Managerial Discretion and Racial Disparities: Evidence from Military Drug Screening

Kyle Greenberg* Carson Homme† Anika Nerella‡
CarlyWill Sloan§

March 20, 2025

Abstract

We analyze racial disparities in U.S. Army commanders' decisions to order discretionary drug tests on subordinate servicemembers, using unannounced random drug tests as a benchmark for actual drug use. Black soldiers are nearly twice as likely as non-Black soldiers to be discretionarily tested for drugs even when measured drug use is the same for both groups. This disparity persists consistently across units and cannot be explained by differences in drug use patterns. Applying Empirical Bayes methods, we find substantial heterogeneity in discretionary testing practices across military commands: after accounting for statistical noise, brigades in the 90th percentile of the racial testing gap overtest Black soldiers at rates two to three times higher than brigades in the 10th percentile. These findings demonstrate how managerial discretion perpetuates racial disparities even within highly structured institutions with formal oversight mechanisms.

JEL J15, J45, M50

Disclaimer: The views expressed in this paper are those of the authors and do not necessarily represent the views, official policy, or position of the United States Military Academy, the Department of the Army, the Department of Defense, or the U.S. Government. Please do not cite or distribute without the authors' permission.

*United States Military Academy at West Point. Email: kyle.greenberg@westpoint.edu.
†United States Military Academy at West Point. Email: carson.homme@westpoint.edu.
‡University of Pennsylvania. Email: anika.nerella@upenn.edu.
§United States Military Academy at West Point. Email: carlywill.sloan@westpoint.edu.

1 Introduction

While a large literature documents discrimination in hiring and promotion decisions, we know far less about how managers may perpetuate discrimination through their day-to-day discretionary choices over employees. This gap is particularly important as punitive workplace policies like drug testing have become increasingly prevalent. According to the National Survey on Drug Use and Health, 41% of U.S. employers have punitive drug testing policies (Le et al. 2024).

Measuring discrimination in managerial decisions is empirically challenging. While studies routinely document disparate outcomes across racial groups, quantifying discrimination often requires strong assumptions about employees' underlying behavior that researchers cannot observe. For example, in hiring or promotion decisions, managers may claim to base their choices on merit, but researchers lack direct visibility into all relevant performance metrics or informal assessments that shape these decisions. In criminal justice contexts, for example, studies of racial bias in police stops or searches must typically make assumptions about baseline rates of criminal behavior, since researchers cannot directly observe who is actually carrying contraband or committing crimes in the broader population.

The U.S. Army's drug testing program offers a unique setting to study how managers exercise discretion and potentially perpetuate discrimination. Servicemembers are required to undergo frequent and unannounced random drug tests. However, if military commanders suspect that a subordinate is using drugs, they can order the servicemember to undergo additional drug tests outside of the random testing regimen. Between 2015-2022, battalion commanders and their subordinate company commanders - who serve as senior managers responsible for executing drug testing programs in battalions that have between 300 to 1000 personnel - tested over 170,000 junior enlisted soldiers, administering nearly 420,000 random tests and 2,500 discretionary tests. The setting provides three key advantages. First, we focus our analysis on junior enlisted soldiers who have little to no ability to select their units or commanders. Second, we have detailed

data on both commanders' discretionary choices and their subordinates' underlying behavior measured through random tests. Third, testing decisions have significant career consequences–soldiers who test positive for drugs without a medical excusal are charged with a crime and separated from the military.

We estimate the causal relationship between soldier race and discretionary testing by leveraging the Army's random drug testing program, which provides an objective benchmark of underlying drug use. Our empirical strategy compares discretionary testing rates between Black and non-Black soldiers while controlling for random test failure rates at the race-by-unit level. Importantly, we show that random testing within our sample is truly random - uncorrelated with soldier characteristics after controlling for basic military characteristics that might influence whether a soldier is available to be randomly tested.

Our key results show that, on average, Black soldiers are nearly twice as likely as non-Black soldiers to face discretionary drug testing. This disparity persists even after accounting for underlying drug use as measured by random tests. Differences in discretionary testing rates remain statistically significant and nearly identical in magnitude after controlling for soldier characteristics and unit-level factors, further underscoring that racial disparities stem from commanders' decision-making rather than any actual differences in drug use not captured by random drug test failure rates among soldiers of the same race who are assigned to the same military unit.

We conduct several empirical checks to assess whether racial disparities in discretionary drug testing reflect factors other than managerial decisions, including nonrandom assignment of soldiers to battalions, the potential noisiness of random drug tests as a measure of underlying drug use, and the possibility that commanders have hidden information on soldier drug use. First, to address the possibility that Black soldiers may be assigned to battalions with a higher discretionary testing propensity, we include battalion commander spell fixed effects and find that racial disparities persist even within the same battalion. Second, we examine battalions with comparable numbers of Black and

non-Black soldiers, finding similar disparities, which suggests that our results are not easily explained by measurement error in drug usage rates estimated from random drug tests. Third, even when restricting the sample to larger battalions to reduce measurement error, Black soldiers remain over-tested. Finally, by comparing racial disparities in battalions with similar drug failure rates among Black and non-Black soldiers, we find that disparities persist, suggesting commanders are not merely responding to unbiased drug use information.

Our battalion-level estimates reveal that, on average, Black soldiers face significantly higher discretionary drug testing rates than non-Black soldiers in the same military units, even after accounting for baseline drug use and individual characteristics. However, this aggregate disparity masks substantial variation across commanders—some commanders show little to no evidence of racial disparity, while others systematically over-test Black soldiers at much higher rates. While these battalion-level estimates provide critical insights, they may also be subject to statistical noise, particularly for commanders with fewer observations.

To better understand how racial disparities in discretionary testing operate at higher levels of command, we aggregate battalion-level estimates to the brigade level (brigades consist of multiple battalions) and apply Empirical Bayes (EB) methods. This approach refines our estimates by borrowing strength across units, allowing us to distinguish true heterogeneity from random variation. The results reveal substantial heterogeneity in racial discretionary testing gaps across brigades. By applying EB shrinkage techniques, we estimate that 31% to 40% of the observed variance in brigade-level disparities is due to genuine heterogeneity rather than statistical noise. Even after accounting for estimation uncertainty, disparities in discretionary testing remain significant: brigades in the 90th percentile of racial testing gaps exhibit overtesting of Black soldiers at rates 130% to 200% higher than brigades in the 10th percentile. These findings suggest that while noise contributes to some variation in observed disparities, systematic differences in discretionary testing practices persist across commands, reinforcing the role of managerial discretion in

shaping racial disparities.

Our findings suggest that while Army regulations and oversight mechanisms are designed to minimize subjectivity in discretionary decisions, they may be insufficient to prevent individual managers from making discriminatory choices. For example, Army drug testing regulations require commanders to establish probable cause before administering discretionary tests. Yet, 73% of these tests yield negative results, suggesting that commanders exercise considerable discretion——potentially in ways that undermine the intended safeguards against bias. This pattern reflects broader research on persistent discrimination in other rule-based environments, such as policing and criminal sentencing, highlighting the limitations of structural safeguards and the need for interventions that directly address individual decision-making.

Our study contributes to several literatures in labor economics and the economics of discrimination that relate to how managers exercise discretion in organizations, racial bias in rules-based institutions, and the economic impacts of workplace drug testing policies.

We expand on the growing literature about managerial discretion and discrimination within organizations, addressing a critical gap in understanding how managers' day-to-day choices affect worker outcomes and perpetuate discrimination. While research has extensively documented biases in hiring and promotion, we know surprisingly little about how managerial discretion impacts minority performance directly through daily interactions. Prior studies in retail contexts, for example, reveal that managers' biases can create self-fulfilling prophecies that negatively affect minority workers (Glover, Pallais, and Pariente 2017), manifesting in hiring through taste-based and screening biases (Benson, Board, and Meyer-ter-Vehn 2024). However, these findings primarily reflect private-sector dynamics, where managers generally have more freedom in personnel decisions than in public-sector settings.

Recent research underscores the influence of managers in shaping productivity within the public sector, particularly through personnel decisions such as retirement (Fenizia 2022). However, little is known about how these discretionary choices impact racial eq-

uity within the constraints that public sector managers face in hiring, firing, and promotion. Addressing this knowledge gap is critical since, even in rule-based environments, managers can meaningfully impact outcomes through the exercise of their limited discretion. The challenge of studying managers' influence on organizational outcomes has historically stemmed from data limitations and identification issues (Bertrand and Schoar 2003; Bloom and Van Reenen 2007). Recent advancements allow researchers to explore settings with detailed personnel data and quasi-random manager-worker assignments (Bandiera et al. 2020; Bloom et al. 2013; Lazear, Shaw, and Stanton 2015; Baltrunaite, Bovini, and Mocetti 2023; Minni 2023). In public sector settings, studies correlate better organizational practices with improved outcomes in hospitals (Tsai et al. 2015), schools (Bloom et al. 2015), and civil service organizations (Rasul and Rogger 2018). Our study builds on this literature by examining military commanders—managers who oversee drug-testing programs and exercise considerable discretion even within a rule-based framework. Our setting offers unique advantages: we observe commanders making repeated, comparable decisions, allowing us to leverage random testing as a baseline to measure underlying patterns of behavior. This design enables us to credibly identify both manager effects and discrimination, contributing valuable insights to the broader discourse on managerial discretion in both the public and private sectors.

Second, we contribute to research on racial bias in rule-based institutions, particularly criminal justice settings. These settings are especially important to study because, as Lang and Kahn-Lang Spitzer 2020 note, taste-based discrimination may be more likely to persist where market forces of competition are weak or absent. Despite clear guidelines and oversight mechanisms, racial disparities persist in policing (Lang and Kahn-Lang Spitzer 2020; Owens and Ba 2021), and bail decisions (Arnold, Dobbie, and Yang 2018). The military setting shares important features with these contexts: decision-makers have discretion within formal rules, face oversight, and make high-stakes choices affecting individuals' careers and freedom. Building on recent methodological advances (Arnold, Dobbie, and Hull 2022; Goncalves and Mello 2021), we adapt techniques to assess dis-

crimination at the individual commander level. Our research design contrasts commanders' discretionary testing decisions against the baseline of random drug tests, enabling a precise measure of discrimination that isolates individual biases. This approach diverges from broader assessments of systemic bias by identifying that only a subset of commanders—approximately 42%—are responsible for the aggregate racial disparity observed. These findings suggest that targeted policy interventions could address the discriminatory practices of specific individuals rather than necessitating wholesale institutional reform.

Finally, we provide new evidence on workplace drug testing policies, which have become increasingly prevalent in the U.S. labor market. Workplace drug testing is quite common: 45% of employers conduct drug testing, and 37% have punitive policies, often leading to immediate termination after a first positive result (Le et al. 2024). Most existing analyses of workplace drug policies emphasize their effects primarily on hiring and initial employment outcomes. Studies indicate that drug testing policies often deter certain applicants, impact the hiring pool, and influence initial employment decisions (Wozniak 2015). By examining the military's extensive drug testing protocols, we document effects that extend beyond hiring decisions, shedding light on how ongoing drug testing policies affect retention, career progression, and broader workplace equity.

Our findings have implications that extend beyond drug testing into broader labor market and policy discussions. First, given that drug-related separations account for 19% of misconduct-based discharges in the military, discriminatory testing practices can significantly impact economic mobility for disadvantaged groups that benefit most from military service (Greenberg et al. 2022). Moreover, these same commanders make numerous other consequential decisions - from performance evaluations to duty assignments to disciplinary actions - suggesting their biases may contribute to racial disparities across multiple dimensions of military careers. Second, our results are particularly relevant for other public sector organizations, especially police departments, since roughly one in four police officers are military veterans - more than four times the rate of veterans among the

general U.S. population (Morin and Mercer 2017). Related research finds that military veteran status is associated with increased odds of shootings among law enforcement officers (Reingle Gonzalez et al. 2019). Thus, discriminatory practices learned or reinforced during military service may influence how authority is exercised in civilian law enforcement. Third, as workplace drug testing becomes increasingly common across sectors, our findings suggest that discretionary testing policies may enable discrimination even in organizations with strong oversight and formal rules. Finally, these results highlight how seemingly neutral workplace surveillance policies can have disparate impacts when implementation authority is delegated to managers, an increasingly important consideration as organizations adopt more sophisticated monitoring technologies.

2 Institutional Details

2.1 Drug Testing in the Army

By enforcing rigorous drug testing policies, the Army presents a structured environment in which we can investigate the fairness of disciplinary practices and explore possible disparities in treatment across demographic groups. The Department of Defense prohibits drug use for all employees (Army Regulation 600-85). Article 112a of the *Uniform Code of Military Justice* prohibits soldiers from using amphetamines, barbiturates, cocaine, ecstasy, opiates, heroin, phencyclidine (PCP), tetrahydrocannabinol (THC), oxycodone/oxymorphone, benzodiazepines, lysergic acid diethylamide (LSD), steroids, synthetic cannabis (Spice), and any related compounds. Prior to enlisting, all soldiers must pass a drug test and remain subject to ongoing urine drug tests throughout their military service.

Given the Army's stringent approach to drug use, drug testing becomes a pivotal, high-stakes event for soldiers, often carrying significant career consequences. Failed drug tests account for 73% of all drug crimes within the Army and often lead to separation. Army Regulation specifically mandates that commanders must initiate separation for ev-

ery soldier that fails a drug test (Army Unit Prevention Leader's Handbook). This process is costly, both to the Army and commanders, with drug testing failure-related separations accounting for 19% of all soldiers who are separated from the Army for misconduct.[1]

2.1.1 The Role of Battalion and Company Commanders in Army Drug Testing

The implementation and enforcement of drug testing policies within the Army rely heavily on the leadership roles of battalion and company commanders, who oversee testing and discipline at the unit level. Battalions are the smallest echelon of Army unit responsible for implementing a drug testing program (Army Regulation 600-85). Battalion commanders manage and oversee the implementation of drug testing policies in their units. These commanders are the first level of approval to separate soldiers on drug charges and adjudicate punishment to soldiers for these crimes under the Uniformed Code of Military Justice. Company commanders (the next echelon of unit size below) are responsible for the scheduling and execution of the drug testing within their units.

For our study, we focus on racial disparities in drug testing at the battalion level, as battalion commanders are responsible for implementing drug testing programs and imposing disciplinary actions under the Uniform Code of Military Justice. However, because company commanders typically oversee the execution of drug tests, our results reflect the collective behaviors of both battalion commanders and their subordinate company commanders. We also report complementary results at the company level, which show similar patterns (see Table A6). In addition to estimating effects at the subordinate unit level (company), we also aggregate and report effects at the next highest echelon of the military unit (a brigade) to capture broader patterns of heterogeneity in discretionary testing and to improve inference through empirical Bayes methods.[2]

1. Soldiers can also be barred from re-enlisting. However, we don't consider this type of separation here, as it is very uncommon for soldiers who fail drug tests.

2. Figure A1 within the appendix helps illustrate this hierarchy within Army units. Brigades are the largest military unit within our analysis, consisting of roughly 6-7 battalions and between 4000-6000 soldiers. Battalions roughly contain 5 companies, and each company contains between 100-150 soldiers.

2.1.2 Random and Unit-Wide Drug Tests.

Most of the drug screening that soldiers are subject to during their military service is non-discretionary. These tests take the form of commanders either testing a random fraction of service members in their unit (an "Inspection Random", or "IR" test), or commanders testing all soldiers in their formation (an Inspection Unit, or "IU", test). The purpose of these tests is to deter drug use. (Department of the Army 2016). Commanders must test at least 10% of their personnel monthly, and each soldier must be tested at least once a year (Army Unit Prevention Leader's Handbook).

We use the failure rates of these arguably random tests to measure drug use within battalions. In this study, we use all "Inspection Random" tests and only "Inspection Unit" tests from January and July. These months likely reflect full unit tests conducted after winter and summer leave periods where units have few scheduled activities and soldiers are encouraged to take leave. Our measure of drug usage does not include unit-wide tests from months other than January and July to avoid concerns that such tests may be influenced by commanders' suspicion of drug use. For simplicity, we collectively refer to random "IR" tests and unit-wide "IU" tests administered during the months of January and July as "random" drug tests going forward.

Soldiers selected for random drug tests must participate. Selection for these tests occurs through a computerized randomization process, which ensures that all soldiers have an equal probability of being chosen.[3] The frequency of these tests is mandated by Army regulations, with a requirement to conduct drug testing on a certain percentage of the unit population at regular intervals throughout the fiscal year. Unit commanders do not directly select individual soldiers but are responsible for overseeing the process. While the timing of these tests is designed to be unpredictable, soldiers may still attempt to anticipate the likelihood of a test based on past frequency or patterns within their unit. In addition to random individual testing, commanders may also implement unit-wide drug

3. If the computer program (DTP Lite software provided by the Army Substance Abuse Program) is unavailable, Army policy suggests alternative randomization methods, such as drawing index cards from a hat or using a 10-sided die.

tests, where all soldiers within a unit are tested on the same day. These unit-wide tests are typically used as a way to ensure comprehensive screening and may be scheduled following deployments or after major holidays.

Only the company commander may excuse a soldier from testing, with the understanding that they are still required to test that soldier as soon as he or she is available. Reasons for missing these drug tests can vary, such as soldiers being on vacation, being tasked with additional duties outside of their unit footprint, or being within an occupation that has different work hours than the rest of the company.[4] For example, cooks within the Army often work longer and very different hours due to the need to provide breakfast, lunch, and dinner. Their ability to get away from their job to participate in unit drug testing is constrained, and the commander may excuse them on the test date and reschedule for later if they were selected for testing. Since we cannot observe soldiers on vacation or any other unobservable characteristics that would influence a soldier's availability for testing, we focus our sample on those with the highest probability of being present for testing, often incorporating occupation and experience fixed effects. We will also show empirical evidence of conditional random testing in Table 2.

2.1.3 Discretionary Drug Testing

Aside from deterring drug use, commanders (both battalion and company) are also given the authority to drug test individual soldiers who they suspect to be using drugs (inspection codes "PO," "CO," and "VO"). These types of tests are rare, however, and there are barriers to using this type of testing. Commanders are required to have probable cause that an individual soldier is suspected of using drugs prior to administering a discretionary drug test. Although commanders have access to and are encouraged to seek advice from legal advisors on whether probable cause exists, this is ultimately the commander's decision. The fact that less than 30% of soldiers fail these tests indicates discretion

4. If a soldier is unable to test when selected, Army regulation mandates that commanders test this soldier at his or her next availability under inspection code "IO." Furthermore, any soldier not tested within the first three quarters of the fiscal year is required to be tested under this code at some point within the fourth quarter (Army Unit Prevention Leader's Handbook).

in how commanders interpret "probable cause."

2.2 How Soldiers and Commanders Are Assigned to Units

The Army's structured assignment of soldiers and commanders minimizes self-selection biases, such as drug-using soldiers choosing commanders who avoid testing, as neither can easily influence their placement. After completing basic training, the Army's Human Resources Command assigns soldiers to brigades, which consist of multiple battalions, based primarily on U.S. Army personnel requirements.[5] Once a soldier arrives to a brigade, the brigade's HR office (also known as the Brigade "Adjutant's Office" or the Brigade "S-1 Shop") is responsible for establishing a process that assigns each soldier to a specific battalion and company according to the soldier's occupation and rank in a manner that generally keeps all subordinate units at similar personnel strength (Army Regulation 614-200). In practice, battalion and company personnel strengths fluctuate daily and are often imperfect due to lags in the time it takes for personnel managers to update soldiers' records.

As a general rule, soldiers have little ability to influence the precise company they are assigned to (Lyle 2006). This is especially true among the junior enlisted soldiers in our analysis sample (Section 3 below describes how we construct our sample). Importantly, it is highly unlikely that the initial-entry soldiers in our sample would have prior knowledge of battalion or company drug testing procedures, much less the ability to sort into units based on such information.

Battalion commanders in the Army are typically selected from a pool of senior officers, often at the rank of lieutenant colonel, through a competitive board process based on their leadership experience, performance evaluations, and career trajectory.[6] Once selected,

5. See Section 2.1 of (Bruhn et al. 2024) and (Army Regulation 614-200) for additional details on how the Army assigns soldiers to brigades.

6. During this time, Battalion Commanders are selected through the Centralized Selection List (CSL) process. All officers within 16-18 years of commissioning are eligible and must either "OPT-IN" or defer their participation. Of those who OPT-IN, their files (photo, education, awards, and previous evaluations) are then evaluated among a committee of 5-12 officers of varying occupations. They evaluate each file

they are assigned to command a battalion, generally for 18 to 24 months. It is uncommon for officers to command a battalion more than once, as most move on to higher-level staff or brigade-level commands after their tenure. Brigade Commanders assign Company commanders, usually captains, to battalion commanders within their brigades. These company-level commanders typically serve for 12 to 18 months before rotating to new assignments or leadership roles.

3 Data and Summary Statistics

3.1 Drug Testing and Personnel Records

We combine Army drug testing data with administrative personnel records to test for racial disparities in discretionary drug testing. Drug testing data comes from the Army's Drug and Alcohol Testing Management Information System (DAMIS) and is only available to us from January 2015 through April 2022. Drug testing data reveals the date the drug test was administered, the type of the drug test, and the results of the test. As described in Section 2.1, the type of drug test allows us to determine if it is a random drug test, if it is ordered at the discretion of the commander, or if it fits in neither category. Drug test results indicate if the soldier tested positive for any banned substance and, if so, the specific substances for which the soldier tested positive.[7]

Army personnel records include individual characteristics of soldiers when they first enlist and a monthly panel of assignment information. Individual characteristics include gender, race, age, Armed Forces Qualification Test (AFQT) scores, educational attainment,

and take in an officer's preferences for an operational, training, installation, or recruiting assignment, and then produce a list of commanders for battalions that will need a new commander. This process happens annually, and they choose future commanders nearly a year before they assume command of the unit.

7. All collected specimens are tested for tetrahydrocannabinol (THC), cocaine, heroin, amphetamines, MDMA/ecstasy, opiates, oxycodone/oxymorphone, benzodiazepines, and synthetic cannabis (Spice)(Army Unit Prevention Leader's Handbook). Commanders can also request some specimens to be additionally tested for barbiturates, phencyclidine (PCP), lysergic acid diethylamide (LSD), and steroids. We cannot observe if a soldier tests negative for barbiturates, phencyclidine (PCP), lysergic acid diethylamide (LSD), and steroids, yet nearly all (99.8%) positive random drug tests are from drugs that are always tested.

military occupation, and indicators for whether the servicemember failed a drug test or had at least one medically disqualifying condition while applying to enlistment.[8] Servicemembers' assignment data include monthly snapshots that indicate, among other things, servicemembers' rank and marital status in a particular month, the Army base where the soldier is permanently stationed, and the specific unit (brigade, battalion, and company) that the servicemember is assigned to.

3.2 Sample Selection

To construct our sample, we start with a monthly panel of all enlisted Army soldiers in the Army between January 2015 and April 2022. Our goal is to identify cases where random drug testing serves as a proxy for underlying drug use. Table A1 provides a detailed description of all sample restrictions we make to achieve this goal, but we summarize three main restrictions here.

First, we limit the sample to soldiers who typically live in the military barracks and are thus generally available for random drug testing. This necessitates restricting the sample to soldiers who are on their first tour of enlistment and excluding soldiers who are married or with dependents, as such soldiers generally do not live in military barracks.[9] We also exclude observations after a criminal charge or positive urinalysis and restrict the sample to male soldiers due to different testing patterns for women.[10]

Second, we eliminate observations where battalion commanders (henceforth referred to as "commanders") served for less than 12 months, as shorter tenures are uncommon and naturally result in fewer random drug tests, which limit our ability to measure drug use. Finally, we remove soldiers and units in two groups: those with many soldiers de-

8. Enlistment is still possible for some soldiers who fail drug tests or medical exams with appropriate enlistment waivers.

9. More precisely, we exclude soldiers who have been in the Army for longer than three years, the modal initial enlistment term-length, and we exclude observations in the month of, or after, a soldier becomes married or has children. By regulation, junior enlisted soldiers without dependents are required to live in the barracks (United States Army 2018). Our sample also excludes commissioned officers (i.e., Lieutenants and Captains), who also typically live in private housing and away from the military barracks.

10. Female soldiers may be tested later if no female urinalysis observer is available.

ployed and those that are training or support units, as both lack a combat mission and drug testing in all these contexts is often irregular.

After these restrictions, our sample includes 2,530,976 soldier-months among 172,339 unique junior enlisted, unmarried, male soldiers assigned to operational units. Each soldier-month is linked to a battalion-commander spell, which lasts about two years on average. Battalion spells refer to specific periods under a commander, not a fixed group of soldiers.

3.3 Summary Statistics

To understand our sample, we begin with soldier-level summary statistics in Table A2. This table provides a snapshot of our sample compared to all enlisted males and first-term males. The key takeaway is that our analysis sample is not significantly different from columns (1) or (2). By design, soldiers in our sample are more likely to be in a brigade combat team, slightly younger, and less educated. Most soldiers are White males in their early twenties. In this paper, we focus on differences between Black and non-Black soldiers, as drug failure rates differ significantly for Black soldiers. Hispanic soldiers comprise about 20% of the non-Black group. The average Armed Forces Qualification Test (AFQT) score of 58 shows that soldiers score above the national average.

Battalion Commander characteristics are summarized in Table A3. The main takeaway is the lack of variation among commanders—most are White, married males with the rank of Lieutenant Colonel. Our final sample includes 782 commander spells.

Average testing rates and counts are shown in Table 1. The first key point from the table is that both Black and non-Black soldiers are subject to random and discretionary tests. Discretionary tests are relatively infrequent, which makes sense given that soldiers know drug use is prohibited, and these tests are not without cost to company commanders. Despite these factors, nearly 2,500 discretionary drug tests were administered during our study period. The failure rate of discretionary tests ("Discretionary Test Fails") is higher than that of random tests, suggesting commanders are more successful than

random guessing in identifying drug use. However, it's important to note that 74% of discretionary tests do not result in failure, indicating that commanders are not always certain about drug use when ordering these tests. This supports the use of discretionary tests as a meaningful outcome for evaluating discrimination, as these are not simply cases where commanders are sure a soldier is using drugs. Discretionary test rates also show that Black soldiers are more likely to be tested through this method compared to non-Black soldiers.

Table 1 shows that random tests are very common and administered according to Army policy, with roughly 417,000 random tests conducted during our sample period. This suggests that our measure of drug test failure is unlikely to suffer from extreme measurement error, though we conduct additional checks to confirm this in the analysis that follows. While this table does not yet account for all controls necessary to ensure truly random testing, baseline random testing rates across different soldier races appear similar. Overall, nearly 1% of random tests result in failure, consistent with the low levels of drug use and the Army's strict policy on drug use. Marijuana is the most common drug found in failures, followed by cocaine (see Table A4).

Black soldiers are more likely to fail random drug tests, indicating a higher rate of drug usage. This disparity is mainly due to marijuana test failures (see Table A4).[11] The disparity is unlikely due to differences in detection, as Moeller, Lee, and Kissack 2008 shows that marijuana and other drugs, including cocaine, are detectable in urine tests for similar durations, regardless of single or frequent use. Table A4 also supports our focus on Black and non-Black disparities, as Hispanic random failure and discretionary testing rates are very similar to those of White soldiers.

While it might seem intuitive that discretionary testing for Black soldiers should be double that of non-Black soldiers considering the wide disparity in drug usage between these groups, panel B of Table 1 reveals that vast discrepancies in discretionary drug test-

11. Hasin, Shmulewitz, and Sarvet 2019 compiles survey sources on drug usage and documents a Black-NonBlack disparity in marijuana usage. Montgomery, Dixon, and Mantey 2022 also provides a summary of recent studies on racial differences in cannabis use.

ing remain even among groups of Black and non-Black soldiers who have similar rates of drug use. In Panel B of Table 1, we present random test failure rates and discretionary testing rates by quintile of failure. The intuition here is that within quintiles of random failure rates, we can better compare discretionary testing for instances where Black and non-Black soldiers have similar drug usage patterns. Across each grouping of similar failure rates, discretionary testing rates are substantially higher for Black soldiers. Within each quintile, Black soldiers are roughly twice as likely to be discretionarily tested as non-Black soldiers even though drug usage rates, as measured through random drug tests, are similar. For example, among groups of Black and non-Black soldiers who never test positive in random drug tests, Black soldiers are 132% more likely to be discretionarily tested than non-Black soldiers. Similarly, within the third quintile, Black Soldiers are 84% more likely to be discretionarily tested than non-Black soldiers despite similar drug usage rates across race within the quintile This suggests potential bias in the decision-making process for discretionary tests.

The distribution of random test failures and discretionary testing rates for Black and non-Black soldiers is shown in Figure A2. These figures reveal that while some battalions have no random failures or drug testing, there is considerable variation in non-zero testing and failure rates both within and across soldier races.

4 Conditional Random Drug Testing

Having outlined the key summary statistics, we now turn to the question of whether random drug test failure rates can be interpreted as reliable estimates of soldiers' underlying drug usage propensity. We begin by showing in Table 2 that soldier characteristics are not correlated with random testing. Each coefficient in the table comes from a separate regression of a soldier's characteristic, measured at the time the soldier enlists in the Army, on the number of random drug tests the soldier receives in a specific month, which is typ-

ically 0 (84.5% of observations) or 1 (14.5% of observations).[12] Specifically, we estimate the following equation:

$$X_i = \alpha + \beta_1 RandomTest_{it} + \tau_{bk(it)} + \pi_t + \epsilon_i \quad (1)$$

The coefficient of interest, β_1, reported in Table 2, represents the average difference in a specific soldier characteristic (X_i) between soldiers who are randomly drug tested and those who are not randomly drug tested. $\tau_{bk(it)}$ are fixed effects for every combination of battalion commander spell (b, which we henceforth refer to as simply battalions), the soldier's occupation, and the soldier's entry-year cohort (where k denotes the combination of occupation and cohort). Soldiers' occupation and entry-cohort rarely change, but soldiers can have different battalion commanders during their enlistment, which is why battalion-by-occupation-by-cohort fixed effects in Equation (1) are determined at the soldier-month level. In some specifications, we also include year-by-month fixed effects, π_t, to account for changes in testing patterns over time. We report heteroskedasticity-robust standard errors (White 1980).

Column 2 of Table 2 reports balance on a more parsimonious version of Equation (1) that regresses soldier characteristics on random drug tests while including only battalion fixed effects, thus omitting occupation and entry-cohort from $\tau_{bk(it)}$. This simple within-battalion comparison does indicate some slight imbalance, but the magnitudes are small. For instance, the coefficient on $BlackSoldier$ shows that a soldier who receives a random drug test is 0.189 percentage points less likely to be Black than a soldier who does not receive a random drug test. Relative to the share of Black soldiers in our sample (16.1%), this point estimate implies that Black soldiers are 1% less likely to be randomly drug tested than non-Black soldiers in the same battalion. However, column 3 reveals that these small imbalances become statistically and economically insignificant when fixed effects for every combination occupation by cohort are introduced. These fixed effects

12. The unit of observation is every combination of soldier-by-month. Soldier characteristics do not vary across months, but the number of random drug tests administered to a soldier each month does. Soldiers receive 2 or more random drug tests in the same month roughly 1% of the time.

are included based on institutional knowledge about drug testing: Soldiers in different military occupations vary significantly not only in demographics but also in roles, assignments, and schedules, which can impact their random drug testing rates. Column 4 shows that the random drug tests remain uncorrelated with soldier characteristics when we include year-by-month fixed effects and that random drug tests do not jointly predict soldier characteristics (the p-value of a joint F-test is 0.58). further includes year-by-month fixed effects.

The findings displayed in Table 2 substantiate our core assumption regarding the conditional randomness of mandatory drug test administration among junior enlisted male soldiers. This randomization in testing protocols validates our use of random drug test failure rates as a proxy measure for underlying substance use prevalence.

5 Methods

5.1 Estimating Random Drug Test Failure Rates

Testing for racial disparities in discretionary drug tests first requires us to determine the average monthly rate of random drug test failures separately for Black and non-Black soldiers within each battalion commander spell.

We define random drug test failure rates at the commander by race level as the share of random drug tests administered to soldiers of race r while they are assigned to battalion commander (spell) b that return a positive result:

$$D_{br} = \frac{\sum_{i \in r, it \in b} RandomTestFailure_{it}}{\sum_{i \in r, it \in b} RandomTest_{it}}$$

where $RandomTest_{it}$ is the number of times that soldier i (of race r) receives a random drug test in month t during months where soldier i is part of commander spell b. Similarly, $RandomTestFailure_{it}$ is the number of times soldier i receives a random drug test

in month t that returns a positive result for drugs.[13]

As discussed above and shown in Table 2 column 3, within battalions, there is a small amount of variance in the frequency of random drug tests across soldier occupations and soldiers' years of enlistment (i.e., soldiers' entry cohorts). As such, we also calculate random drug test failure rates within every combination of commander spell (b), race (r), and every combination of military occupation-by-cohort (k)

$$R_{brk} = \frac{\sum_{i \in rk, it \in b} RandomTestFailure_{it}}{\sum_{i \in rk, it \in b} RandomTest_{it}}$$

5.2 Regression Analysis - Estimate of Average Over-testing

We first document whether there are disparities in discretionary testing and whether these disparities exist even after controlling for drug usage by race as measured through random testing. Formally, we estimate the following equation at the soldier-by-month level:

$$Y_{it} = \alpha + \beta BlackSoldier_i + \gamma D_{br(it)} + \delta R_{brk(it)} + X_i'\Gamma + \pi_t + \epsilon_{it} \quad (2)$$

where $Y_{it} = 1$ if soldier i receives a discretionary drug test in month t, $BlackSoldier_i = 1$ if soldier i is Black, $D_{br(it)}$ is the commander-by-race random drug test failure rate (which is common to all soldiers of race r who are assigned to battalion b in month t), and $R_{brk(it)}$ is the commander-by-race-by-occupation-by-cohort random drug test failure rate. We also include year-by-month fixed effects, π_t, to account for changes in testing patterns over time, and X_i is a vector of individual-level controls that include occupation-by-cohort fixed effects and additional individual soldier controls in some specifications. We continue to report heteroskedasticity-robust standard errors (White 1980).

13. Technically $RandomTest_{it}$ equals 0 or 1 in all but 31 observations. The few instances where soldiers are randomly tested more than once during the same month occur because Commanders are not required to conduct all random drug tests on the same day of the month, and there is randomization with replacement across testing days.

After controlling for random test failure rates, we argue that β captures overtesting of Black soldiers if $\beta > 0$, and overtesting of non-Black soldiers if $\beta < 0$. The coefficients γ and δ reflect the responsiveness of discretionary drug testing to random drug test failure rates. By construction, $D_{br(it)}$ and $R_{brk(it)}$ are highly correlated, but we include both terms to minimize the possibility that estimates of β include bias from differential random testing rates by race owing to small variations in random testing across soldiers occupations and age. However, we also report results from specifications that omit $R_{brk(it)}$ to more accurately estimate how discretionary testing varies with drug usage at the spell by race level.[14]

We also explore whether disparate testing of Black or non-Black soldiers varies across random failure rates. First, we report results from regressions that add the interaction of $BlackSoldier_i$ and $D_{br(it)}$ (and the interaction of $BlackSoldier_i$ and $D_{brk(it)}$) to equation (2). Second, we estimate specifications that include fixed effects for quintiles of $D_{br(it)}$ and interact the quintile fixed effects with $BlackSoldier_i$, rather than controlling for $D_{br(it)}$ as a continuous variable.

5.3 Regression Analysis - Estimating Commander-Specific Over-testing

We also estimate commander-specific instances of over-testing by estimating a variation of Equation 2 that includes fixed effects for the interaction of battalion and soldier race. This specification also uses a more granular measure of battalion random drug test failure rates, calculated monthly by race, by using indicator variables for whether there was a random test failure in a specific month by race. We control for 12 months of failure rate history within a battalion, although results are consistent with different time windows.

$$Y_{it} = \alpha + \sum_{b \in B} \delta_b J_{b(i)} + \sum_{b \in B} \gamma_b \big(J_{b(i)} \times BlackSoldier_i \big) + \sum_{m=-11}^{m=0} \sigma_{brm} + X_i' \Gamma + \pi_t + \epsilon_{it} \qquad (3)$$

14. We also report results from specifications that exclude $D_{br(it)}$, as explained below.

The unit of observation remains a soldier-by-month, and the outcome indicates whether a discretionary test was administered for a specific soldier-month observation. The $J_{b(i)}$ terms are fixed effects for all battalion commander spells in B (the set of all battalions). The σ_{brm} terms are fixed effects for every combination of battalion, race, and twelve variables that account for random failures in the past twelve months. The model controls for year-by-month (π_t) fixed effects and soldier-level characteristics (X_i) that include occupation-by-cohort fixed effects.[15] We do not report standard errors as most battalion-level estimates are too imprecise to statistically distinguish from zero. This limitation partly motivates our use of brigade-level analysis later in the paper.

The coefficient of interest is γ_b. By including battalion commander-spell fixed effects (δ_b), this coefficient captures the difference in discretionary testing between Black and non-Black soldiers, accounting for individual soldier characteristics and the battalion's history of random test failures. To assuage concerns that our monthly controls do not accurately account for drug usage in a battalion, we also estimate results for a sample of battalions with very similar failure rates, and larger battalions.

These commander-specific estimates serve as key inputs for our Empirical Bayes analysis, where we refine inference by leveraging information across brigades to account for statistical noise and estimation uncertainty.

5.4 Empirical Bayes Estimation of Disparate Testing

To improve inference regarding disparities in discretionary drug testing, we employ empirical Bayes (EB) methods, following Kline, Rose, and Walters (2022, 2024) and Walters (2024). This method leverages information from the distribution of racial disparities in discretionary drug testing across military brigades to help distinguish between genuine heterogeneity and statistical noise.

15. The vector X_i includes separate indicator variables that equal one in months where a battalion commander lacks 12 months of command history (e.g., soon after taking command). We assign these months the mean monthly failure rate within the sample to avoid dropping observations. Equation (3) also includes a binary variable for each month-by-race combination to indicate if any testing of soldiers in a specific race occurred for that commander during the specific month.

We limit our analysis to brigades, denoted by j, that have at least 10 different battalion commander spells, resulting in a sample of 33 brigades covering 669 of 782 (86%) battalion commander spells. Our objective is to estimate brigade-specific disparities in discretionary testing, θ_j, and use the inferred distribution to refine these estimates.

5.4.1 Estimation Strategy

Brigade-level estimates of disparate testing are constructed as the mean of battalion-specific estimates:

$$\hat{\theta}_j = \frac{1}{n_j} \Sigma_{b=1}^{B} D_{bj} \hat{\gamma}_b$$

where $D_{bj} = 1\{D_b = j\}$ indicates if battalion commander spell b belongs to brigade j, and $\hat{\gamma}_b$ represents the battalion-specific estimates obtained via equation (3). We weight $\hat{\gamma}_b$ by the number of soldier-month observations in each battalion.

The standard error for each estimate $\hat{\theta}_j$ is computed as:

$$s_j = \sqrt{\frac{1}{n_j(n_j-1)} \Sigma_{b=1}^{n_j} \left(\hat{\gamma_{jb}} - \hat{\theta}_j\right)^2}$$

If we assume that true levels of disparate testing in brigades are independent and identically distributed draws from a normal mixing distribution, then we can construct a bias-corrected estimate of the variance of the underlying distribution of disparate testing as:

$$\hat{\sigma_\theta^2} = \frac{1}{J-1} \sum_{j=1}^{J} \left[(\hat{\theta}_j - \hat{\mu_\theta})^2 - \frac{J-1}{J} s_j^2 \right]$$

where $\hat{\mu_\theta} = \frac{1}{J} \sum_{j=1}^{J} \hat{\theta}_j$, the mean of brigade-level disparate testing estimates. We subtract $\frac{J-1}{J} s_j^2$ in our estimate of the variance of the mixing distribution as a bias correction to account for the expected noise in raw brigade-level estimates.

5.4.2 Linear Shrinkage and Nonparametric Estimation

We use the estimated mixing distribution to construct linear shrinkage posterior estimates of brigade-level disparate testing:

$$\hat{\theta_f^*} = \Big(\frac{\hat{\sigma_\theta^2}}{\hat{\sigma_\theta^2} + s_f^2}\Big)\hat{\theta_f} + \Big(\frac{s_f^2}{\hat{\sigma_\theta^2} + s_f^2}\Big)\hat{\mu_\theta}$$

We additionally relax our assumption that brigade-level disparate testing estimates follow a normal distribution by estimating a nonparametric mixing distribution, along with corresponding nonparametric posterior estimates for each brigade, that use the log spline deconvolution estimator of Efron (2016).

5.4.3 Multiple Hypothesis Testing

We also use Empirical Bayes methods to show that brigade-level estimates of disparate drug testing are robust to multiple hypothesis testing. We estimate the population share of true nulls (i.e., brigades where there is no disparate testing of Black soldiers) as:

$$\hat{\pi_0} = \frac{\sum_{j=1}^{J} 1\{p_j > \lambda\}}{(1-\lambda)J},$$

where p_j is the p-value for brigade j and λ is a tuning parameter that reflects which p-values in the sample that we assume to be true nulls. We initially adopt a very conservative value of $\lambda = 0.05$, implying that disparate discretionary testing of Black soldiers is not present in any brigades with p-values above 0.05.

We to can then use this estimate to construct q-values for each brigade, which reflect the probability that there is no disparate testing of Black soldiers conditional on rejecting the null hypothesis (i.e., obtaining a p-value less than 0.05) as:

$$q_j = \frac{p_j \hat{\pi_0}}{\hat{F}(p_j)},$$

where $\hat{F}(p_j)$ is the share of brigade p-values less than or equal to p_j.

6 Results

6.1 Aggregate Results

First, we present our main results in Figure 2. The graph displays local averages of discretionary test rates and random failure rates, categorized by soldier race., among soldiers assigned to the same battalion. Each square and circle represents an equal number of observations by race. We also include fitted lines to the underlying data for each racial group. Rates are calculated using the method outlined in Section 5.1.

There are two key takeaways from this figure. First, on average, Black soldiers experience higher discretionary testing rates compared to their counterparts, regardless of the random testing failure rate in their battalion (red squares are consistently above gray circles for any given random failure rate). Second, in Figure 2, the linear fit exhibits a similar upward slope for both Black and non-Black soldiers. This indicates that commanders increase discretionary testing in response to higher random failure rates and that this response is consistent across racial groups at least in the raw data.

More formal regression results are reported in Table 3. Coefficients are multiplied by 100 for ease of interpretation. The first column reports the average difference in discretionary testing rates between Black and non-Black soldiers from a regression that includes fixed effects for every combination of occupation-by-cohort and year-by-month-fixed effects. The coefficient of 0.0947, which is statistically significant with a t-stat that exceeds 10, implies that discretionary drug tests are 0.0947 percentage points more common among Black soldiers than among Non-Black soldiers. Relative to the mean rate of discretionary drug testing of 0.083% among non-Black soldiers, Black soldiers are roughly 114% more likely to be discretionarily tested than non-Black soldiers who are of the same occupation and who enlisted in the Army in the same year.

Some of this disparity in discretionary drug testing is explained by differences in drug usage between Black and Non-Black soldiers. Column 2 reports estimates from a regression that controls for average random drug test fail rates at the commander spell-by-race

level, or D_{br} as described in Section 5.1 above. The point estimate of 0.0861 implies that when comparing Black soldiers to non-Black soldiers assigned to commander-spell-by-race groups with similar random drug test failure rates, Black soldiers are still roughly twice as likely to be discretionarily drug-tested as non-Black soldiers. The positive and statistically significant point estimate on $D_{br(it)}$ indicates that a one standard deviation increase in random test failure rates at the commander-spell-by-race level translates to about a 0.2 standard deviation increase in discretionary drug test rates.

Column 3 of Table 3, which exactly matches Equation (2) above, shows that racial differences in discretionary drug testing rates do not change when further controlling for random drug test failure rates within every combination of commander spell, race, occupation, and cohort (i.e., R_{brk}), which accounts for the small amount of within-battalion variation in the frequency of random drug tests by occupation and enlistment year.[16] The inclusion of additional soldier-level controls for AFQT scores, education levels, age, waivers for a prior felony conviction or other disqualifying conduct, and drug test and medical exam failures during enlistment screening also have little impact on the estimated difference in discretionary drug test rates by race (column 4). Finally, column 5 shows no significant interaction between random drug test failure rates and race, indicating that commanders' discretionary testing decisions do not vary based on the race of the soldiers. This finding should alleviate concerns that commanders interpret drug usage rates differently across racial groups.

While Army policy establishes drug testing protocols at the battalion level, company commanders may also play a significant role in monitoring and addressing substance use within their units. To address potential concerns about the appropriate organizational level for our analysis, Table A6 demonstrates that our findings remain consistent when controlling for company-by-race estimates of drug usage. A separate analytical challenge arises from instances where no random tests were conducted within specific Commander

16. For instances where there are no random drug tests within a combination of battalion-by-race-by-occupation-by-cohort, we assign R_{brk} the random drug test failure rate at the battalion-by-race level. Table A5 reports estimates of Equation (2) when we exclude observations where there are no random drug tests within a combination of battalion-by-race-by-occupation-by-cohort.

Spell × Race × MOS × Entry-Year cells. In our main analysis, we address these missing values by substituting them with the broader Commander Spell × Race failure rates. To verify this approach doesn't drive our results, Table A5 shows our findings remain robust even when we completely exclude these observations with missing cell-level random test data. These complementary analyses strengthen confidence in both our organizational-level approach and our treatment of missing test data.

6.2 Alternative Explanations: Addressing potential non-random assignment to battalions and measurement error

In Table 4, we address concerns that soldiers may not be perfectly randomly assigned to battalions and that drug test failure rates do not adequately measure differences in underlying drug usage by race. Column 1 replicates column 4 from Table 3 for reference. To explore whether our results could be explained by a tendency for Black soldiers to be disproportionately assigned to battalions that have a higher propensity to administer discretionary drug tests regardless of soldier race or underlying drug usage, column 2 of Table 4 estimates a specification that includes fixed effects for battalion commander spell. Notably, the disparity in discretionary drug test rates by race remains nearly identical even when comparing Black and non-Black soldiers who are assigned to the same battalion, suggesting that any non-random assignment of soldiers to battalions is unlikely to explain racial disparities in discretionary drug testing.

Even though the average soldier in our sample is randomly tested twice per year, random drug test failures are relatively rare. This raises the possibility that random drug test failure rates measure actual drug usage with error. An additional concern in our setting is that this measurement error may vary by race because there are fewer Black soldiers, and therefore fewer random drug tests of Black soldiers, than non-Black soldiers. As an initial check that measurement error is unlikely to explain differences in discretionary drug testing by race, Column 3 of Table 4 reports results when we restrict our sample to battalion-by-race combinations where the number of Black and Non-Black soldiers are

similar. Specifically, we exclude soldiers from battalions where their numbers exceed the highest number of Black soldiers in any battalion (64% of the sample) and we exclude Black soldiers from battalions where their numbers are lower than the lowest number of Non-Black soldiers in any battalion (11% of the sample).[17] In specification we also control for quantiles of battlion-spell-race size.[18] Within this subsample, the difference in discretionary drug testing by race is very similar to the difference reported in our baseline specification.

Concerns about measurement error should also be less significant in large battalion spells. Column 4 of Table 4 therefore reports differences in discretionary drug testing by race when we restrict the sample to battalions with an above-median number of total soldiers. Even in large battalions, Black soldiers are over-tested relative to non-Black soldiers, indicating that our results are unlikely to be driven by measurement error.

Another concern is that commanders might be forming unbiased beliefs about drug use based on information other than random drug test results. This could potentially explain why our results do not suggest that commanders respond to random drug testing in a 1-for-1 manner, meaning that when drug failures increase by 10%, commanders don't necessarily increase testing by 10%. Instead, our results suggest a smaller response.

However, there are reasons to doubt that this explains our findings. First, about three-fourths of all soldiers charged with drug crimes had previously tested positive for drugs through formal Army tests, with most (73%) of these tests being random or unit-wide. Second, we can narrow our analysis to situations where Black and non-Black soldiers in the same battalion have similar drug failure rates, or where non-Black soldiers fail at higher rates. Unless commanders have knowledge of drug use that is not reflected in positive random drug tests and that typically does not lead to criminal or misdemeanor drug charges, it's hard to justify why commanders would test Black soldiers more in

17. There are battalions where only Black soldiers are in the subsample of column 3 of Table 4; other battalions only have non-Black soldiers in the column 3 subsample. We also note that this specification does not include battalion fixed effects and thus does not force comparisons of Black and non-Black soldiers in the same battalion.

18. The number of soldiers in a spell-race group is correlated with race, which may lead to some attenuation in β. However, despite this potential attenuation, our main result remains robust.

these situations. Yet, as shown in columns 5 and 6 of Table 4, if anything, the racial gap in discretionary drug testing as a percent of mean non-Black discretionary test rates *expands* under these sample restrictions.

6.3 Battalion Spell Estimates

This section examines heterogeneity in disparate testing across military units. We begin by estimating battalion-specific disparities to capture variation at the unit level where discretionary testing decisions are made. Battalion-level estimates are particularly relevant because battalions are the institutional level at which drug testing policy is implemented, with battalion commanders overseeing both random and discretionary testing practices. Additionally, these estimates serve as the foundation for our subsequent Empirical Bayes analysis, which aggregates information across battalions to refine our understanding of disparate testing effects while mitigating statistical noise. However, inference at this granular level presents challenges due to limited statistical power, resulting in excessively wide confidence intervals. To address this limitation while preserving our ability to characterize heterogeneity, we subsequently aggregate battalion-level estimates to the brigade level.

Next, we present estimates of differential testing patterns across battalion spells. Figure 4 plots γ_b from Equation 3, where each circle represents a spell's over-testing measure (coefficient). Coefficients above zero indicate higher rates of discretionary testing for Black soldiers relative to predicted drug use, while those below zero indicate higher relative testing rates for non-Black soldiers. Circle size corresponds to the frequency of estimates within a 0.00005 bin. This distribution reveals a systematic pattern of higher relative testing rates for Black soldiers, with larger magnitudes of differential testing for this group. The average disparity across spells is 0.00096, approximately 50% of the mean discretionary testing rate for Black soldiers.

There are also a non-trivial number of commanders with large testing disparities: 67 commanders test non-Black soldiers more than 1/2 of a standard deviation above the

mean, while 14 commanders test non-Black soldiers more than one full standard deviation above the mean. However, large over-testing is more common for Black soldiers: 281 commanders over-test Black soldiers by more than 1/2 a standard deviation, and 170 commanders exceed one full standard deviation. These results highlight substantial variation in individual commander discretion, with a clear pattern of systemic racial disparities in discretionary testing practices

To further strengthen our findings, we examine two key subsamples. Figure 5(a) focuses on battalions where differences in underlying substance use between racial groups are minimal, allowing us to directly address concerns that observed testing disparities might still be driven by racial differences in drug use. We examine battalions in the bottom quartile of Black-non-Black random failure rate differences, where the average difference is just 0.0034 (compared to the full sample average of 0.0137). Even in these 188 spells where racial differences in drug use are smallest, we continue to observe systematic disparities: Black soldiers experience discretionary testing rates that are 0.0071 higher than non-Black soldiers in the same battalion spell. The persistence of these disparities in units with minimal drug use differences strengthens our conclusion that differential testing cannot be explained by underlying behavioral differences.

Figure 5(b) addresses potential measurement error concerns by examining battalions with over 3,050 soldier months, where larger sample sizes provide more precise estimates of underlying drug use patterns. In these 391 larger units, the pattern becomes even more pronounced: discretionary testing disparities increase to 0.0092, further reinforcing the robustness of our findings. Additionally, Figure 4 demonstrates the substantial variation in testing disparities across battalion spells, highlighting that while some commanders exhibit little to no racial disparity in testing, others engage in significantly higher over-testing of Black soldiers. This variation underscores the role of individual commander discretion in shaping observed racial disparities in discretionary testing.

The consistency of these results across different subsamples confirms that the observed disparities are not simply artifacts of unit-level measurement error or statistical

noise but reflect meaningful differences in battalion-level discretionary testing policies. Next, we use these battalion-level estimates—aggregated through Empirical Bayes methods—to construct brigade-level estimates, allowing for a more precise assessment of disparate testing patterns at higher levels of command.

6.4 Empirical Bayes Estimates of Brigade-Level Disparate Testing

The previous analysis estimated disparities in discretionary drug testing at the battalion level, identifying significant heterogeneity across commanders. However, battalion-level estimates may be subject to noise, particularly when sample sizes are small or when variation in testing behavior is limited. To address this, we apply empirical Bayes (EB) estimation to brigade-level discretionary testing rates, refining our estimates by borrowing strength from the full distribution of disparate testing across brigades. This allows us to distinguish genuine heterogeneity from statistical noise and improve inference regarding disparate testing patterns.

Columns (1) and (2) of Table A7 report the raw brigade-level estimates ($\hat{\theta_j}$) and their corresponding standard errors (s_j), which are constructed as the mean of battalion-specific estimates weighted by soldier-month observations. Column (3) presents p-values from one-tailed hypothesis tests for the null hypothesis that brigades do not disproportionately test Black soldiers. While disparate testing occurs in all brigades, statistical significance at the 5% level is only observed in two-thirds of them, suggesting that some observed disparities may be driven by small-sample noise rather than systematic differences in testing behavior.

To further refine our estimates, Figure 6, panel (a), presents the distribution of raw brigade-level estimates of racial disparities in drug testing (white bars), overlaid with the bias-corrected mixing distribution (black line), which adjusts for expected statistical noise and assumes an underlying normal distribution. Bias correction reduces the standard deviation of racial disparities in drug testing from 0.0629 percentage points to 0.0349 percentage points. This suggests that $(1 - (\frac{0.0349}{0.0629})^2) \times 100 = 69\%$ of the observed variance

in brigade-level testing rates is attributable to statistical noise while the remainder likely reflects true heterogeneity.

Our estimated mixing distribution allows us to derive linearly shrunk estimates of brigade-level testing disparities, which are depicted as red bars in panel (a) of Figure 6. As expected, shrinkage reduces variation, but notable differences persist across brigades: disparities at the 90th percentile remain 130% higher than those at the 10th percentile (0.123 vs. 0.053), indicating that substantial heterogeneity exists in discretionary testing behavior across brigades.

Relaxing the normality assumption suggests that heterogeneity in disparate drug testing across brigades is likely to be larger than what is implied by linear shrinkage posteriors. As reported in panel (b) of Figure 6, the standard deviation of nonparametric posterior estimates is 63% as large as the unadjusted standard deviation of brigade-level estimates, implying that 40% of the variance in raw brigade-level testing rates reflect true heterogeneity in disparate testing. Nonparametric posteriors also suggest that disparate testing at the 90th percentile is roughly 200% higher than disparate testing at the 10th percentile (0.159 vs. 0.052).

Finally, we assess robustness to multiple hypothesis testing using the empirical Bayes false discovery rate (FDR) approach. Overall, 22 of 33 brigades exhibit statistically significant evidence of disparate discretionary drug testing of Black soldiers relative to non-Black soldiers (see column (3) of Table A7). This conclusion does not change when adjusting for multiple hypothesis testing. Setting $\lambda = 0.05$ generates an estimate of true nulls of $\hat{\pi_0} = 0.35$ (as described in Section 5.4), meaning that at least 65% of brigades must be disparately testing Black soldiers to rationalize the observed brigade-level p-values in our sample. Additionally, corresponding q-values, presented in column (4) of Table A7, confirm that brigade-level estimates of disparate testing remain statistically significant after adjusting for false discovery rates.

Overall, the empirical Bayes results suggest that while a substantial portion of observed heterogeneity in racial drug testing disparities may reflect statistical noise, there

remain real differences across brigades. This finding underscores the importance of targeted policy interventions aimed at commanders in units exhibiting extreme disparities, rather than broad institutional reforms.

7 Conclusion

This study provides robust evidence of racial disparities in military drug testing that cannot be explained by differences in drug use. Analyzing over 420,000 random tests and 2,500 discretionary tests administered to 170,000 junior enlisted soldiers (2015–2022), we find that Black soldiers are nearly twice as likely as non-Black soldiers to face discretionary testing, despite similar drug use rates. Battalion-level estimates reveal substantial variation across commanders, with some systematically over-testing Black soldiers. This heterogeneity underscores the role of individual discretion in shaping racial disparities.

Applying Empirical Bayes methods to brigade-level testing patterns, we estimate that between 31% and 40% of the observed variance in brigade-level disparities is due to genuine heterogeneity rather than statistical noise. Even after correcting for noise, drug testing disparities by race remain large and meaningful: the 90th percentile of brigade-level disparate testing is 130% to 200% larger than the 10th percentile of disparate testing, confirming that racial bias persists beyond individual decision-makers.

These findings have broad implications for organizational oversight and accountability. First, they show how managerial discretion can reinforce racial disparities even in highly structured institutions with formal oversight. Second, given that 19% of misconduct-based discharges are drug-related, biased testing may disproportionately impact the economic mobility of Black soldiers, particularly those who rely on military service for career advancement. Finally, since military veterans comprise one in four law enforcement officers, discriminatory practices in the Army may carry over into civilian policing, shaping racial disparities in law enforcement.

As workplace drug testing becomes more prevalent across sectors, these findings highlight the risks of unchecked managerial discretion in disciplinary policies. Targeted inter-

ventions focused on individual decision-makers may be more effective than broad institutional reforms in mitigating racial disparities in enforcement.

References

Army Substance Abuse Program. 2018. *Unit Prevention Leader's Urinalysis Collection Handbook.* Washington D.C.

Arnold, David, Will Dobbie, and Peter Hull. 2022. "Racial Discrimination in Bail Decisions." *The Quarterly Journal of Economics* 137 (3): 1885–1932. https://doi.org/10.1093/qje/qjab018.

Arnold, David, Will Dobbie, and Crystal S Yang. 2018. "Racial bias in bail decisions." *The Quarterly Journal of Economics* 133 (4): 1885–1932.

Baltrunaite, Audinga, Giulia Bovini, and Sauro Mocetti. 2023. "Managerial talent and managerial practices: are they complements?" *Journal of Corporate Finance* 79:102348.

Bandiera, Oriana, Andrea Prat, Stephen Hansen, and Raffaella Sadun. 2020. "CEO behavior and firm performance." *Journal of Political Economy* 128 (4): 1325–1369.

Benson, Alan, Simon Board, and Moritz Meyer-ter-Vehn. 2024. "Discrimination in hiring: Evidence from retail sales." *Review of Economic Studies* 91 (4): 1956–1987.

Bertrand, Marianne, and Antoinette Schoar. 2003. "Managing with style: The effect of managers on firm policies." *The Quarterly Journal of Economics* 118 (4): 1169–1208.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does management matter? Evidence from India." *The Quarterly journal of economics* 128 (1): 1–51.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, and John Van Reenen. 2015. "Does management matter in schools?" *The Economic Journal* 125 (584): 647–674.

Bloom, Nicholas, and John Van Reenen. 2007. "Measuring and explaining management practices across firms and countries." *The Quarterly Journal of Economics* 122 (4): 1351–1408.

Bruhn, Jesse, Kyle Greenberg, Matthew Gudgeon, Evan K Rose, and Yotam Shem-Tov. 2024. "The effects of combat deployments on veterans' outcomes." *Journal of Political Economy* 132 (8): 000–000.

Department of the Army. 2016. *Army Regulation 600-85: The Army Substance Abuse Program.* Washington D.C.

———. 2019. *Army Regulation 614-200: Enlisted Assignments and Utilization Management.* Washington D.C.

Efron, Bradley. 2016. "Empirical Bayes deconvolution estimates." *Biometrika* 103 (1): 1–20.

Fenizia, Alessandra. 2022. "Managers and productivity in the public sector." *Econometrica* 90 (3): 1063–1084.

Glover, Dylan, Amanda Pallais, and William Pariente. 2017. "Discrimination as a self-fulfilling prophecy: Evidence from French grocery stores." *The Quarterly Journal of Economics* 132 (3): 1219–1260.

Goncalves, Felipe, and Steven Mello. 2021. "A Few Bad Apples? Racial Bias in Policing." *American Economic Review* 111 (5): 1406–1441. https://doi.org/10.1257/aer.20181613.

Greenberg, Kyle, Matthew Gudgeon, Adam Isen, Corbin Miller, and Richard Patterson. 2022. "Army service in the all-volunteer era." *The Quarterly Journal of Economics* 137 (4): 2363–2418.

Hasin, Deborah S, Dvora Shmulewitz, and Aaron L Sarvet. 2019. "Time trends in US cannabis use and cannabis use disorders overall and by sociodemographic subgroups: a narrative review and new findings." *The American journal of drug and alcohol abuse* 45 (6): 623–643.

Kline, Patrick, Evan K Rose, and Christopher R Walters. 2022. "Systemic discrimination among large US employers." *The Quarterly Journal of Economics* 137 (4): 1963–2036.

Kline, Patrick, Evan K Rose, and Christopher R Walters. 2024. "A Discrimination Report Card." *American Economic Review* 114 (8): 2472–2525.

Lang, Kevin, and Ariella Kahn-Lang Spitzer. 2020. "Race Discrimination: An Economic Perspective." *Journal of Economic Perspectives* 34 (2): 68–89. https://doi.org/10.1257/jep.34.2.68.

Lazear, Edward P, Kathryn L Shaw, and Christopher T Stanton. 2015. "The value of bosses." *Journal of Labor Economics* 33 (4): 823–861.

Le, Aurora B., Emily Urban-Wojcik, Meghan Seewald, and Briana R. Mezuk. 2024. "The Relationship Between Workplace Drug Policies, Opioid Misuse, and Psychological Distress: Evidence From the 2020 National Survey on Drug Use and Health." *New Solutions: A Journal of Environmental and Occupational Health Policy* 34 (1): 22–37. https://doi.org/10.1177/10482911241231523.

Lyle, David S. 2006. "Using Military Deployments and Job Assignments to Estimate the Effect of Parental Absences and Household Relocations on Children's Academic Achievement." *Journal of Labor Economics* 24 (2): 319–350.

Minni, Virginia. 2023. "Making the invisible hand visible: Managers and the allocation of workers to jobs." *Working Paper*.

Moeller, Karen E, Kelly C Lee, and Julie C Kissack. 2008. "Urine drug screening: practical guide for clinicians." In *Mayo clinic proceedings*, 83:66–76. 1. Elsevier.

Montgomery, LaTrice, Shapree Dixon, and Dale S. Mantey. 2022. "Racial and Ethnic Differences in Cannabis Use and Cannabis Use Disorder: Implications for Researchers." *Current Addiction Reports* 9 (1): 14–22. https://doi.org/10.1007/s40429-021-00404-5. https://pmc.ncbi.nlm.nih.gov/articles/PMC8896813/pdf/nihms-1770498.pdf.

Morin, Rich, and Andrew Mercer. 2017. "A closer look at police officers who have fired their weapon on duty." *Pew Research Center* 8.

Owens, Emily, and Bocar Ba. 2021. "The Economics of Policing and Public Safety." *Journal of Economic Perspectives* 35 (4): 3–28. https://doi.org/10.1257/jep.35.4.3.

Rasul, Imran, and Daniel Rogger. 2018. "Management of bureaucrats and public service delivery: Evidence from the Nigerian civil service." *The Economic Journal* 128 (608): 413–446.

Reingle Gonzalez, Jennifer M, Stephen A Bishopp, Katelyn K Jetelina, Ellen Paddock, Kelley Pettee Gabriel, and M Brad Cannell. 2019. "Does military veteran status and deployment history impact officer involved shootings? A case–control study." *Journal of Public Health* 41 (3): e245–e252.

Tsai, Thomas C, Ashish K Jha, Atul A Gawande, Robert S Huckman, Nicholas Bloom, and Raffaella Sadun. 2015. "Hospital board and management practices are strongly related to hospital performance on clinical quality metrics." *Health Affairs* 34 (8): 1304–1311.

United States Army. 2018. "Army Barracks Management Program Handbook." Available at: https://home.army.mil/imcom/2715/1639/0097/ABMP_Handbook_Final_9_January_2018.pdf, Last accessed: October 9, 2024.

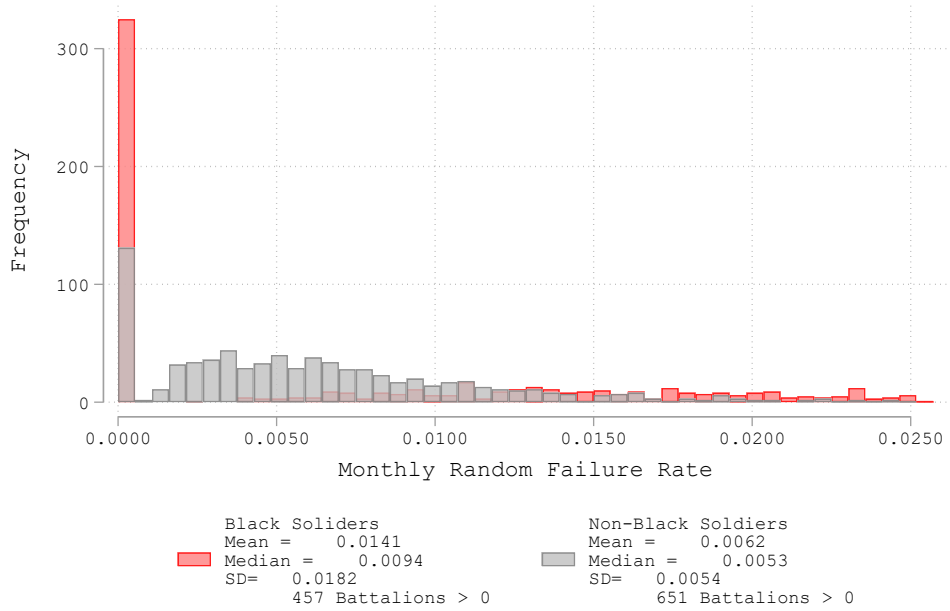
Walters, Christopher. 2024. "Empirical Bayes methods in labor economics." In *Handbook of Labor Economics*, 5:183–260. Elsevier.

White, Halbert. 1980. "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." *Econometrica: journal of the Econometric Society*, 817–838.

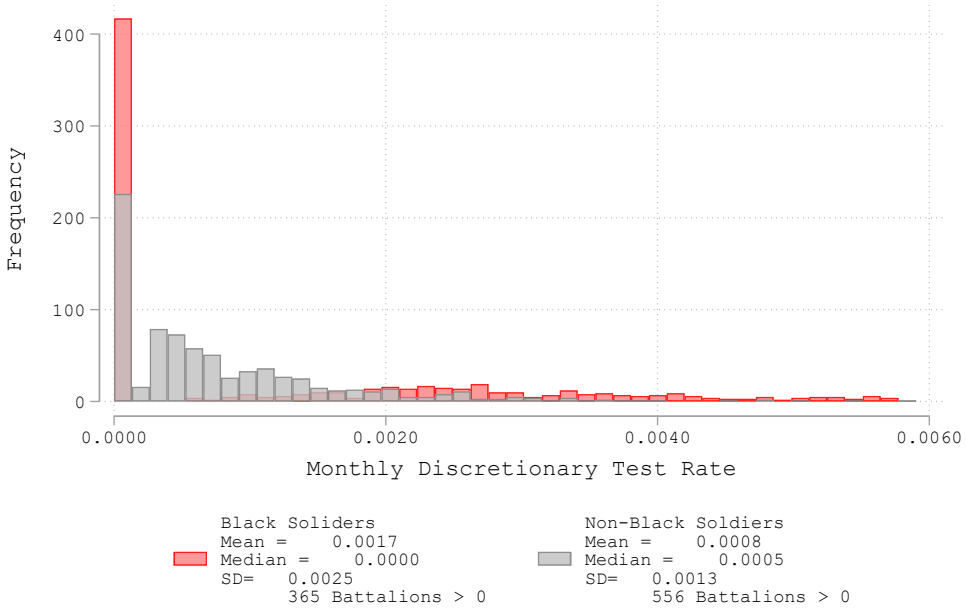
Wozniak, Abigail. 2015. "Discrimination and the Effects of Drug Testing on Black Employment." *Review of Economics and Statistics* 93 (7): 548–566.

8 Figures & Tables

Figure 1: Distribution of Battalion Commander Rates



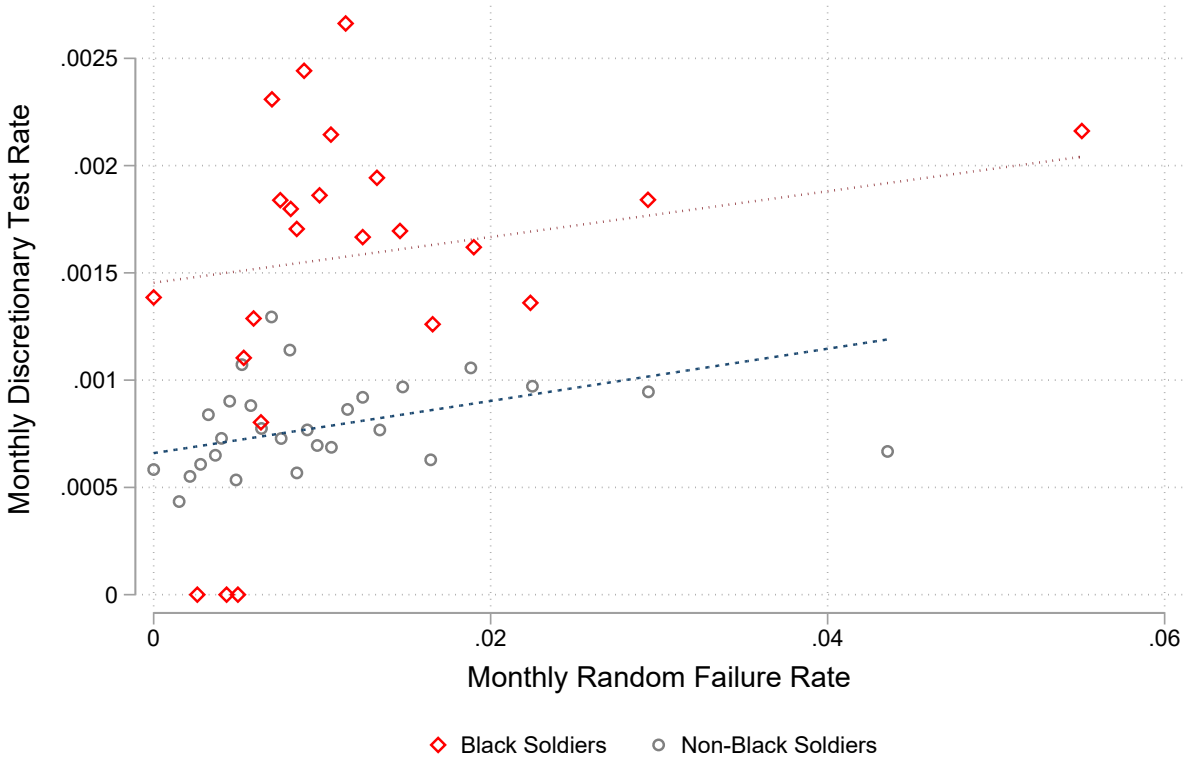
(a) Random Failure Rates



(b) Discretionary Testing Rates

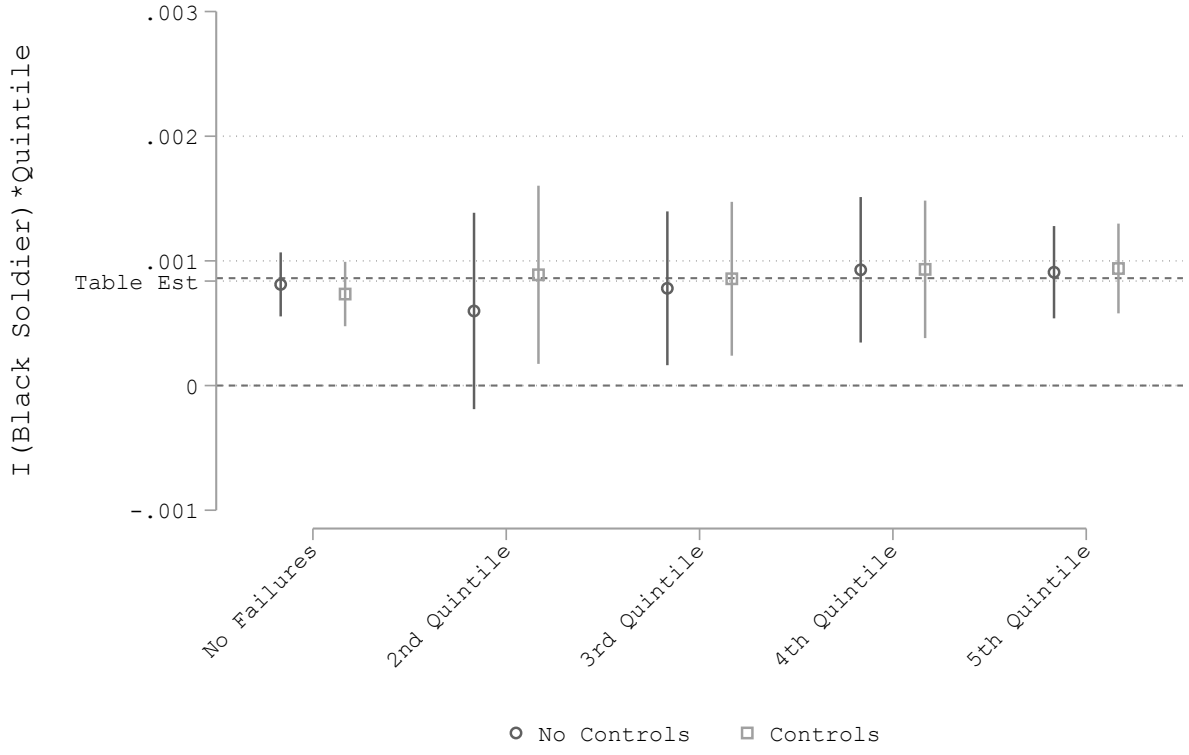
Notes: Figures plot battalion commander spell-level estimates of random and discretionary testing rates. The width of each bar is 1/10th of a standard deviation.

Figure 2: Scatter of Discretionary Testing and Random Failure Rates



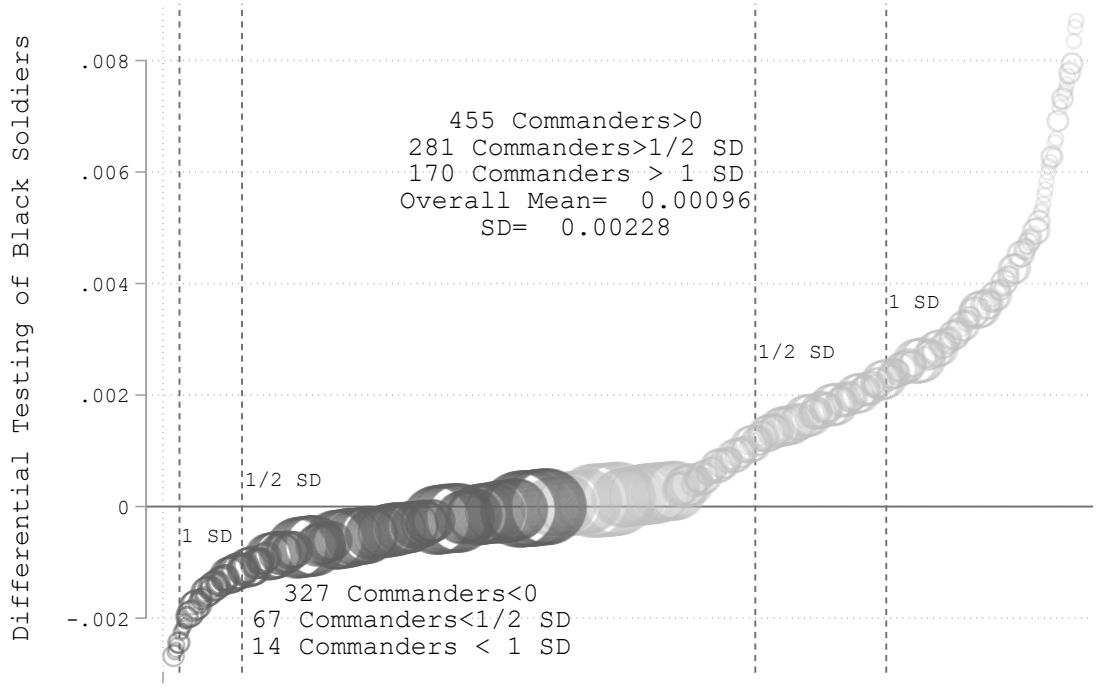
Notes: Observations are grouped so that each point includes an equal number of battalion spells by race. The fitted line is a linear fit across all random drug test failure rates. Rates are average monthly discretionary testing rates or average monthly random drug test failure rates.

Figure 3: Heterogeneity by Failure Rates



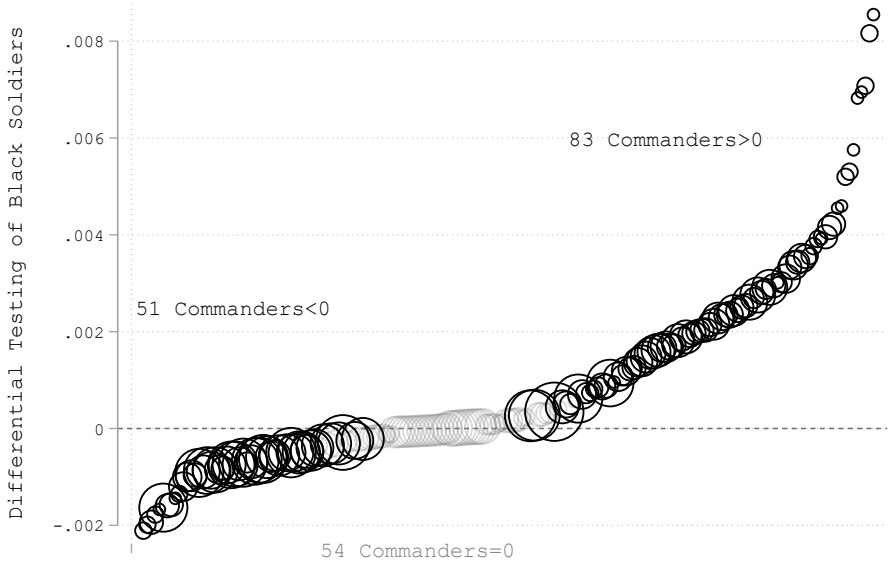
Notes: This figure measures the overtesting of Black soldiers by quintile of random drug test failure rates. For example, the second coefficient plotted is the disparity in discretionary testing for all battalions in the 2nd quintile of failure rates. Estimates from regressions with and without controls (occupation-by-entry year, year-by-month fixed effects, and individual soldier controls) are shown. Robust standard errors are reported in parentheses.

Figure 4: Commander Level Results

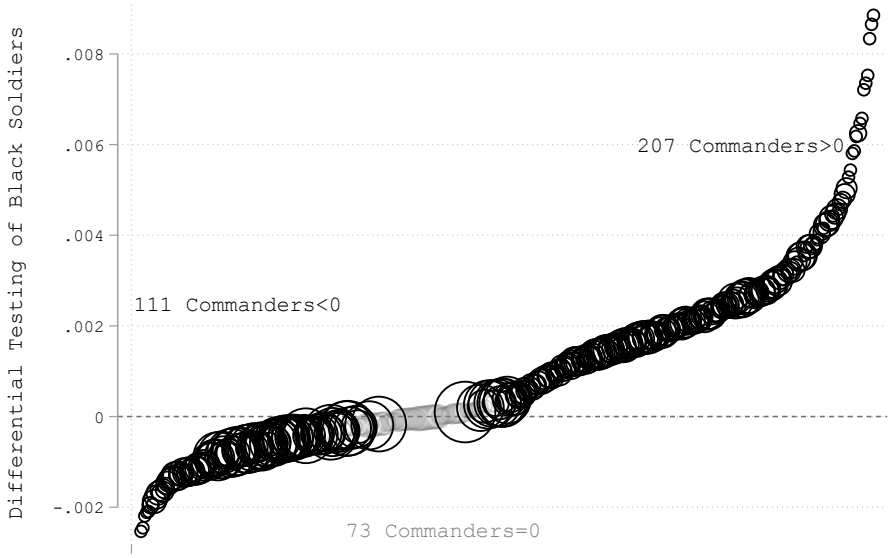


Notes: Notes: This figure presents commander-specific estimates of the overtesting of Black soldiers from Equation 3. Black-colored coefficients are less than zero. The regression includes fixed effects for occupation-by-entry year and year-by-month, as well as controls for individual soldier characteristics. Additionally, the regression accounts for the monthly history of random test failures within a battalion. The figure includes four reference lines: two at 1/2 and 1 standard deviation above zero, and two at 1/2 and 1 standard deviation below zero, to illustrate the distribution of overtesting estimates. The average coefficient is 0.00096, with a standard deviation of 0.00228. The size of each circle corresponds to the frequency of estimates within 0.00005-width bins, with larger circles indicating bins that contain more estimates.

Figure 5: Commander Level Results Subgroups



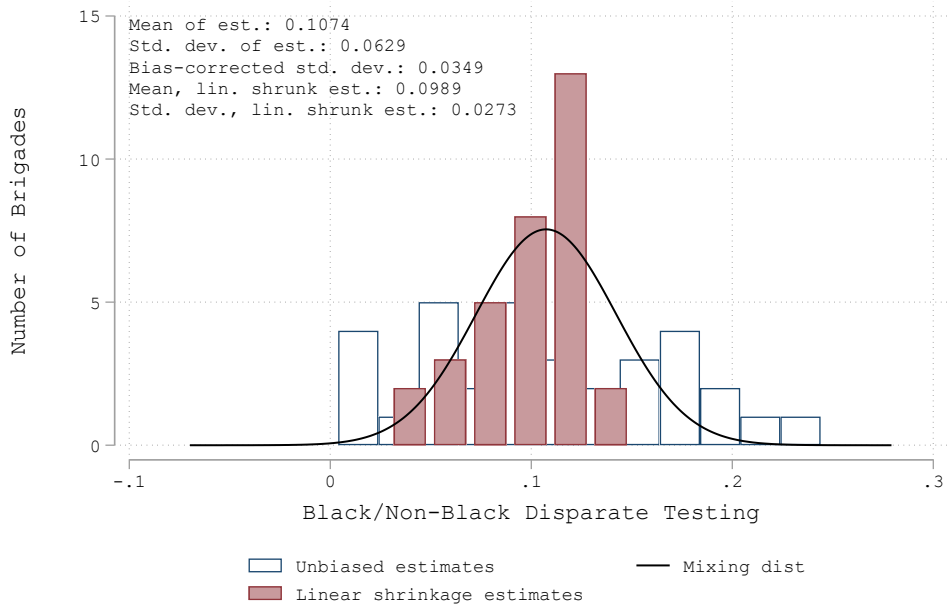
(a) Battalions with similar Black and non-Black failure rates



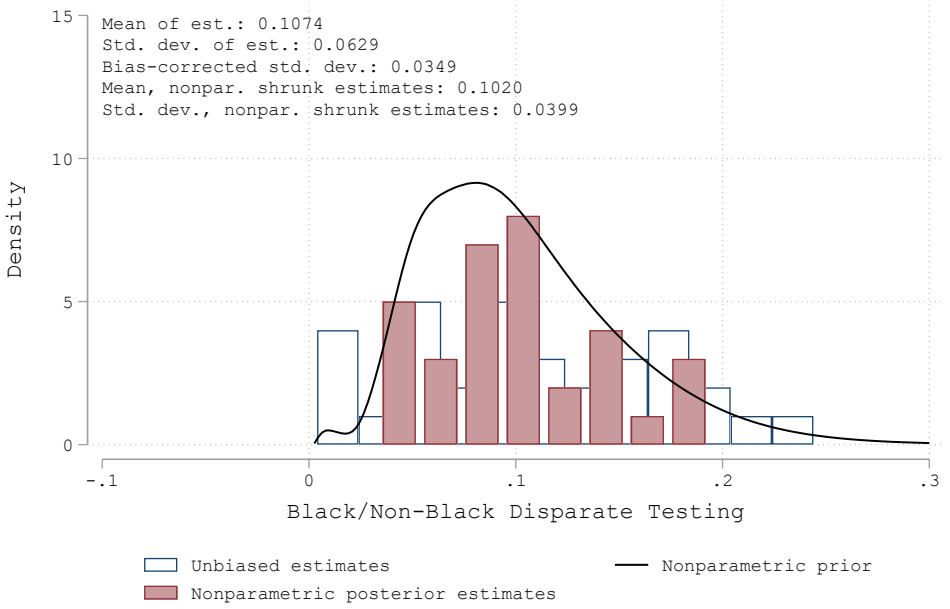
(b) Larger Battalions

Notes: This figure presents commander spell-specific estimates of the overtesting of Black soldiers. Black-colored coefficients are statistically significant at the five percent level. The regression analysis includes fixed effects for occupation-by-entry year and year-by-month, as well as controls for individual soldier characteristics. Additionally, the regression accounts for the monthly history of random test failures within a battalion. Panel (a) restricts to battalions where Black and non-Black soldiers have similar failure rates (188 spells). Panel (b) restricts battalions to over 3050 soldier-months, the median number of soldier-months among spells in our sample.

Figure 6: Empirical Bayes Estimates of Brigade-Level Disparate Testing



(a) Estimates vs. Linear Shrinkage Posteriors



(b) Estimates vs. Nonparametric Posteriors

Notes: This figure presents brigade-level estimates of racial disparities in discretionary drug testing, using empirical Bayes shrinkage to adjust for estimation noise. Panel (a) compares raw estimates to their posterior distributions under linear shrinkage, while panel (b) contrasts linear and nonparametric shrinkage estimates. The mixing distribution, shown as a smoothed density, represents the inferred underlying distribution of disparate testing rates after accounting for statistical noise, while the nonparametric prior provides an alternative estimate that does not impose a specific functional form, allowing for greater flexibility in capturing variation across brigades.

Table 1: Summary Statistics Testing & Failure Counts and Rates

	All Soldiers	Black Soldiers	Non-Black Soldiers
Panel A: Counts			
Discretionary Tests	2469	697	1772
Discretionary Test Fails	582	209	373
Random Tests	416792	64719	352073
Random Test Fails	2897	835	2062
Unique Soldiers	172339	29902	142437
Panel B: Rates			
Random Tests	16.468%	15.895%	16.577%
Random Test Failure	0.745%	1.352%	0.628%
Discretionary Tests	0.098%	0.171%	0.083%
Discretionary Test Failure	25.562%	30.911%	24.798%
Panel C: Rates by Random Failure Rate Quintile			
Random Test Failure Rate Quintile 1	0%	0%	0%
Discretionary Test Rate Quintile 1	0.087%	0.142%	0.061%
Random Test Failure Rate Quintile 2	0.271%	0.367%	0.270%
Discretionary Test Rate Quintile 2	0.062%	0.122%	0.062%
Random Test Failure Rate Quintile 3	0.539%	0.554%	0.538%
Discretionary Test Rate Quintile 3	0.096%	0.171%	0.093%
Random Test Failure Rate Quintile 4	0.860%	0.903%	0.856%
Discretionary Test Rate Quintile 4	0.112%	0.196%	0.103%
Random Test Failure Rate Quintile 5	1.966%	2.643%	1.583%
Discretionary Test Rate Quintile 5	0.130%	0.189%	0.096%

Notes: Discretionary and random test failure rates are measured conditional on test administration. We calculate the failure rate as the total number of random/discretionary drug tests failures divided by the total number of random/discretionary drug tests.

Table 2: Balance Test: Correlation between Random Drug Testing and Soldier Characteristics

	Mean			
	(1)	(2)	(3)	(4)
Black	16.087	-0.189***	-0.040	-0.009
	[36.741]	(0.057)	(0.050)	(0.052)
Hispanic	17.834	-0.047	-0.059	-0.074
	[38.280]	(0.061)	(0.058)	(0.060)
Other Race	6.850	0.062	0.048	0.051
	[25.260]	(0.041)	(0.038)	(0.039)
HSG+	94.093	0.088**	0.020	-0.009
	[23.575]	(0.038)	(0.036)	(0.037)
Initial Approx. Age	20.291	-0.015***	-0.006	-0.004
	[2.886]	(0.005)	(0.004)	(0.004)
Initial AFQT	58.167	-0.053*	0.016	0.003
	[19.012]	(0.029)	(0.026)	(0.027)
Moral Waiver	1.748	0.003	-0.032	-0.026
	[13.107]	(0.021)	(0.020)	(0.021)
Drug Fail at Enlistment	0.310	-0.013	-0.012	-0.012
	[5.559]	(0.009)	(0.008)	(0.009)
Medical Fail at Enlistment	14.771	-0.012	0.008	0.024
	[35.481]	(0.057)	(0.054)	(0.056)
(Battalion Commander Spell)		Y	Y	Y
(BN CDR Spell)X(Occ)X(Enlist Yr)			Y	Y
(Year-x-Month FE)				Y
N (Soldier-x-Month)		2530976	2528643	2528643
P-value from Joint-F		0.00	0.33	0.59

Notes: Each coefficient is derived from a separate regression of soldier characteristics on the monthly number of random tests. The mean of the demographic covariate in each regression is listed in column (1); standard deviations are listed in brackets. Robust standard errors are reported in parentheses. Coefficients are scaled by a factor of 100 for ease of interpretation.

Table 3: Soldier Month Level Disparities in Drug Testing

	Outcome: Discretionary Test				
	(1)	(2)	(3)	(4)	(5)
Black Soldier=1	0.095***	0.086***	0.086***	0.081***	0.087***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)
(CDR) x (Race) Random Fail Rate		1.125***	0.865***	0.840***	1.282***
		(0.288)	(0.292)	(0.292)	(0.396)
(CDR) x (Race) x (Occ) x (Enlist Yr) Random Fail Rate			0.261***	0.258***	0.244**
			(0.088)	(0.088)	(0.116)
((CDR) x (Race) Random Fail Rate)*Black Soldier					-0.673
					(0.552)
((CDR) x (Race) x (Occ) x (Enlist Yr) Rand Fail Rt)*Black Soldier					0.023
					(0.172)
N	2530976	2530976	2530976	2530976	2530976
Outcome Mean*100	0.098	0.098	0.098	0.098	0.098
NonBlack Soldier Mean*100	0.083	0.083	0.083	0.083	0.083
% Sample Mean	97	88	88	83	89
% NonBlack Soldier Mean	114	103	103	98	104
(Occ)x(Enlist Yr) FE, Month FE	Y	Y	Y	Y	Y
Soldier Controls				Y	Y

Notes: This table reports differences in discretionary drug testing between Black and Non-Black soldiers from regression estimates corresponding to Equation (2). The unit of observation for each regression in this table is Soldier x Month. The RHS variable "(CDR)x(Race) Random Fail Rate" is the random drug test failure rate at the commander-by-race level ($D_{br(it)}$ from Equation (2)). The RHS variable "(CDR)x(Race)x(Occ)x(Enlist Yr) Random Fail Rate" is the random drug test failure rate at the commander-by-race-by-occupation-by-cohort level ($R_{brk(it)}$ from Equation (2)), where cohort reflects year of enlistment. For instances where there are no random drug tests within a combination of commander-by-race-by-occupation-by-cohort, we replace "(CDR)x(Race)x(Occ)x(Enlist Yr) Random Fail Rate" with the random drug test failure rate at the commander-by-race level. Robust standard errors are reported in parentheses. Coefficients are multiplied by 100 for ease of interpretation.

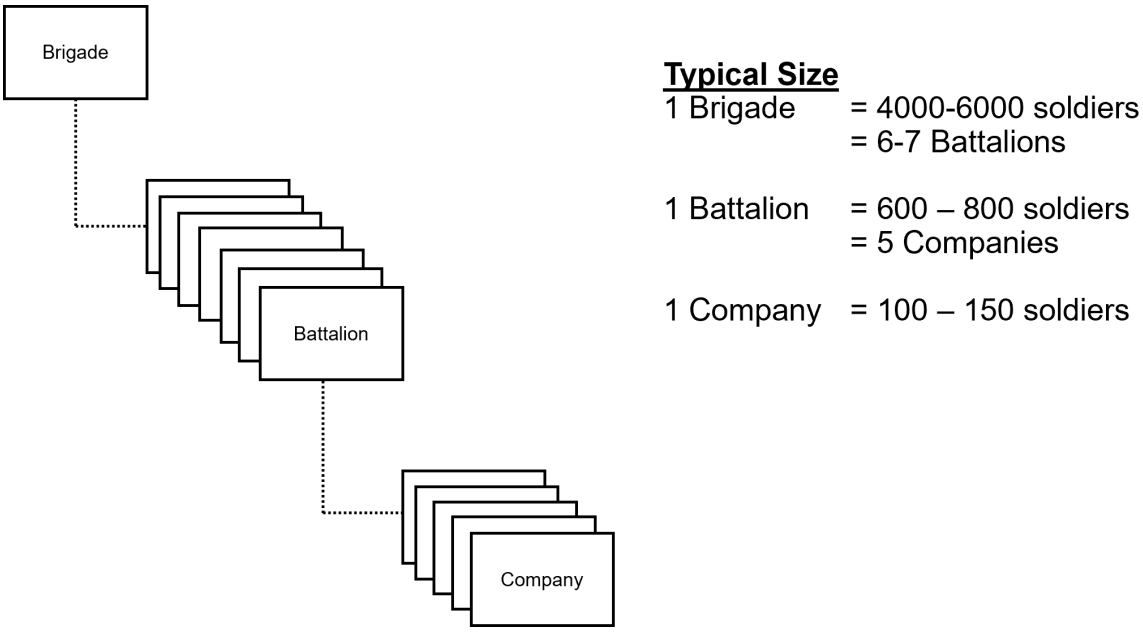
Table 4: Addressing Potential Alternative Explanations

	Outcome: Discretionary Test					
	(1)	(2)	(3)	(4)	(5)	(6)
	All	*Spell FE*	*Close # Soldiers*	*Large Battalions*	*Close Failure Rts*	*Higher NonBlack Failure Rt*
Black Soldier=1	0.081***	0.084***	0.077***	0.094***	0.081***	0.102***
	(0.007)	(0.008)	(0.016)	(0.010)	(0.013)	(0.018)
N	2530976	2530976	609624	1586271	633094	923936
Outcome Mean*100	0.098	0.098	0.098	0.097	0.077	0.099
NonBlack Soldier Mean*100	0.083	0.083	0.083	0.082	0.063	0.088
% Sample Mean	83	86	79	97	106	104
% NonBlack Soldier Mean	98	100	92	114	130	116
Occupation & Time FE	Y	Y	Y	Y	Y	Y
Soldier Controls	Y	Y	Y	Y	Y	Y
Spell-FE		Y		Y	Y	Y

Notes: Column 1 replicates Column 4 from Table 3 for reference. Column 2 includes fixed effects for battalion commander spells to address concerns that non-random assignment of soldiers to battalions might influence results. Column 3 restricts the sample to battalions where the number of Black and non-Black soldiers are similar and controls for quantiles of battalion-race group size, addressing potential measurement error due to unequal numbers of soldiers tested. Column 4 restricts the sample to battalions with more than the median number of total soldiers to assess whether results are driven by measurement error in smaller battalions. Columns 5 and 6 examine discretionary drug testing rates in scenarios where Black and non-Black soldiers have similar drug failure rates or where non-Black soldiers have higher failure rates, to further explore potential biases in discretionary testing. Robust standard errors are reported in parentheses. Coefficients are multiplied by 100 for ease of interpretation. "(CDR)x(Race)" and"(CDR)x(Race)x(Occ)x(Enlist Yr)" random test failure rate controls are included in each column.

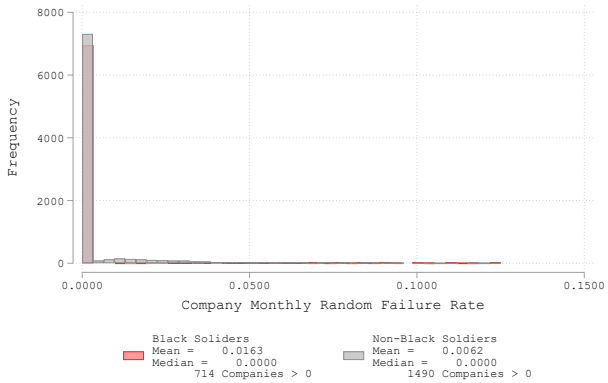
Appendix Results

Figure A1: Army Unit Hierarchy

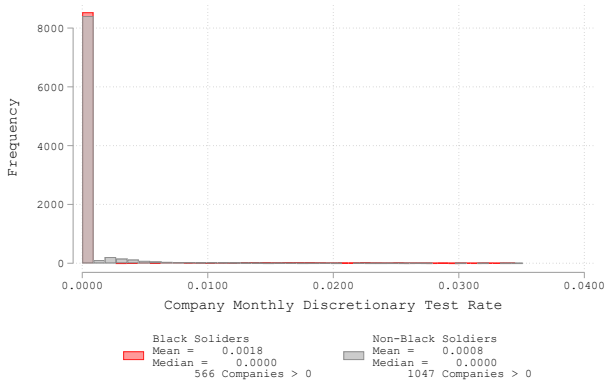


Notes: This figure displays the typical hierarchy and organization of units within an Army Brigade. Drug testing is managed and overseen within the Army at the battalion level, and implemented at the company level.

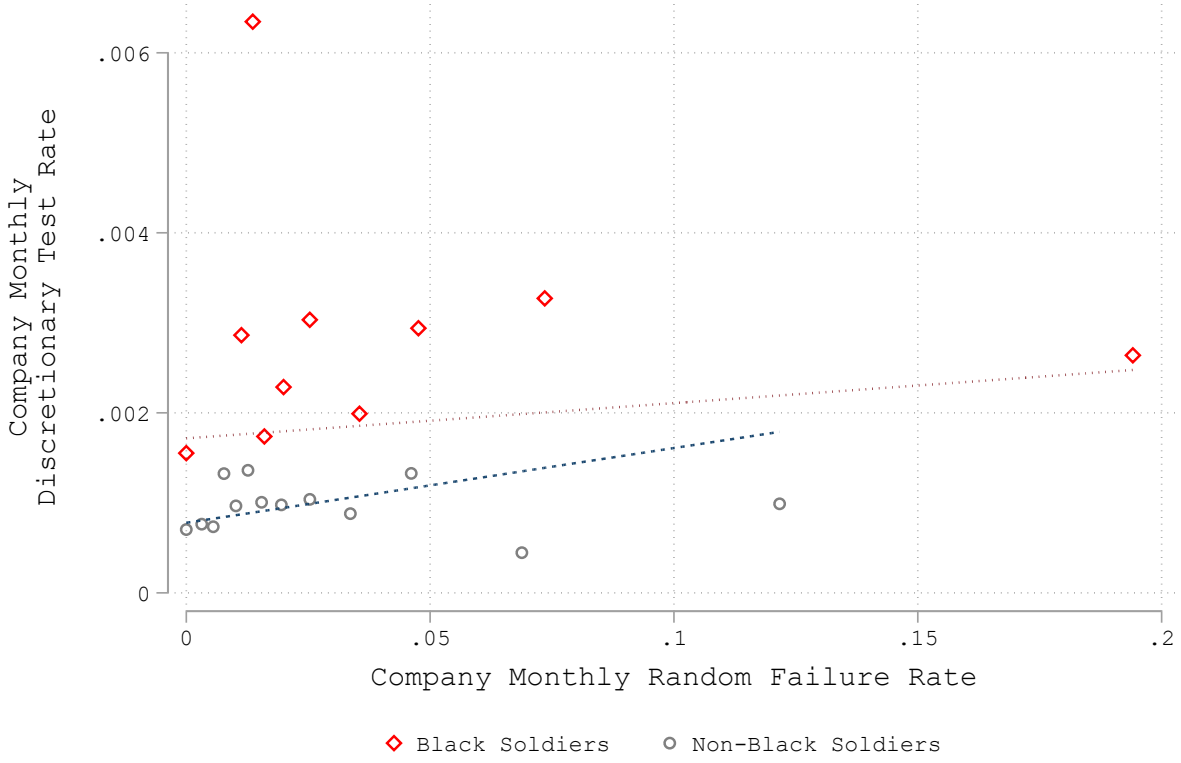
Figure A2: Company Commander-Level Results



(a) Random Rates, Full Sample



(b) Discretionary Rates, Full Sample



(c) Binscatter of Company Level Results.

Notes: Random failure and discretionary testing rates are calculated at the company level, rather than at the battalion spell level.

Table A1: Sample Restrictions

Restriction	Entries Lost	Remaining (Soldier)x(File-Month) Entries	Remaining Unique Soldiers
1. Drop battalion commander-spell if spell length is less than 12 months.	2,326,314	30,914,292	819,159
2. Drop spell if share of soldiers deployed within the spell > 0.5.	57,044	30,857,248	818,757
3. Drop non-MTOE units.	7,160,218	23,667,030	684,521
4. Drop non-first-term soldiers.	14,277,145	9,389,885	460,991
5. Drop anyone who is initially married or has dependents.	1,279,853	8,110,032	396,845
6. Drop anyone once they become married or have dependents.	1,948,729	6,161,303	347,448
7. Drop entries post-previous positive urinalysis (UA).	75,562	6,085,741	346,865
8. Drop entries post-CID charge of any type.	386,818	5,698,923	344,324
9. Drop if missing unit.	1,411	5,697,512	344,135
10. Drop if deployed soldier.	388,216	5,309,296	340,648
11. Drop if battalion commander is unidentifiable.	1,260,448	4,048,848	278,124
12. Drop battalion commander-spell if average monthly battalion size is not within $[100, 2000]$ soldiers.	1,260,101	2,788,747	191,974
13. Drop female soldiers.	257,753	2,530,994	172,350

Notes: We begin with the universe of all first-term, initial-entry enlisted soldiers in the Army between January 2015 and April 2022. This is a sample of approximately 33.2 million (soldier)-x(file-month) entries, representative of approximately 850,000 unique soldiers. An MTOE unit is a training or support unit without a combat mission First-term soldiers include enlisted personnel with 3-year or 4-year terms with less than 3 years of service.

Table A2: Soldier Level Demographics

	(1) *All Enlisted Males, 10/2018*	(2) *First-Term Enlisted Males, 10/2018*	(3) *Analysis Sample*
Share Black	0.206	0.197	0.174
Share White	0.560	0.553	0.589
Share Hispanic	0.166	0.183	0.174
Share Non-Black	0.794	0.803	0.827
Share HSG/GED	0.758	0.896	0.933
Share BA+	0.084	0.057	0.033
Share Moral Waiver	0.033	0.026	0.017
Initial AFQT	59.520	59.396	58.075
Average Age	27.060	21.950	20.729
Unique Soldiers	321206	132046	172339

Notes: Column (1) includes all enlisted male soldiers in the Army as of October 2018. Column (2) restricts the sample to first-term enlisted males, defined as those with less than three years of service. Column (3) presents our analysis sample, which is further refined from the first-term enlisted male sample based on the restrictions outlined in Table A1.

Table A3: Battalion Commander Spell Level Summary Statistics

Share Female	0.04
Share Black	0.09
Share Nonblack	0.91
Share BA+	1
Share Married	0.96
Average Age	41.12
Average # of Dependents	3.20
Average Years of Service	18.41
Average Tenure as BNCDR	22.78
Ever Promoted to COL	0.56
Share in Combat Arms	0.91
Unique Battalions	264
Unique BNCDRs	776
Unique BNCDR-Spells	782

Notes: This table lists demographic information for the battalion commanders (BNCDRs) in our analysis sample. Battalion commanders are typically of the rank of lieutenant colonel (LTC) during service. The outcome "Ever Promoted to Colonel (COL)" measures future promotion in rank and continued tenure in the Army. Combat arms is inclusive of infantry, armor, field artillery, engineering, air defense artillery, and aviation branches.

Table A4: Soldier-Level Testing Statistics *by Drug Type and Detailed Race*

Black Random Failure Rate	0.0135
Nonblack Random Failure Rate	0.0063
Black Discretionary Test Rate	0.0017
Nonblack Discretionary Test Rate	0.0008
White Random Failure Rate	0.0064
Black Random Failure Rate	0.0135
Hispanic Random Failure Rate	0.0059
White Discretionary Test Rate	0.0008
Black Discretionary Test Rate	0.0017
Hispanic Discretionary Test Rate	0.0009
Black Marijuana Failure Rate	0.0105
Nonblack Marijuana Failure Rate	0.0031
Black Cocaine Failure Rate	0.0017
Nonblack Cocaine Failure Rate	0.0017
Black Other Drug Failure Rate	0.0016
Nonblack Other Drug Failure Rate	0.0018
Soldier Months	2530976
Unique Soldiers	172339
Battalion Commander Spells	782

Notes: This table lists random failure and discretionary testing rates by demographic group and drug type from the analysis sample as described in Column (3) of Table A2. Random failure rates are calculated as the total number of positive results from random drug tests divided by the total number of random tests administered within each battalion commander spell.

Table A5: *Robustness:* Aggregate Results Omitting CDR Spell-Race-MOS-Entry-Year Observations with No Random Testing

	Outcome: Discretionary Test				
	(1)	(2)	(3)	(4)	(5)
Black Soldier=1	0.092***	0.083***	0.083***	0.078***	0.082***
	(0.007)	(0.008)	(0.008)	(0.008)	(0.009)
(CDR) x (Race) Random Fail Rate		1.130***	0.870***	0.841***	1.129***
		(0.307)	(0.310)	(0.310)	(0.404)
(CDR) x (Race) x (Occ) x (Enlist Yr) Random Fail Rate			0.261***	0.257***	0.244**
			(0.088)	(0.088)	(0.116)
((CDR) x (Race) Random Fail Rate)*Black Soldier					-0.456
					(0.583)
((CDR) x (Race) x (Occ) x (Enlist Yr) Rand Fail Rt)*Black Soldier					0.022
					(0.172)
N	2435748	2435748	2435748	2435748	2435748
Outcome Mean*100	0.098	0.098	0.098	0.098	0.098
NonBlack Soldier Mean*100	0.083	0.083	0.083	0.083	0.083
% Sample Mean	94	85	85	80	84
% NonBlack Soldier Mean	110	99	99	93	98
(Occ)x(Enlist Yr) FE, Month FE	Y	Y	Y	Y	Y
Soldier Controls				Y	Y

Notes: This table reports results analogous to the results reported in Table 3, except that *we drop soldier-year observations for instances where there are no IR tests within a combination of CDR Spell x Race x MOS x Entry-Year*. Robust standard errors are reported in parentheses. Coefficients are multiplied by 100 for ease of interpretation. MOS (Military Occupational Specialty) refers to a specific job or role that a soldier is trained for.

Table A6: *Robustness:* Aggregate Results with Company Commander Level Controls

	Outcome: Discretionary Test				
	(1)	(2)	(3)	(4)	(5)
Black Soldier=1	0.095***	0.094***	0.092***	0.087***	0.090***
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)
Company Monthly Random Failure Rate		0.437***	0.376***	0.372***	0.613***
		(0.116)	(0.113)	(0.113)	(0.187)
(CDR) x (Race) x (Occ) x (Enlist Yr) Random Fail Rate			0.284***	0.280***	0.282**
			(0.091)	(0.091)	(0.113)
(Company x Race Rand Fail Rt)*Black Soldier					-0.354
					(0.236)
((CDR) x (Race) x (Occ) x (Enlist Yr) Rand Fail Rt)*Black Soldier					-0.002
					(0.176)
N	2530976	2421413	2421413	2421413	2421413
Outcome Mean*100	0.098	0.098	0.098	0.098	0.098
NonBlack Soldier Mean*100	0.083	0.083	0.083	0.083	0.083
% Sample Mean	97	96	94	89	92
% NonBlack Soldier Mean	114	112	110	104	108
(Occ)x(Enlist Yr) FE, Month FE	Y	Y	Y	Y	Y
Soldier Controls				Y	Y

Notes: This table reports results analogous to the results reported in Table 3, except we estimate the Random Failure Rate at the Company Commander- Race level. Robust standard errors are reported in parentheses. Coefficients are multiplied by 100 for ease of interpretation.

Table A7: Empirical Bayes Estimates of Disparate Testing against Black Soldiers

	Estimate	Std. err.	P-value	Q-value $\lambda = 0.05$	Q-value $\lambda = 0.1$	Q-value $\lambda = 0.2$	Linear shrinkage	Non-par. post. mean
Brigade	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
BDE 2	0.130	0.032	0.000	0.001	0.001	0.000	0.119	0.070
BDE 8	0.175	0.045	0.000	0.001	0.001	0.000	0.131	0.104
BDE 9	0.225	0.054	0.000	0.002	0.001	0.001	0.140	0.127
BDE 16	0.145	0.047	0.004	0.009	0.007	0.003	0.120	0.103
BDE 22	0.122	0.042	0.004	0.011	0.008	0.004	0.113	0.097
BDE 4	0.154	0.066	0.015	0.011	0.009	0.004	0.117	0.125
BDE 3	0.115	0.047	0.011	0.012	0.009	0.004	0.110	0.114
BDE 5	0.177	0.064	0.010	0.012	0.009	0.004	0.122	0.137
BDE 17	0.198	0.083	0.015	0.012	0.009	0.004	0.120	0.174
BDE 33	0.171	0.070	0.014	0.013	0.010	0.004	0.119	0.154
BDE 13	0.192	0.075	0.010	0.013	0.010	0.004	0.121	0.144
BDE 23	0.079	0.029	0.007	0.013	0.010	0.004	0.092	0.061
BDE 18	0.144	0.061	0.013	0.013	0.010	0.004	0.116	0.127
BDE 19	0.091	0.042	0.019	0.013	0.010	0.004	0.101	0.081
BDE 31	0.094	0.042	0.018	0.013	0.010	0.004	0.102	0.087
BDE 21	0.221	0.081	0.008	0.013	0.010	0.004	0.123	0.177
BDE 14	0.063	0.025	0.010	0.014	0.011	0.005	0.079	0.066
BDE 6	0.095	0.048	0.031	0.019	0.015	0.006	0.103	0.096
BDE 11	0.105	0.054	0.034	0.020	0.015	0.006	0.107	0.101
BDE 15	0.177	0.089	0.031	0.020	0.015	0.006	0.116	0.174
BDE 7	0.101	0.052	0.038	0.021	0.016	0.007	0.106	0.108
BDE 12	0.126	0.071	0.046	0.024	0.019	0.008	0.111	0.146
BDE 29	0.088	0.055	0.062	0.031	0.024	0.010	0.102	0.107
BDE 1	0.057	0.039	0.079	0.038	0.029	0.012	0.086	0.076
BDE 25	0.074	0.056	0.102	0.046	0.035	0.015	0.099	0.114
BDE 26	0.046	0.035	0.099	0.046	0.035	0.015	0.078	0.063
BDE 27	0.054	0.046	0.126	0.054	0.041	0.017	0.089	0.093
BDE 28	0.044	0.040	0.140	0.058	0.044	0.019	0.082	0.080
BDE 10	0.027	0.030	0.188	0.073	0.056	0.023	0.063	0.056
BDE 24	0.022	0.024	0.183	0.073	0.056	0.024	0.052	0.044
BDE 32	0.019	0.027	0.243	0.091	0.070	0.029	0.055	0.053
BDE 30	0.008	0.018	0.326	0.118	0.090	0.038	0.031	0.033
BDE 20	0.004	0.028	0.444	0.156	0.120	0.050	0.046	0.044

Notes: This table presents brigade-level estimates of racial disparities in discretionary drug testing, constructed as the mean of battalion-level estimates weighted by soldier-month observations. The first two columns report raw estimates and standard errors, while the third column provides p-values from one-tailed hypothesis tests for the null hypothesis that brigades do not disproportionately test Black soldiers. Subsequent columns report q-values for multiple hypothesis correction, and the final two columns show posterior estimates under linear and nonparametric shrinkage. Linear shrinkage adjusts estimates by pulling them toward the overall mean based on the ratio of signal to noise, reducing estimation variance while preserving systematic differences. Nonparametric shrinkage applies a data-driven approach that does not impose a functional form, allowing for greater flexibility in capturing heterogeneity across brigades. Brigade numbers do not correspond to real brigade identities.