer ethnicity on policing outcomes[74,75,76,77] and differs from some of it[78,79]. Instead, the ethnic composition of officers' crime suspect encounters (suspect share) and of the officers' patrolling areas (patrol share) are associated with the ethnic composition of searches.

In exploring team compositions, we uncover a nuanced effect of officer ethnicity. We find evidence that teams' ethnic compositions influence officers' search behaviour: Teams that are more homogeneously white search fewer ethnic minorities. Officers preferring to interact (or being tasked with interacting) with members of their own ethnicity alone cannot explain this effect because more diverse teams search more Black people, yet most of this diversity is due to Asian officers and not Black officers, of whom there are very few.

In a second step, we infer an officer's bias of over-searching an ethnic group relative to crime suspects or to patrol. Almost all officers over-search Black people both relative to how they encounter them as crime suspects and relative to the areas officers patrol. Similarly, almost all officers under-search white people relative to crime suspects or to patrol. For Asian people, we find that almost all officers search Asian people more than they encounter them as crime suspects. Relative to their patrol areas however, the picture shifts and officers are split into two groups, one that over-searches and another that under-searches Asian people which cancels out on the aggregate.

Such disproportionate contact with police relates back to use of force. Ross, Winterhalder and McElreath[58] demonstrate that pooled analyses of use of force conditioned on the rates with which police encounter civilians can be confounded if officers differ in how often they encounter minorities. We find that officers indeed differ in how often they come into contact with ethnic minorities (for example, by searching them) and this cannot be explained by differential crime rates. Furthermore, even if officers were to use force on ethnic groups equally conditional on coming into contact with them, the fact that they have more contact with ethnic minorities means that these groups are subjected to higher levels of police use of force (Eckhouse, unpublished manuscript). Of course, this not only applies to use of force but also other policing activities such as misdemeanour enforcement or arrests and emphasises the importance of documenting these disparities.

Regarding our findings of over-searching minorities relative to patrol, it is important to note that the patrol share is based on residential data from the 2011 ONS Census. The population available on the street, the 'available population', can be markedly different from the residential population[80]. In particular, the ethnic make up of the available population can be different from the residential population and potentially account for the bias against ethnic minorities[81,82]. On the other hand, the available population explanation can be another self-fulfilling prophecy similar to the crime explanation[83,84]: If officers are deployed to areas with ethnically diverse available populations then the available population will predictably 'explain away' the bias compared to the residential population. That does not make the deployment decision bias-free. Additionally, search decisions have to be based on sufficient grounds that a specific person is suspicious, not general availability of an ethnic group[22,23].

Deployment decisions are relevant to our analysis. Minority communities are over-patrolled: The median officer patrols an area which is 1.16 times more Asian and 1.37 times more Black than all of West Midlands. The overall over-representation of ethnic minorities in stop and search decomposes into officer bias and over-patrolling. With officers over-searching minorities and command deploying officers to more diverse areas, the effects of officer biases are exacerbated by these deployment decisions. This results in more over-searching of minorities than can be attributed to officer biases alone.

The over-policing of minority communities documented in our study is supported by a wide range of other studies finding the same phenomenon[15,31,32,85]. Addressing the common question if these deployment biases can be explained by crime patterns is difficult. By their presence in an area, police are more likely to observe and record crime there. The observation of crime then is not independent from patrolling and searching patterns (and the ethnic biases therein). With the data available to us we cannot make any statement as to the mechanism that causes minority areas to be over-patrolled or the role of crime in that. Here, we only note that over-patrolling accounts for a considerable part of the overall over-searching of ethnic minorities.

There are clear limitations to our analysis, especially related to the generality of our findings. The policing context in the United Kingdom is particular, due to public and political scrutiny of police forces and the specific nature of the relationship between minority communities and the police. More officer-level analyses are needed and we hope that more police forces make officer-level data available to researchers. Furthermore, we hope that future work can clarify the process of deployment decisions.

For policy-makers, police forces and advocates looking to address the over-representation of ethnic minorities in stop and search, our results are both concerning and promising. Concerning, because our results show that 1. officer bias is a key factor in the over-representation of ethnic minorities in stop and search and 2. this officer bias is exacerbated by where police officers are deployed to. Promising, because our results could mean a multiplier effect of institutional change. Clearly, police forces should carefully examine their deployment policies as an amplifier in the over-representation of ethnic minorities in stop and search. Additionally though, we find that teams' ethnic compositions impact the composition of officers' searches. Addressing the norms and environment of officer teams could then change officers' behaviour rather than just reduce its effect[51]. Our work shows that police forces need to reconcile the tension between officer behaviour and department-level decisions in creating ethnic disparities in stop and search.

Data & Methods

Data

Our data covers the period between 01/04/2014 and 30/09/2018 which we split into nine periods of 6 months each, beginning from 01/04/2014. We chose this time resolution because periods shorter than 6 months result in sparse officer-level information. Officers which performed searches in fewer than 50% of the half-year periods, i.e., in fewer than five half-year periods out of the nine in our study period were excluded to avoid data sparsity issues. The final file covers 1,194 officers observed in 29 teams, 203,176 reported crimes and 36,028 searches.

In our model, we use the following variables: counts of officers' searches; counts of officers' crime suspect encounters; counts of residents in officers' patrolling areas, all broken down by ethnic group; officer gender (dummy encoded); officer age; officer experience and two dummy variables indicating whether officer i is Asian or Black. We standardise officer age and officer experience to have mean 0 and standard deviation 1. We summarise these variables in Table 1.

Officers transfer between teams during our study period. We account for this in our model with the team-specific intercept α_j. All officers are assigned to the team j they were part of for the majority of the time in each 6-month period.

Multinomial model

Our data are counts of searches of ethnic group e by officer i in time period t. For each officer we thus have a vector $Y_{it} \in \mathbb{N}_0^E$ where $E = 3$ are the three ethnic groups we

Table 1: Means, standard deviations (SD), minima and maxima of the variables in the final data file.

Variable	Mean	SD	Min	Max
Time-varying variables per half-year				
Search counts				
Asian	2.41	4.55	0.00	76.00
Black	1.72	4.13	0.00	94.00
White	5.39	9.18	0.00	125.00
Suspect counts				
Asian	10.61	12.30	0.00	80.00
Black	8.29	8.87	0.00	97.00
White	43.41	40.71	0.00	212.00
Patrolling counts				
Asian	85.56	46.95	0.00	349.00
Black	32.30	19.45	1.00	145.00
White	200.84	49.62	37.00	343.00
Officer age in years	37.78	7.34	19.08	62.00
Standardized officer age	0.00	1.00	-2.55	3.30
Officer experience	10.81	5.25	0.17	30.42
Standardized officer experience	0.00	1.00	-2.02	3.73
Fixed variables				
Female officer	0.12	0.32	0.00	1.00
Asian officer	0.06	0.25	0.00	1.00
Black officer	0.01	0.10	0.00	1.00
Number of observed half-years per officer	8.46	1.14	5.00[a]	9.00
Share of White officers in team	0.93	0.06	0.81	1.00
Total number of observations (officers × half-years)	N =	10,103		
Number of officers	N =	1,194		
Number of teams	N =	29		

[a] We exclude officers with fewer than 5 half-years' worth of observations (see Data section)

consider: Asian, Black and white, which we abbreviate to A, B, W for ease of notation.

We are then interested in the proportions of each ethnic group in the total number of searches by officer i in t as a function of covariates. Formally, we model the allocation of total number of searches by i in t, $\sum_{e\in\{A,B,W\}} Y_{ite}$ (shortened to $\sum_e Y_{ite}$ for ease of notation), into $E = 3$ ethnic groups as follows:

$$Y_{it} \sim \text{Multinomial}\left(\sum_e Y_{ite}, p\right), \qquad p = \text{Softmax}(\theta_{it}). \tag{1}$$

In words, Equation (1) states that each observation vector Y_{it} is modeled by the the vector $\theta_{it} \in \mathbb{R}^E$ where θ_{it} gives an officer's propensity to search ethnic group e as a function of some covariates. To obtain valid proportions, we use the Softmax($\cdot$) function which normalises a vector of real numbers into a vector of proportions that sum to 1. This means that $p = \text{Softmax}(\theta_{it})$ gives the proportion of each ethnic group e in $\sum_e Y_{ite}$, the quantity of interest.

However, θ_{it} is not yet identifiable because the same values of $p = \text{Softmax}(\theta_{it})$ can be induced by different θ_{it}. This is easily resolved by setting $\theta_{it\,\text{white}} = 0$. In doing so, $\theta_{it\,\text{Asian}}$ and $\theta_{it\,\text{Black}}$ then represent an officer's propensity to search Asian or Black individuals compared to searching white people and θ_{it} is uniquely identified.

We model θ_{it} as a function of the demographic covariates listed in Table 1. The coefficients of these covariates represent their relative contribution to an officer's propensity to search Asian or Black people over white people. In modelling θ_{it} we are particularly interested in the contribution of an ethnic group's proportion in the officer's crime suspect population and the contribution of an ethnic group's proportion in the officer's residential population in the patrolling area.

We observe a vector of counts of crime suspects and a vector of counts of residents encountered on patrol. We then infer the proportions of each group in those vectors. To this end, we introduce four additional terms: S_{it}, ζ_{it}, P_{it} and ρ_{it}. Similarly to Y_{it}, $S_{it} \in \mathbb{N}_0^E$ is a vector holding counts of crime suspect encounters by officer i in t for $E = 3$ ethnic groups. Because we do not use any covariates to model the allocation of S_{it}, we can directly model the proportions rather than using the Softmax($\cdot$) transformation from before. ζ_{it} is the vector directly giving the suspect shares, that is the proportions of each ethnic group e in S_{it}. The remaining two terms follow the same logic: P_{it} gives counts of residents encountered on patrol by officer i in time period t. ρ_{it} directly models the patrol shares—the proportions of each ethnic group in P_{it}. More formally,

$$
\begin{aligned}
S_{it} &\sim \text{Multinomial}\left(\sum_e S_{ite}, \zeta_{it}\right), \\
P_{it} &\sim \text{Multinomial}\left(\sum_e P_{ite}, \rho_{it}\right).
\end{aligned}
$$

Taken together, this corresponds to the following model:

$$
\begin{aligned}
\theta_{it\,\text{Asian}} &= \alpha_{j[it]A} + \beta_A\, x'_{it\,A} + \gamma_A\, \zeta_{itA} + \delta_A\, \rho_{itA} + \omega_A w_{j[it]} \\
\theta_{it\,\text{Black}} &= \alpha_{j[it]B} + \beta_B\, x'_{itA} + \gamma_B\, \zeta_{itB} + \delta_B\, \rho_{itB} + \omega_B w_{j[it]} \\
\theta_{it\,\text{white}} &= 0,
\end{aligned}
$$

where $\alpha_{j[it]e}$ is an ethnicity-specific group-level intercept corresponding to the team officer i was part of in time period t. x'_{ite} is a vector holding i's covariate information at t specific to ethnic group e. $w_{j[it]}$ gives the share of white officers in the team officer i was in in time period t.

Modeling suspect and patrol shares as the allocation of suspect and patrolling counts allows us to account for measurement error. For example, if an officer encounters only few crime suspects, then the uncertainty in the suspect shares will be large because the estimates are based on few data points. The uncertainty in the shares will then be propagated forward to the inference on γ and δ such that noisier, less certain shares receive less weight than shares inferred from sufficient amounts of data.

We specify prior distributions on model parameters as follows: The group-level intercepts $\alpha_j \sim N(\mu_\alpha, \sigma_\alpha)$ where $\mu_\alpha \sim N(0,1)$ and $\sigma_\alpha \sim N^+(0,1)$ (half-normal) and all regression coefficients $\beta, \gamma, \delta, w \sim N(0,2)$. For ζ we use weakly informative Dirichlet priors parametrised with the respective share of each ethnic group in all arrests in England in the year 2016/17. (The Home Office does not publish crime by ethnicity.) This yields the prior $\zeta \sim \text{Dirichlet}(0.43, 0.61, 5.00)$ corresponding to country-wide shares of (0.07, 0.10, 0.82). Similarly, for ρ we use the share of each ethnic group in England in the 2011 ONS census: $\rho \sim \text{Dirichlet}(0.39, 0.21, 5.00)$ which corresponds to shares of (0.07, 0.04, 0.89)[86,87].

We fit the full model with `Stan` in R version 3.6.3 using `rstan` version 2.19.3[88,89]. Hamiltonian Monte Carlo sampling was performed on four chains with each 1,000 warm-up draws and 1,000 sampling draws, resulting in 4,000 draws from the posterior distribution in total.

Code and data availability

This research is based on data resources provided by West Midlands Police. Data were originally collected as part of routine police record keeping. The data are not available publicly and were provided to the authors under an Information Sharing Agreement with West Midlands Police. Under the terms of this agreement, the authors are not at liberty to share the data. Other researchers can contact West Midlands Police to obtain a data sharing agreement.

All code used to produce the results is available online on Github at `https://github.com/laravomfell/ethnic_bias_stop_and_search`. Since the original data from West Midlands Police may not be shared publicly, we generate synthetic data to demonstrate our code. The repository includes a folder `/data` which contains the synthetic data as well as the file `code/generate_synthetic_data.R` used to generate the data. The distributions of the variables in the synthetic data match the distributions in our data.

References

[1] Home Office. *Police powers and procedures England and Wales year ending 31 March 2018* (Home Office Statistical Bulletin, London, 2018).

[2] Quinton, P. The formation of suspicions: Police stop and search practices in England and Wales. *Policing and Society* **21**, 357–368 (2011).

[3] Brayne, S. Big data surveillance: The case of policing. *American Sociological Review* **82**, 977–1008 (2017).

[4] May, T., Gyateng, T. & Hough, M. *Differential treatment in the youth justice system* (Equality and Human Rights Commission, London, 2010).

[5] Kohler-Hausmann, I. Misdemeanor justice: Control without conviction. *American Journal of Sociology* **119**, 351–393 (2013).

[6] Sharp, D. & Atherton, S. To serve and protect? The experiences of policing in the community of young people from black and other ethnic minority groups. *The British Journal of Criminology* **47**, 746–763 (2007).

[7] Lerman, A. E. & Weaver, V. Staying out of sight? Concentrated policing and local political action. *The ANNALS of the American Academy of Political and Social Science* **651**, 202–219 (2014).

[8] Tyler, T. R., Fagan, J. & Geller, A. Street stops and police legitimacy: Teachable moments in young urban men's legal socialization. *Journal of Empirical Legal Studies* **11**, 751–785 (2014).

[9] Bradford, B. Unintended consequences. In Delsol, R. & Shiner, M. (eds.) *Stop and Search: The Anatomy of a Police Power*, 102–122 (Palgrave Macmillan, Basingstoke, 2015).

[10] Bradford, B. *Stop and search and police legitimacy* (Routledge, London, 2017).

[11] Laniyonu, A. Police, politics and participation: The effect of police exposure on political participation in the United Kingdom. *The British Journal of Criminology* **58**, 1232–1253 (2018).

[12] Skogan, W. G. Asymmetry in the impact of encounters with police. *Policing & Society* **16**, 99–126 (2006).

[13] Geller, A., Fagan, J., Tyler, T. & Link, B. G. Aggressive policing and the mental health of young urban men. *American Journal of Public Health* **104**, 2321–2327 (2014).

[14] Delsol, R. Effectiveness. In Delsol, R. & Shiner, M. (eds.) *Stop and Search: The Anatomy of a Police Power*, 79–101 (Palgrave Macmillan, Basingstoke, 2015).

[15] Sampson, R. J. & Lauritsen, J. L. Racial and ethnic disparities in crime and criminal justice in the United States. *Crime and Justice* **21**, 311–374 (1997).

[16] Shiner, M., Carre, Z., Delsol, R. & Eastwood, N. *The colour of injustice: 'Race', drugs and law enforcement in England and Wales* (StopWatch & Release, London, 2018).

[17] Shiner, M. Post-Lawrence policing in England and Wales: Guilt, innocence and the defence of organizational ego. *The British Journal of Criminology* **50**, 935–953 (2010).

[18] Equality and Human Rights Commission. *Stop and think again, towards equality in police PACE stop and search* (EHRC, London, 2013).

[19] The Centre for Social Justice. *It can be stopped–A proven blueprint to stop violence and tackle gang and related offending in London and beyond* (The Centre for Social Justice, London, 2018).

[20] Scarman, J. *The Brixton Disorders, 10–12th April (1981)* (Her Majesty's Stationery Office, London, 1981).

[21] Macpherson, W. *The Stephen Lawrence Inquiry* (Home Office, London, 1999).

[22] Delsol, R. & Shiner, M. Regulating stop and search: A challenge for police and community relations in England and Wales. *Critical Criminology* **14**, 241–263 (2006).

[23] Bowling, B. & Phillips, C. Disproportionate and discriminatory: Reviewing the evidence on police stop and search. *The Modern Law Review* **70**, 936–961 (2007).

[24] Tiratelli, M., Quinton, P. & Bradford, B. Does stop and search deter crime? Evidence from ten years of London-wide data. *The British Journal of Criminology* **58**, 1212–1231 (2018).

[25] Quinton, P., Bland, N. & Miller, J. *Police stops, decision-making and practice* (Home Office London, 2000).

[26] Rudovsky, D. Law enforcement by stereotypes and serendipity: Racial profiling and stops and searches without cause. *University of Pennsylvania Journal of Constitutional Law* **3**, 296–366 (2001).

[27] of Constabulary, H. M. I. *Stop and search powers: Are the police using them effectively and fairly?* (Her Majesty's Inspectorate of Constabulary, London, 2013).

[28] Phillips, C. & Bowling, B. *Racism, ethnicity, crime and criminal justice*, 579–619 (Oxford University Press, Oxford, 2007), 4 edn.

[29] Smith, S. J. *Crime, space and society* (Cambridge University Press, Cambridge, 1986).

[30] Elliott, D. S. *et al.* *Lies, damn lies, and arrest statistics* (Center for the Study and Prevention of Violence, Boulder, CO, 1995).

[31] Fagan, J. & Davies, G. Street stops and broken windows: Terry, race, and disorder in New York City. *Fordham Urban Law Journal* **28**, 457–504 (2000).

[32] Kane, R. J. Social control in the metropolis: A community-level examination of the minority group-threat hypothesis. *Justice Quarterly* **20**, 265–295 (2003).

[33] Lammy, D. *The Lammy review: An independent review into the treatment of, and outcomes for, Black, Asian and Minority Ethnic individuals in the Criminal Justice System* (Ministry of Justice, London, 2017).

[34] Gounev, P. & Bezlov, T. The Roma in Bulgaria's criminal justice system: From ethnic profiling to imprisonment. *Critical Criminology* **14**, 313–338 (2006).

[35] Richardson, R., Schultz, J. & Crawford, K. Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *New York University Law Review Online* (2019). Forthcoming.

[36] Mawby, R. C. & Wright, A. The police organisation. In Newburn, T. (ed.) *Handbook of policing*, 224–228 (Taylor & Francis, Abingdon, 2008), 2 edn.

[37] Oberfield, Z. W. Socialization and self-selection: How police officers develop their views about using force. *Administration & Society* **44**, 702–730 (2012).

[38] Smith, D. J. & Gray, J. *Police and people in London: The PSI report* (Gower Publishing, London, 1985).

[39] Waddington, P. A. Police (canteen) sub-culture. an appreciation. *The British Journal of Criminology* **39**, 287–309 (1999).

[40] Eberhardt, J. L., Goff, P. A., Purdie, V. J. & Davies, P. G. Seeing black: Race, crime, and visual processing. *Journal of Personality and Social Psychology* **87**, 876 (2004).

[41] Alpert, G. P., MacDonald, J. M. & Dunham, R. G. Police suspicion and discretionary decision making during citizen stops. *Criminology* **43**, 407–434 (2005).

[42] Warren, P., Tomaskovic-Devey, D., Smith, W., Zingraff, M. & Mason, M. Driving while black: Bias processes and racial disparity in police stops. *Criminology* **44**, 709–738 (2006).

[43] Correll, J. *et al.* Across the thin blue line: police officers and racial bias in the decision to shoot. *Journal of Personality and Social Psychology* **92**, 1006 (2007).

[44] Morris, W., Burden, A. & Weekes, A. *The Report of the Morris Inquiry–The case for change: People in the Metropolitan Police Service* (Metropolitan Police Authority, London, 2004).

[45] American Civil Liberties Union (ACLU). *The Persistence of Racial and Ethnic Profiling in the United States* (Rights Working Group, New York, 2009).

[46] Adebowale, V. *Independent Commission on Mental Health and Policing Report* (Independent Commission on Mental Health and Policing, London, 2013).

[47] Quinton, P. Race disproportionality and officer decision-making. In Delsol, R. & Shiner, M. (eds.) *Stop and Search: The Anatomy of a Police Power*, 57–78 (Palgrave Macmillan, Basingstoke, 2015).

[48] Lea, J. The Macpherson report and the question of institutional racism. *The howard journal of criminal justice* **39**, 219–233 (2000).

[49] Reiner, R. *The politics of the police* (Oxford University Press, Oxford, 2010).

[50] Shiner, M. Regulation and reform. In Delsol, R. & Shiner, M. (eds.) *Stop and Search*, 146–169 (Palgrave Macmillan, Basingstoke, 2015).

[51] Shiner, M. & Thornbury, P. *An Evaluation of the Northamptonshire Police Reasonable Grounds Panel: Regulating Police Stop and Search* (Open Sciety Justice Initiative, London, 2019).

[52] Sekhon, N. Dangerous warrants. *Washington Law Review* **93**, 967–1017 (2018).

[53] Ross, C. T. A multi-level Bayesian analysis of racial bias in police shootings at the county-level in the United States, 2011–2014. *PloS One* **10**, e0141854 (2015).

[54] Edwards, F., Esposito, M. H. & Lee, H. Risk of police-involved death by race/ethnicity and place, United States, 2012–2018. *American Journal of Public Health* **108**, 1241–1248 (2018).

[55] Cesario, J., Johnson, D. J. & Terrill, W. Is there evidence of racial disparity in police use of deadly force? Analyses of officer-involved fatal shootings in 2015–2016. *Social Psychological and Personality Science* **10**, 586–595 (2019).

[56] Johnson, D. J., Tress, T., Burkel, N., Taylor, C. & Cesario, J. Officer characteristics and racial disparities in fatal officer-involved shootings. *Proceedings of the National Academy of Sciences* **116**, 15877–15882 (2019).

[57] Fryer Jr, R. G. An empirical analysis of racial differences in police use of force. Tech. Rep., National Bureau of Economic Research (2016). `http://www.nber.org/papers/w22399`.

[58] Ross, C. T., Winterhalder, B. & McElreath, R. Resolution of apparent paradoxes in the race-specific frequency of use-of-force by police. *Palgrave Communications* **4**, 61 (2018).

[59] Simpson, E. H. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society: Series B (Methodological)* **13**, 238–241 (1951).

[60] Neil, R. & Winship, C. Methodological challenges and opportunities in testing for racial discrimination in policing. *Annual Review of Criminology* **2**, 73–98 (2018).

[61] Gelman, A., Fagan, J. & Kiss, A. An analysis of the New York City Police department's 'stop-and-frisk' policy in the context of claims of racial bias. *Journal of the American Statistical Association* **102**, 813–823 (2007).

[62] Ridgeway, G. Analysis of racial disparities in the New York Police Department's stop, question, and frisk practices. Tech. Rep., Rand Corporation, Santa Monica (2007). `https://www.rand.org/pubs/technical_reports/TR534.html`.

[63] Pierson, E. *et al.* A large-scale analysis of racial disparities in police stops across the united states. *Nature Human Behaviour* (2020).

[64] Alpert, G. P., Smith, M. R. & Dunham, R. G. Toward a better benchmark: Assessing the utility of not-at-fault traffic crash data in racial profiling research. *Justice Research and Policy* **6**, 43–69 (2004).

[65] Lamberth, J. Revised statistical analysis of the incidence of police stops and arrests of black drivers/travelers on the new jersey turnpike between exits or interchanges 1 and 3 from the years 1988 through 1991. Report by Lamberth Consulting (1994).

[66] Lamberth, J. C. Racial profiling data analysis study: Final report for the San Antonio Police Department. Report by Lamberth Consulting (2005).

[67] Rojek, J., Rosenfeld, R. & Decker, S. Policing race: The racial stratification of searches in police traffic stops. *Criminology* **50**, 993–1024 (2012).

[68] Withrow, B. L. & Williams, H. Proposing a benchmark based on vehicle collision data in racial profiling research. *Criminal Justice Review* **40**, 449–469 (2015).

[69] Ridgeway, G. & MacDonald, J. M. Doubly robust internal benchmarking and false discovery rates for detecting racial bias in police stops. *Journal of the American Statistical Association* **104**, 661–668 (2009).

[70] Office for National Statistics. Census geography—an overview of the various geographies used in the production of statistics collected via the UK census (2016). `https://web.archive.org/web/20190715091529/https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography`.

[71] Simoiu, C., Corbett-Davies, S. & Goel, S. The problem of infra-marginality in outcome tests for discrimination. *The Annals of Applied Statistics* **11**, 1193–1216 (2017).

[72] Knox, D., Lowe, W. & Mummolo, J. The bias is built in: How administrative records mask racially biased policing. *American Political Science Review* (2020). Forthcoming.

[73] Kruschke, J. K. & Liddell, T. M. The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review* **25**, 178–206 (2018).

[74] Mastrofski, S. D., Parks, R. B. & Worden, R. E. *Community policing in action: Lessons from an observational study* (US Department of Justice, Washington, DC, 1998).

[75] Engel, R. S., Sobol, J. J. & Worden, R. E. Further exploration of the demeanor hypothesis: The interaction effects of suspects' characteristics and demeanor on police behavior. *Justice Quarterly* **17**, 235–258 (2000).

[76] Terrill, W. *Police coercion: Application of the force continuum* (LFB Scholarly Publishing, New York, 2001).

[77] Antonovics, K. & Knight, B. G. A new look at racial profiling: Evidence from the Boston Police Department. *The Review of Economics and Statistics* **91**, 163–177 (2009).

[78] McCrary, J. The effect of court-ordered hiring quotas on the composition and quality of police. *American Economic Review* **97**, 318–353 (2007).

[79] Legewie, J. & Fagan, J. Group threat, police officer diversity and the deadly use of police force. Tech. Rep. 14-512, Columbia Public Law Research Paper (2016).

[80] Malleson, N. & Andresen, M. A. The impact of using social media data in crime rate calculations: Shifting hot spots and changing spatial patterns. *Cartography and Geographic Information Science* **42**, 112–121 (2015).

[81] Miller, J., Le Masurier, P. & Wicks, J. *Profiling populations available for stops and searches* (Research, Development and Statistics Directorate, Home Office, London, 2000).

[82] Waddington, P. A., Stenson, K. & Don, D. In proportion: Race, and police stop and search. *British Journal of Criminology* **44**, 889–914 (2004).

[83] Equality and Human Rights Commission. *Stop and think: A critical review of the use of stop and search powers in England and Wales* (EHRC, London, 2010).

[84] Shiner, M. & Delsol, R. The politics of the powers. In Delsol, R. & Shiner, M. (eds.) *Stop and Search: The Anatomy of a Police Power*, 31–56 (Palgrave Macmillan, Basingstoke, 2015).

[85] Whitfield, J. *Unhappy dialogue. The Metropolitan Police and Black Londoners in post-war Britain* (Willan Publishing, Cullompton, 2004).

[86] Home Office. Table A.01c: Number of persons arrested by ethnic group year ending 31 march 2016. `https://www.gov.uk/government/statistics/police-powers-and-procedures-england-and-wales-year-ending-31-march-2016`.

[87] Office for National Statistics. 2011 census: QS201EW ethnic group: countries. `https://www.nomisweb.co.uk/census/2011/qs211ew`.

[88] Carpenter, B. *et al.* Stan: A probabilistic programming language. *Journal of Statistical Software* **76**, 1–32 (2017).

[89] Stan Development Team. RStan: the R interface to Stan (2020). URL `http://mc-stan.org/`. R package version 2.19.3.

[90] Office for National Statistics. *Ethnic group statistics–A guide for the collection and classification of ethnicity data* (Office for National Statistics, London, 2003).

[91] Bowsher, K. The code systems used within the Metropolitan Police Service (MPS) to formally record ethnicity. Tech. Rep. 03/07, Metropolitan Police Service (2007).

[92] Office for National Statistics. *Final recommended questions for the 2011 Census in England and Wales–National Identity* (Office for National Statistics, London, 2009).

Acknowledgements

Support for this research was provided by a grant from the Leverhulme Trust (L.V.), the West Midlands Police and Crime Commissioner and Economic and Social Research Council grants ES/P008976/1 and ES/N018192/1 (N.S.). Special thanks goes to Jeremias Knoblauch for invaluable discussions. We also thank Philip Newall, Jenny Richards, Jennifer Trueblood, Gillian Wall, Dave Whordley, and the West Midlands Police Stop and Search Commission for comments.

Author contributions

L.V. analysed data, designed and executed the research. L.V. and N.S. wrote the paper.

Competing interests

The authors declare no competing interests.

Appendix A Additional results

In this section we provide some additional results which do not currently have a place in the main text but may be of interest to the reader.

A.1 provides the coefficients from Equation (2) in table form.

A.2 shows Figure 1 and Figure 3 disaggregated by time.

A.3 shows the model fit of $p = \text{Softmax}(\theta_{it})_e$ by comparing predicted search counts based on p to observed search counts.

A.4 presents our analysis on the stability of D^S and D^P over time.

A.5 includes an area-specific plot of stop and search activity relative to crime against the non-white population share.

A.1 Coefficients

The summarised posterior distributions of the coefficients in the estimation of θ are already visualised in Figure 2. Additionally, they are provided in table format in Table A.1.

A.2 Time disaggregation

As mentioned in the main part of the paper, there are essentially no time dynamics across the nine 6-month periods in our study. We therefore aggregated our results in Figure 1

Table A.1: Estimates and 90% uncertainty intervals (UI) for model parameters in Equation (2). The estimates are also displayed graphically in Figure 2. Officer age and experience are standardised.

	Asian		Black	
Parameter	Median	90% UI	Median	90% UI
Global intercept	-0.11	[-1.30, 1.12]	-0.24	[-1.46, 0.96]
Female officer	-0.15	[-0.20, -0.11]	-0.10	[-0.16, -0.04]
Officer age	0.02	[-0.00, 0.04]	0.01	[-0.01, 0.04]
Officer experience	-0.03	[-0.05, -0.01]	-0.06	[-0.09, -0.04]
Officer of same ethnicity	-0.03	[-0.08, 0.03]	0.02	[-0.20, 0.24]
White share in team	-1.73	[-3.09, -0.44]	-2.10	[-3.42, -0.78]
Officer-level suspect share	1.94	[1.53, 2.31]	-1.18	[-1.64, -0.72]
Officer-level patrol share	1.89	[1.59, 2.21]	9.35	[8.69, 10.01]
SD of team-specific intercept (σ_α)	0.35	[0.27, 0.46]	0.43	[0.34, 0.57]

and Figure 3 by time. For the interested reader we provide the disaggregated results in Figure A.1 and Figure A.2.

A.3 Model fit

Next, we comment on the model fit visualised in Figure A.3.

A.4 AR(1) model

We comment on the autocorrelation or serial correlation of D^S and D^P over time. If an officer exhibited a similar degree of bias against an ethnic group at all times, then we would observe a high degree of autocorrelation. Equally, if an officer's bias at a previous period does not give us any information about the officer's bias now then the bias is not stable and we would observe no autocorrelation.

We infer the officers' degree of autocorrelation using a autoregressive time series model of order 1, an AR(1) model. Briefly, an AR(1) model is a linear model that predicts the value of the time series at time t using the previous value of the series at $t-1$. The coefficient b_{ie} on the previous value gives us the degree of autocorrelation. For each officer and ethnicity we infer a separate coefficient such that we account for different degrees of autocorrelation between ethnic groups within the same officer.

In a second step, we model the stability of our disparity measures D because it allows us to draw conclusions about the stability of officer bias. For each officer we obtain a posterior distribution over D_{ite} which is officer i's log disparity of searching ethnicity e in time period t relative to ethnicity e's prevalence in the officer's baseline. We then fit an autoregressive time series process of order 1, an AR(1) process to the time series

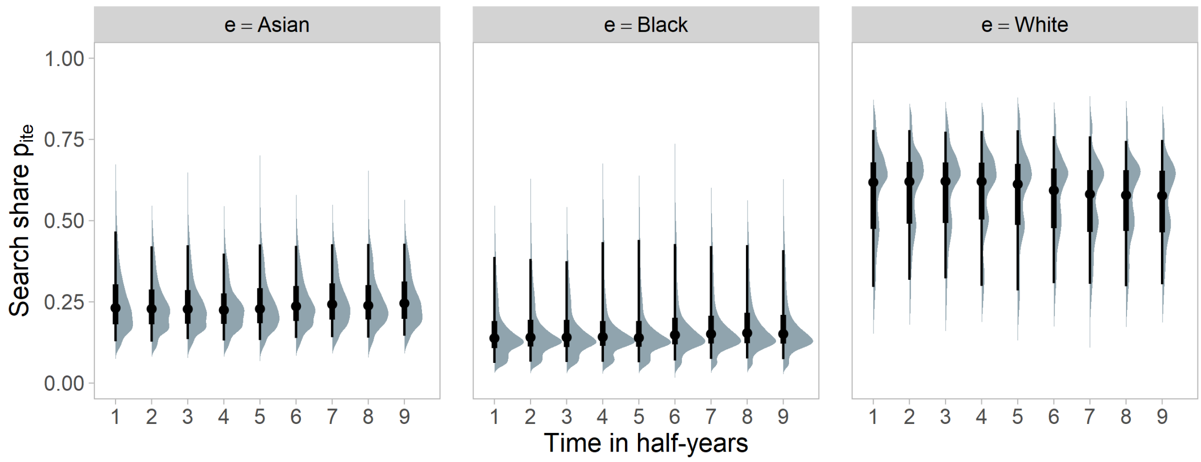

Figure A.1: Densities of posterior distributions of search shares over all officers for each 6-month time period. Each grey pixel represents one sample of the posterior distributions of search share p_{ite} for each officer i, ethnic group e and time t. The black dot represents the median of the distributions aggregated over officers and the black lines show 50% and 90% uncertainty intervals.

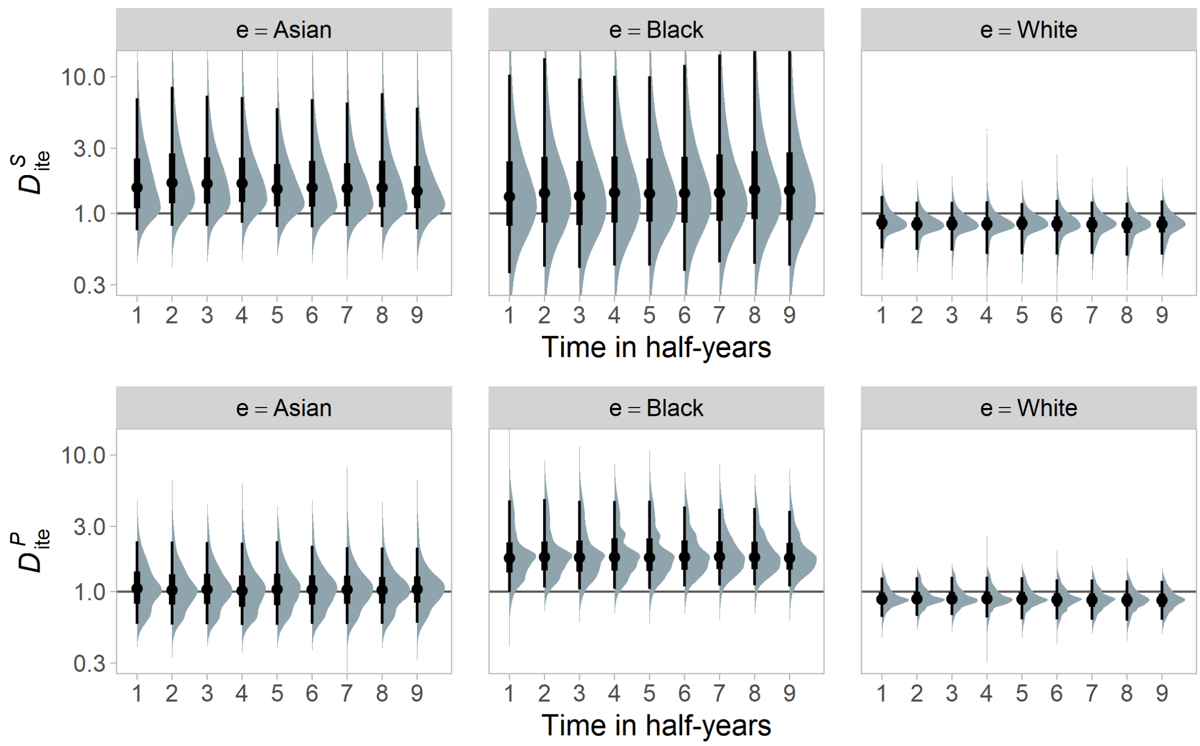

Figure A.2: Posterior distributions of D^S and D^P. For visual clarity, we only show values between $[-5, 7.5]$ which includes the minimum of all distributions. Note that the y-axis is on the log scale. 1.8% of all posterior draws are excluded by this choice. Each grey pixel represents a posterior draw. The black dots represent the medians; the black lines represent 50% and 90% uncertainty intervals.

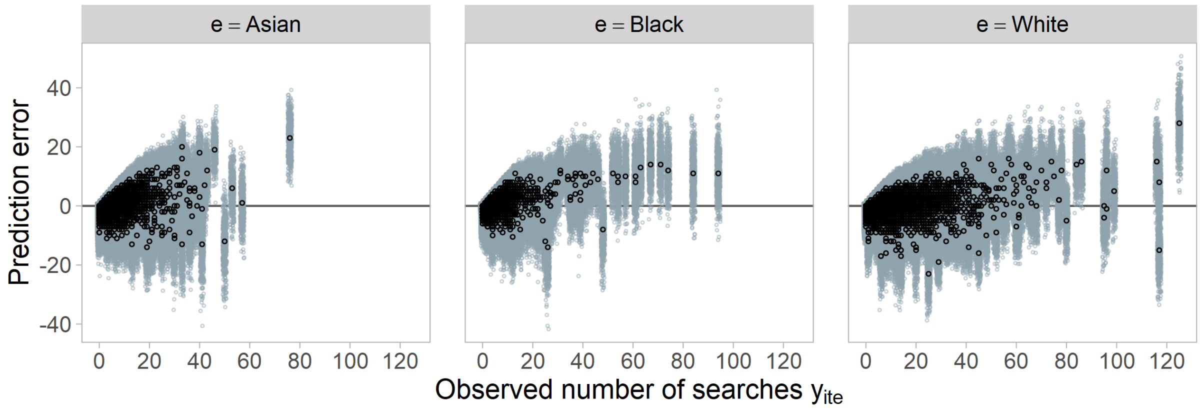

Figure A.3: Comparison of observed search counts to predicted search counts based on inferred search shares p. Each grey dot is the observed search count by an officer in time period t and ethnic group e against the prediction error (observed − predicted). The black dots are show the observed data against the error from the median prediction for that observation. The plot shows that key features of the data are captured in the model.

of summarised D_{ite} over t. Since we cannot fit a time series to every single posterior draw we instead fit the time series at three summary points of the posterior distributions of D^S_{ite} and D^P_{ite}: The median and the lower and upper 90% uncertainty intervals. We consider two measures, D^S and D^P and we fit AR(1) models to both time series at three summary points. This results in 1,194 officers × 3 ethnic groups × 2 disparity measures × 3 summary points = 14,316 AR(1) coefficients. For ease of notation, we describe our model with respect to a generic disparity measure D:

$$\begin{aligned} D_{ite} &= a_{ie} + b_{ie} D_{i(t-1)e} + \varepsilon_{ite}, \qquad t > 1 \\ \varepsilon_{ite} &\sim N(0, \sigma_{D_{ie}}) \end{aligned} \tag{A.1}$$

where b_{ie} gives the degree of autocorrelation. If $b_{ie} > 0$, i.e., the autocorrelation is positive, then the D_{ite} move in the same direction over time. Negative autocorrelation indicates that the terms move in opposite directions over time. If b_{ie} is zero then the process is driven entirely by a_{ie} and the error term.

Our time series is very short with only nine time periods. Additionally, some officers are not observed in the entire study period so we have even fewer observations for these officers. Altogether, the data sparsity makes the estimation of the officer-specific terms a_{ie}, b_{ie} and $\sigma_{D_{ie}}$ challenging. We therefore introduce a hierarchical prior structure where all officer- and ethnicity-specific intercept and slope a and b have an ethnicity-specific hyper-prior A_e or B_e on their mean. Additionally, the standard deviation $\sigma_{D_{ie}}$ also has

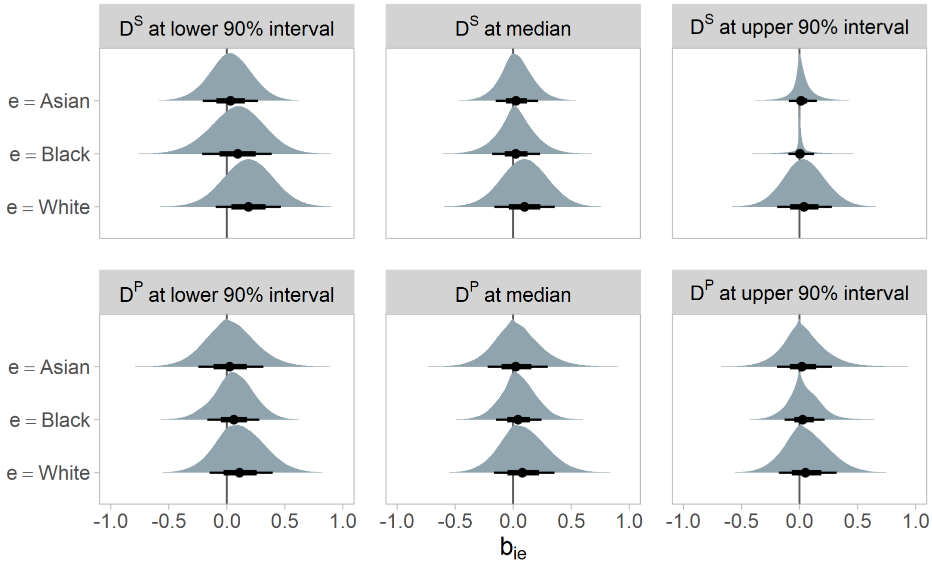

Figure A.4: Densities of AR(1) coefficients for time series of D^S and D^P at different summary points of the posterior distributions of D^S_{ite} and D^P_{ite}.

an ethnicity-specific hyper-prior σ_e. More formally:

$$\begin{aligned} a_{ie} &\sim N(A_e, 0.25) & A_e &\sim N(0,1) \\ b_{ie} &\sim N(B_e, 0.25) & B_e &\sim N(0, 0.25) \\ \sigma_{D_{ie}} &\sim N^+(0, \sigma_e) & \sigma_e &\sim N^+(0,1), \end{aligned}$$

where N^+ denotes a half-normal distribution.

Unfortunately, our time series of only nine half-years is too short to allow strong conclusions about the stability of bias. Figure A.4 shows the distributions of AR(1) coefficients aggregated by officers at the different summary points. The 90% uncertainty intervals around the AR(1) coefficients are simply too wide: the average range between the lower end of the 90% UI and the upper end of the 90% is 0.66 which is considerable given that the coefficient lies between -1 and 1.

Evidence of autocorrelation is particularly weak for D^S, where the 90% UI for only 56 of the 3579 coefficients excluded zero. In contrast, 1,355 out of 3579 90% UI for the coefficients estimating the stability of D^P excluded zero. All coefficients are positive. In other words, we have weak evidence that approximately one third of officers are consistent in their search bias against ethnic groups relative to their patrol baseline. We cannot comment on the strength of this consistency because of the aforementioned poor estimation of the coefficients. We simply do not have enough data to make conclusive statements about the stability of bias for the majority of officers.

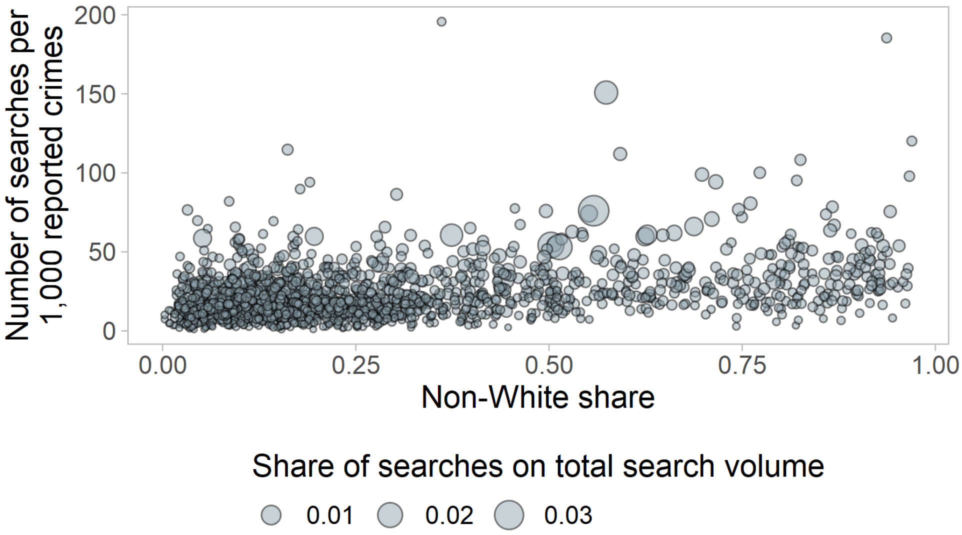

Figure A.5: Scatterplot of search activity relative to crime against the non-white population share for 1,666 Lower layer Super Output Areas (LSOA). Dot size indicates how much the searches in this area contribute to the overall search volume by West Midlands police. This figure shows that while some areas do see higher volumes of stop and search relative to crime incidence, these searches do not disproportionately contribute to overall searches.

A.5 Stop and search activity

A recent evaluation of stop and search behaviour in England noted that some officer teams focused their attention to specific areas "with a large proportion of minority residents"[51]. Given our findings of over-patrolling a natural question arises: Are there areas—ethnically diverse areas in particular—which are subject to high numbers of stop and search that cannot be ed by the incidence of crime?

We note here that this is a purely descriptive figure. The very presence of police in an area can drive the observation of crime. Therefore, the above result does *not* mean that police do not over-search minority areas after adjusting for crime.

Appendix B Supplementary material

B.1 Ethnic classification

In our analysis, we compare stop and searches between ethnic groups. During any interaction with police, individuals are asked to define their ethnicity into five broad categories: White, Mixed, Asian/Asian British, Black/Black British and Other. The White category encompasses encompasses British White, Irish and any other White background; the Asian/Asian British category encompasses Indian, Pakistani, Bangladeshi and any other Asian background and the Black/Black British category encompasses Caribbean, African and any other Black background. This classification system used by the police is based on the ONS 2001 Census[90, 91].

In the ONS 2011 Census, the Office for National Statistics changed the classification system to include Chinese people in the Asian/Asian British category rather than in the Other code as they did in 2001[92]. To harmonise the ONS and the police's classification system, we follow the police's classification and exclude Chinese people from the census counts of Asian people.

B.2 Sample selection

The final dataset is compiled from four data files provided by West Midlands Police: `crimes`, `incidents`, `searches` and `officers`. The `searches` data are based on search forms which each officer has to fill out at the time of search. The searched person is then provided with a receipt and serial number of this record. We assign search decisions to all officers who jointly made the decision on patrol together, independently of who logged the search.

We link officers to crime suspects using the `incidents` and `crimes` data. Officers attend incidents throughout their work day. Some of these incidents will be logged as crimes and the `crimes` data holds information on the person suspected of having committed the crime. If an incident with officers A and B present is logged as a crime with suspect C present, we say that both officers A and B interacted with suspect C. We cannot ascertain whether suspect C was identified at the time of the crime incident or later on following an investigation. We exclude all crimes with more than five years between the crime incident and the crime report since it is unlikely that the officers encountered C as part of their investigation. All exclusions and matches between the data files are reported in Table B.1. Our reliance on the `incidents` data to match officers to crime cases means that our final data does not contain any crimes which were reported at police stations. Our analysis also excludes all crimes which were recorded as a consequence of stop and search. This means that our measure of the criminal population is not confounded by the process of stop and search.

Dataset	Number of cases
`incidents` and `crimes`	
total incidents between 01/04/2014–30/09/2018	2,315,348
resulting in crime report	598,837
with any crime suspect information	341,297
with Asian, Black or White suspect	313,365
excluding old cases	312,651
by qualifying officers	203,176
`searches`	
total stops and searches between 01/04/2014–30/09/2018	62,804
with stopped person's ethnicity	59,739
with Asian, Black or White stopped person	56,021
requiring reasonable grounds of suspicion	55,740
by qualifying officers officers	36,028
`officers`	
total active police officers in West Midlands	5,081
were active in the police force in at least 5 our of 9 half-years	3,916
performed searches	1,194

Table B.1: Description of matching and exclusion criteria applied to `incidents`, `crimes`, `stops` and `officers`. Indented conditions are chained: the last row of this table are all officers who were active in at least 5 out of 9 half-years AND performed at searches in that time.