A Noisy Signal, a Consequential Decision: Observational and Experimental Evidence on Criminal Background Checks *

Amanda Agan[†] David Autor[‡] Emma Rackstraw[§]

October 28, 2026

Abstract

Criminal background checks are a standard feature of hiring in the United States, yet little evidence exists on which record characteristics shape screening decisions or whether those characteristics predict subsequent worker outcomes. We partner with a large online services marketplace and a background check company to link applicants' criminal records to screening decisions, as well as outcomes on the marketplace such as client engagement and incidents. Visible record characteristics explain about half of the variation in allowance decisions, with felony status, charge type, and recency strongly associated with exclusion. Among allowed applicants, those with criminal records engage less with clients and have higher incident rates than those without records. However, detailed record characteristics have little predictive power for engagement or incidents among allowed applicants. Because these outcomes are observed only for applicants who pass screening, we implement a randomized experiment that hides selected criminal charges, increasing allowance by 3.5 percentage points among applicants with records. We estimate outcomes for applicants allowed because of the experiment and find no statistically detectable differences in engagement or incident rates relative to applicants with records whom the marketplace already allows, although the estimates are imprecise. Together, these findings suggest that the record details strongly associated with exclusion do not reliably distinguish worse performers among allowed applicants or at the admission margin changed by the experiment.

*Thanks to Austin Lentsch, Lauren Qu, and Can Yesildere for invaluable research assistance. Thanks also to the employees of the online market place company and the background check company who provided us with and helped us understand these data and were integral at implementing the experiment described here. The authors are grateful for feedback from Max Kapustin, Doug Miller, Ezra Oberfield, and seminar participants at Opportunity Insights, University of Toronto, APPAM, and the Minneapolis Fed. The RCT was pre-registered on the AEA RCT Registry: AEARCTR-0010917. IRB approval was obtained from MIT: COUHES Protocol 2302000885.

[†]Cornell University & NBER

[‡]MIT & NBER

[§]Swarthmore College

1 Introduction

Criminal background checks are one of the most common forms of labor market screening in the United States.[1] These background checks potentially affect a large share of U.S. workers: estimates imply one-third of U.S. adults have been arrested over one-third of working-age adults have been arrested (Goggins and DeBacco, 2024) and roughly 8% of US adults, and 33% of Black males, have a felony conviction record (Shannon et al., 2017). Firms use criminal background checks for reasons that are easy to understand: past criminal charges may contain information about future reliability, productivity, client safety, or reputational risk. At the same time, criminal records are complicated objects that may make their signals difficult to interpret as they can contain a bundle of charges that may each differ by severity, type, disposition, age, intensity, and legal visibility. Firms, however, must turn this high-dimensional object into a binary decision of whether to allow the applicant to the next stage or exclude them due to their criminal record.

In this paper, we study how a particular firm uses criminal records to make screening decisions, and whether the record information that shapes those decisions predicts the worker outcomes firms care about. Background check screening creates a real trade-off. Excluding applicants with concerning records may help firms avoid costly matches, protect clients, and reduce reputational or liability risk. But exclusion can also cause firms to forgo workers who would have been productive and safe. We ask (1) which aspects of a criminal record are associated with exclusion (e.g. charge type, severity, recency, number of charges)?; (2) whether these record characteristics predict outcomes the firm cares about among allowed applicants; and (3) how allowing applicants who would otherwise be excluded affects worker activity and incident risk. To address the third question, we conduct an experiment that randomly hides some charges during screening. Our data provide measures of incident risk and productivity for allowed applicants, but cannot capture reputational or liability costs. For applicants with criminal records, these decisions can limit access to work, and because employment opportunities may reduce subsequent criminal activity (Schnepel, 2018; Yang, 2017), the consequences of screening can extend beyond the firm-worker match.

We study these questions in partnership with a large online services marketplace company and a background check company. The marketplace screens all prospective service providers, allowing us to observe allowance/exclusion decisions for the entire applicant pool. Through

[1] In a 2021 survey, 90 percent of firms reported conducting background checks on all full-time employees and 59 percent reported conducting background checks on all contractors (Professional Background Screening Association, 2021). A survey of people with hiring experience in the US shows that 70% say their firm did background checks on all entry-level workers (Agan et al., Forthcoming). ETC.

this partnership, we are able to observe detailed criminal background check information on all applicants, including their full criminal records and the record information available to the firm during screening, whether the applicant was excluded from the marketplace due to their criminal record, and marketplace outcomes for allowed applicants for 12 months, including client contacts, reviews, ratings, and reported incidents.

For new applicants, subsequent marketplace outcomes are observed only for those allowed onto the marketplace, so the data cannot show what would have happened to excluded applicants had they been allowed. Associations between record details and outcomes among allowed applicants may not reflect how those details relate to potential outcomes across the full applicant pool, because the firm has already selected whom to allow. To understand the consequences of excluding applicants, we therefore need to learn about the outcomes of applicants who would otherwise be excluded. Through this partnership we implement a randomized experiment that reduces the visibility of selected criminal charges for a random subset of applicants. The adjudicators making screening decisions were unaware of the experiment and treatment assignments. This intervention effectively allows some new applicants with records access to the marketplace who would otherwise have been turned away, allowing us to study the outcomes of service providers at the margin of exclusion affected by the experiment.

Our results show that the firm's criminal record adjudicators screen systematically on criminal record characteristics: 45% of workers with a visible record are excluded, and that rate is much higher among applicants with felony, property, or violent charges, more recent records, or more charges. Visible record characteristics explain about half of the variation in allowance decisions. Allowed applicants with any legally reportable criminal record are 25% less likely to contact a client, make 23% fewer contacts, and have a 33% higher chance of having an incident (over a baseline of 1.8%) than those without a reportable record. However, record details have little predictive power for outcomes among allowed applicants. For example, we find little evidence that applicants with felony convictions have higher incident rates or lower client engagement than those with only misdmeanors even though those with felonies are much less likely to be allowed. Analyzing the results of our experiment, we find no detectable difference in incident rates between providers allowed due to the experiment and applicants with records whom the marketplace was already allowing. Marginally allowed providers also engage with clients, at rates we cannot distinguish from those of already-allowed providers with records, although these estimates are imprecise. This pattern suggests that the firm's screening rule may exclude some applicants who would add activity to the marketplace without evidence of greater incident risk than providers with

records whom it already allows.

At the margin, allowing applicants with the types of criminal records the experiment allows onto the marketplace does not measurably increase incidents; and on average, the detailed features of a criminal record carry little information about outcomes among those allowed. There are at least two possible interpretations for these findings. One is that the screen adjudicators are using may be too strict for the measured outcomes the marketplace cares about, and for incidents in particular. This may make sense practically: criminal records are high-dimensional signals that can be hard to interpret, and limited feedback about allowed workers, together with the absence of outcomes for excluded applicants, may make it difficult for adjudicators to update or recalibrate their screening thresholds. Another interpretation is that the firm is screening for costs we do not observe well, which could include reputational risk or client preferences, including discriminatory preferences. The experiment may also lack the statistical power to detect changes in rare but costly tail-incident events. These explanations are not necessarily mutually exclusive. Because the marketplace faces no capacity constraint on applicants allowed onto the marketplace, relaxing some criminal record based exclusions would add providers who engage with clients without requiring the exclusion of other applicants, while giving more people with records access to work, with potential social benefits beyond the marketplace.

As complementary evidence, we study incumbent service providers who joined the marketplace before background checks were required and completed a check during our sample period. For these providers, we observe activity before background check screening, including for some who are subsequently excluded. In this sample also detailed record characteristics have little association with prior engagement or reported incidents, and we find no statistically detectable differences in prior contact or incident rates between incumbents subsequently allowed and excluded. Although these incumbents are a selected group, their pre-screening outcomes provide an additional perspective on the relationship between criminal records and marketplace outcomes.

A large literature documents correlations between criminal records and subsequent outcomes, including recidivism (Blumstein and Nakamura, 2009; Bushway, 2024), non-employment (Mueller-Smith and Schnepel, 2021; Rose, 2021; Agan et al., Forthcoming), and employer callbacks (Pager, 2003; Uggen et al., 2014; Agan and Starr, 2018). Employers also report substantial reluctance to hire individuals with records (Holzer et al., 2006; Cullen et al., 2022). A separate literature examines whether criminal characteristics predict subsequent criminal justice contact. Prior work shows that risk declines sharply with time since last

criminal contact and eventually converges toward population averages (Blumstein and Nakamura, 2009; Bushway et al., 2009; Kurlychek et al., 2012). But recidivism is not the only decision relevant outcome for employers. Firms also care about productivity, reliability, and client- or coworker-facing risks, and there is comparatively little evidence linking criminal histories to these on-the-job outcomes. Additionally, firms worry about very rare but salient tail risks like negligent hiring lawsuits.[2]

A smaller literature examines how criminal record characteristics are related to actual screening decisions. Kurlychek et al. (2019) use data on applicants provisionally hired for non-licensed health-care jobs in New York state and find that the number, severity, type, and recency of convictions are associated with the initial decision to allow an applicant to work. Weiss et al. (2025) use applicant data from one employer to study how applicants move from a conditional offer of employment to being hired by offense type and severity. We contribute to this literature by studying how criminal record characteristics relate to screening decisions in a different environment and whether those characteristics also predict subsequent outcomes.

Existing evidence on on-the-job performance of people with records is limited and largely observational. Lundquist et al. (2018) study US military enlistees and find that those with felony records are no more likely to attrit and, if anything, are promoted more quickly than those without. Minor et al. (2018) use data from a call center and find that workers with records are less likely to quit and have similar or slightly better performance. While informative, these studies focus on selected samples of hired workers and rely on comparisons conditional on observables, leaving open how criminal records relate to outcomes at the margin of hiring decisions. Neither has information on applicants who applied but were rejected, and thus cannot study the rejection margin. Our experiment complements this evidence by also revealing outcomes for marginal applicants who would otherwise have been excluded by the firm's criminal record screening process.

The study closest to ours, Denver et al. (2017), examines applicants provisionally hired for non-licensed healthcare jobs in New York State, where applicants initially denied because of their criminal record can submit evidence of rehabilitation before a final decision. Using the initial decision as an instrument for the final one, they find that being allowed to work reduces subsequent arrests, particularly for men. They cannot, however, observe performance or harm on the job.

[2]While Douglas and Maltby (2023) find only 435 cases of employers liable for negligent hiring because they employed someone with a conviction over a 50-year period, Pyle (2023) finds that reducing firms' exposure to negligent hiring legal liability raises the hiring of workers with criminal records.

Our paper also relates to the personnel economics literature on screening applicants and evaluating hiring criteria (Autor and Scarborough, 2008; Hoffman et al., 2018). A central difficulty is that job performance is observed only for applicants who are selected under existing hiring rules (Brown, 1982) (a problem that arises in evaluating decisions outside of employment as well, e.g. Kleinberg et al. (2018)). Thus relationships between screening criteria and outcomes among hired workers do not necessarily reveal how excluded applicants would have performed. Like work that experimentally extends employment opportunities to reveal unobserved worker quality (Pallais, 2014), our experiment allows us to estimate performance for applicants who would otherwise be excluded. In our setting, this allows us to connect criminal record information used in screening to performance and risk outcomes, at least at the margin of the firm's existing screening policy.

Our paper contributes to this literature in three ways. First, we provide experimental evidence on the consequences of relaxing criminal record screening thresholds, allowing us to isolate outcomes for marginally allowed workers rather than relying on comparisons among observationally selected workers. Second, we study a setting with rich measures of on-the-job performance, including engagement, productivity, and safety outcomes that are directly relevant to firm decision making. Third, because the intervention operates through the firm's existing adjudication rules, we are able to link specific features of criminal records, such as charge type and time since offense, to both hiring decisions and subsequent outcomes. In addition to examining their subsequent outcomes, we characterize the criminal records of the applicants whose allowance changes under the experiment. In the spirit of Gruber et al. (1999), this allows us to understand what types of applicants with records are at the margin of the firm's screening policy and how their records compare with those of applicants the firm already allows.

2 Institutional Setting

We partner with an online marketplace company and an external background check company. Through this partnership, we have data all background checks run for the marketplace company between 7/23/2023-9/1/2024 (the "study period"), and we have data on marketplace outcomes for allowed applicants until 9/1/2025. More details on this data are provided in Appendix Section D.

2.1 Online Marketplace Company

The online marketplace company we partner with connects service providers with users needing assistance with various short-term, mostly in-person, tasks. Service providers on the marketplace contact potential clients who have posted tasks they need assistance with. Clients then can choose a service provider (or no service provider, or a service provider outside the marketplace). Similar to many online platforms and marketplaces, clients can leave a review for the service provider, which includes a 1-5 star rating and potentially a short written assessment. If they have any problem with the services provided (such as incomplete or shoddy work, or a safety issue), they can report an incident to the marketplace company, which the company will investigate and may take action on if substantiated (such as removing a service provider from the marketplace).

2.2 The Expansion of Background Check Requirements

The background checks conducted during our study period come from two distinct populations. For new entrants, screening occurs at the beginning of their marketplace participation and determines whether they have continued access to the marketplace. For incumbent service providers already on the marketplace at the time our study began, screening may have occurred after a potentially substantial period of marketplace activity and determines whether they may remain active. We analyze these groups separately because the timing, selection, and interpretation of the screening decision differ substantially across them.

In our study period, all newly registering service providers were required to consent to a background check as part of the onboarding process. Service providers could begin contacting clients while the background check was processing, but a service provider who did not pass the check would subsequently lose access to new client opportunities.[3]

During our study period, the marketplace was also extending the mandatory background check requirement in waves to incumbent service providers who had already joined under earlier background check screening rules. These service providers were already on the marketplace offering services without having to undergo a mandatory background check. Incumbents who did not pass a check by their deadline were placed on hold and could no longer contact new clients.

Before screening became mandatory for a given incumbent cohort, some service providers could voluntarily complete a background check. Service providers who were allowed on

[3]Processing by a human adjudicator for a report with visible criminal charges takes on average 9 days, with a median of 8 days.

under the voluntary check received a visible background check badge, while those who were excluded were removed from the marketplace. As a result, incumbents who first complete a check during our study period are a selected group: they entered under earlier screening rules, remained active until the mandatory rollout reached them, and had generally not previously completed a voluntary check. We examine these incumbents – both those subject to a mandatory check and those who volunteer for it – in Section 7.

2.3 The Background Check Process

The marketplace uses an external consumer reporting agency to conduct court record searches. Once a background check is completed, the resulting record is processed using a combination of rule-based filtering and, when necessary, human review. For new entrants, the resulting adjudication determines whether they may continue contacting clients after onboarding. For incumbent service providers, it determines whether they may retain access to new client opportunities.

Because background check companies are subject to the Fair Credit Reporting Act (FCRA), they must inform job candidates whenever information from their criminal background is used to make decisions that adversely affect them. The goal is to allow job candidates an opportunity to flag inaccurate information that appeared on their record or to appeal the decision by providing information about extenuating circumstances.[4] This notice is called an "adverse action" notice.

Manually reviewing high volumes of background checks is a time-intensive process. To codify and streamline their review processes, the marketplace company has pre-specified the relevance of each charge. For categories, dispositions, and classifications they find irrelevant, the consumer reporting agency then removes these charges from new background checks before sending those background check results to that firm's adjudication team. This lightens the burden of costly manual review and allows adjudicators to focus their discretion on the remaining records.

Thus, for our marketplace company and criminal background check agency partners, the standard adjudication process proceeds as follows (described visually in Figure 1):

1. The background check company retrieves the applicant's full criminal record (and applies any relevant state and federal rules that regulate the reporting of different charges, e.g. FCRA).

[4]Criminal histories frequently include information fuzzily matched to an individual through inaccurate entity matching (Lageson and Stewart 2023).

2. A charge visibility ruleset, chosen by the marketplace company, is applied by the background check company and determines which charges are visible.

3. If no visible charges remain, the applicant is automatically passed through the screening process ("allowed" on the marketplace) with no human intervention.

4. If visible charges remain, a human adjudicator reviews the record and either allows the applicant on the marketplace or issues a pre-adverse action notice. Federal guidance suggests waiting five business days for a response before continuing to the post-adverse action stage.[5]

The charge visibility ruleset is based on crime type (one of 10 charge categories and 55 charge subcategories), level of crime (felony, misdemeanor, or petty offense), and years since the offense. The full pre-experimental ruleset this marketplace company implemented is described in Appendix A.

The firm has a small cadre of part-time adjudicators whose only job for the firm is to make decisions on criminal records. Adjudicators face no binding capacity constraint and have no additional information about the applicant besides their name, date of birth, location, and visible criminal record history. Conceptually, they allow applicants if expected benefits exceed expected costs.

Marketplace companies benefit from having additional individuals providing services on the marketplace because this can lead to more engagement with potential clients, and potentially higher client satisfaction based on the quality of the applicant. At the same time, allowing applicants who do not perform well or cause incidents could lead to harm in terms of fewer clients or reputational risks.

Throughout the paper, we describe a service provider as "allowed" if the background-check process does not result in a final adverse action and the service provider is therefore permitted to contact new clients after adjudication.

2.4 Experimental Change to the Visibility Ruleset

Between July 24, 2023 and September 1, 2024, all applicants who completed a background check were randomly assigned to:

- **Control ruleset**: Status quo charge visibility rules.

[5]It is rare for candidates to reverse an adverse action process in this five-day period. In our data, fewer than 5% manage this.

- **Treatment ruleset**: Additional categories of charges hidden.

The rulesets differ in their look-back windows: the number of years a given type of charge remains visible to adjudicators, where a window of zero means the charge is never shown. The treatment ruleset hides every charge the control ruleset hides, plus additional charges. Treatment shortens the look-back window for selected misdemeanors (mostly vehicle and traffic charges and DUIs) and for felony drug and alcohol charges. The most common types of charges made invisible by the treatment ruleset are car-related misdemeanors: most are license & registration and unsafe operation charges, followed by DUIs, speeding, and vehicle equipment charges. Table 1 lists these changes, and Appendix A reports the full control ruleset.

3 Data

The data combine charge level criminal background check records from the background check company with service provider level activity and outcomes from the online marketplace. The background-check data include all checks processed during our sample period: July 24, 2023 through September 1, 2024. The marketplace data allow us to link each check to the service provider's signup date, screening outcome, and marketplace activity. Because the service providers completing checks during this period include both newly registering service providers and incumbent service providers who had already been active on the marketplace, we define our analysis samples separately below. The provider level activity and outcome data on the marketplace include data from the sample period followed by 1 year of outcome data for any applicant newly allowed onto the marketplace (the new entrant sample).[6]

3.1 Criminal Background Check Data

The background check company provides the applicant's full legally reportable criminal record. By "full record," we mean all charges that the background check company was legally permitted to report after applying the Fair Credit Reporting Act (FCRA) and relevant state and local reporting restrictions, but before applying the marketplace company's own charge visibility rules.[7]

[6]Incumbent service providers will also have earlier data available dating back to when they joined the marketplace. This provides a complementary source of evidence on how service providers with criminal records performed during a period in which they were already operating on the marketplace but had not yet completed the subsequently required check. We examine this data separately from the post-background-check period in Section 7.

[7]See Agan et al. (Forthcoming) for additional detail on the legal regulation of reportable criminal records.

For each legally reportable charge, we observe its disposition, such as conviction, pending, alternative adjudication, or dismissal; the date of the charge; its crime type, classified into 10 broad categories and 55 subcategories[8]; and its level, classified as a felony, misdemeanor, or petty offense. We also observe whether the charge was hidden or visible under the applicant's assigned visibility ruleset and the ultimate allowance outcome, including whether the service provider received an adverse-action notice.

Because we know the rules of both the control and treatment visibility rulesets, we can reconstruct the record that would have been shown to adjudicators under either assignment. Thus, for each service provider, we observe both the full legally reportable record and the counterfactual visible record under each experimental condition.

We do not observe self-reported race or gender, education, or many other characteristics commonly available in labor market data. The marketplace company's human adjudicators see only the background check itself; they are not directly shown race or gender, although they did observe the applicant's name, date of birth, and location along with the visible criminal record. Because the background check company has access to names, they were able to provide us with predicted race and gender based on names and location information.[9] While the prediction is likely to be noisy, given that the background checks nor the marketplace company's application includes race/ethnicity directly, these likely align with the implicit assumptions of the firm's adjudicators about the race/ethnicity of the applicant from their names. The background check company also provides date of birth for each applicant which allows us to calculate their age at the time of the background check.

3.2 Marketplace Activity and Outcomes

The data from the marketplace company allows us to link each background check to the corresponding service provider account on the marketplace. For applicants allowed onto the marketplace, we observe their signup date and marketplace activity, including the dates on which the service provider contacted potential clients, reviews and star ratings received,

[8]The 10 broad categories are: violence, vehicles & traffic, theft & property, statutory, sexual, security, homicide, fraud & deception, drugs & alcohol, criminal intent

[9]At the research team's request, the background check company predicted race and gender using open-source algorithms based on historical national-level data for the sole purpose of this research and for sole use by the research team. Predicted race uses last name, state, and county (by three-digit county FIPS code) as inputs to the algorithm (Imai et al. 2024). Predicted gender uses first name and year of birth (Mullen 2021). We then assign race and gender probabilistically, where each candidate is allocated to the (race/gender) category for which they have the highest predicted probability of being correctly assigned. Note that the availability of predictions at the margin are predicated on the quality of these inputs; for example, gender predictions could not be produced for candidates reporting single-letter first names or with uncommon names in the U.S. context. These individuals are assigned to an 'unknown' category.

and any incidents reported to the marketplace. For the main analyses of new entrants, we measure outcomes over the 12 months following completion of the background check. The marketplace data extend through September 1, 2025, providing at least 12 months of follow-up for every background check in the sample period.

Our primary engagement outcomes are whether the service provider contacts any potential client and the total number of client contacts. A contact records an instance in which a service provider responds to a client's request for services. We do not observe whether the service provider and client ultimately contract, nor do we observe the price paid. Most payments occur outside the marketplace, and we do not have access to the contents of private communications between service providers and clients.

We also observe whether a service provider receives a review, the review's star rating, and whether the review is verified as originating from activity on the marketplace. Service providers may import reviews for work performed outside the marketplace, so our main review outcomes focus on verified reviews from work originating on the marketplace. Finally, we observe incidents reported to the marketplace and their broad classification; e.g. a complaint of incomplete work, lateness, or a safety violation. The marketplace investigates these reports and may take action, including removing the service provider, when an incident is substantiated.

3.3 Analysis Samples

The background check data contain 253,218 individuals whose background checks were completed between July 24, 2023 and September 1, 2024. We drop data from the final two weeks of December 2023 due to data anomalies we describe further in Appendix D, which leaves us with 246,082 background checks.

This universe combines two distinct service provider populations. The first consists of newly registering service providers who were required to initiate a background check as part of signup. The second consists of incumbent service providers who had joined under earlier screening rules and were later required to complete a background check as the marketplace expanded mandatory screening to existing providers during our sample period. Because the timing and interpretation of screening differ across these populations, we analyze them separately.

We use two main analysis samples:

1. **Status quo entrant sample.** Sections 4 and 5 use newly registering service providers

assigned to the control visibility ruleset. Restricting to the control group allows us to study the marketplace's status quo visibility and adjudication decisions.[10] This sample contains 94,566 service providers, including 13,415 (14.2%) with at least one legally reportable criminal charge and 8,833 (65.84% of those with a reportable record) that were allowed onto the marketplace. We use this sample to describe which features of visible records drive screening decisions and whether the corresponding characteristics of full criminal histories predict outcomes among service providers allowed under that policy.

2. **Experimental entrant sample.** Section 6 uses the same population of newly registering service providers but includes both applicants assigned to treatment and control rulesets. This sample contains 189,519 service providers, of whom 27,039 (14.3%) have at least one legally reportable criminal charge and 18,287 (67.63% of those with a record) were allowed onto the marketplace. The status quo entrant sample is therefore nested within the experimental entrant sample. Comparing applicants assigned to the treatment and control visibility rules allows us to estimate how reducing the visibility of selected charges affects allowance and subsequent marketplace outcomes. We also use the randomization to characterize the applicants whose allowance changes due to treatment (marginal applicants) and estimate their outcomes if allowed.

And a supplemental sample analyzed in Section 7:

1. **Incumbent sample.** Section 7 uses 56,766 service providers who had joined the marketplace before a background check was required for their cohort and who subsequently completed a check during the sample period. Of these service providers, 7,535 (13%) have at least one legally reportable criminal charge and 1,964 (26% of those with a record) end up excluded after their eventual background check. These service providers may have accumulated substantial marketplace activity before screening, allowing us to examine how the criminal histories revealed by their later background checks relate to their earlier activity and reported incidents on the marketplace.

The incumbent sample is selected in ways that differ from the entrant sample. The marketplace company introduced screening to incumbent service providers gradually, and a service provider's appearance in our data may depend on the timing of that rollout, continued par-

[10]The experiment was going on simultaneously. Firm adjudicators were unaware of the experiment, and they were reviewing applicants assigned to both treatment and control rulesets at the same time. Because only half of applicants were assigned to treatment, and treatment changed the visible record for only a subset of those applicants, we expect any change in the overall mix of records seen by adjudicators to have been difficult to detect.

ticipation on the marketplace, and any earlier voluntary or targeted screening. In particular, service providers who had previously completed a voluntary check and failed would already have been removed. We therefore do not treat the incumbent sample as representative of all service providers who would have entered in the absence of screening. Instead, we use it as complementary evidence from a selected group whose marketplace activity is observed before their criminal histories are adjudicated.

Both the experimental sample and the incumbent sample provide complementary ways to learn about outcomes that are typically unobserved because of the "selective labels" problem: under the status quo screening process, marketplace outcomes are only observed for applicants allowed onto the marketplace (Brown, 1982; Lakkaraju et al., 2017). The experiment newly allows some applicants who would otherwise be excluded, letting us to estimate outcomes for those applicants on the margin of the experiment. For incumbent service providers, we observe activity before their background checks, including for some service providers who are subsequently excluded. Each approach therefore provides evidence on outcomes for some applicants that would be excluded under the status quo decision rules, though the groups and periods we observe differ for each.

3.4 Sample Characteristics and Experimental Balance

Figure 2 shows the monthly volume of background checks in the experimental entrant sample, separately for service providers with and without a legally reportable criminal charge. Application volume is relatively stable over the sample period, with modest month-to-month variation but no clear trend or discontinuity. Service providers with criminal records also constitute a relatively stable share of the entrant sample over time.

Table 2 evaluates experimental balance in the entrant sample. The first panel reports characteristics for all entrants, assigning zeros to criminal-record measures for service providers without a legally reportable charge. The second panel conditions on having at least one charge. Approximately half of service providers were assigned to each visibility ruleset, and average pre-treatment criminal-record and demographic characteristics are similar across treatment and control. Joint tests fail to reject equality of the listed characteristics across experimental arms in either panel.[11]

The balance table also provides summary statistics on the experimental entrant sample. Criminal histories are typically multidimensional. Among entrants with a legally reportable record, service providers have approximately four charges on average, often spread across

[11] Conditional on having a record, the control group is slightly more likely to have a drug charge.

several distinct charge dates. Nearly all have at least one misdemeanor or petty offense, while approximately 30 percent have at least one felony. About 27 percent are always clear under both rulesets, 32 percent become newly clear under treatment, 17 percent retain visible charges but experience reduced visibility, and 24 percent have the same visible record under both rulesets. Thus, the experimental ruleset changes some aspect of the record shown to adjudicators for nearly half of entrants with legally reportable charges.

To illustrate the variation in criminal records and subsequent outcomes among allowed applicants, Table 3 reports 12-month outcomes for different types of applicants with records who were allowed onto the marketplace. The examples span differences in charge type, number of charges, and recency. For each group, the table reports the share of applicants who contact a client, receive a verified review, or have an incident reported to the marketplace. Appendix Table B.1 presents outcomes for a broader set of record groups and additionally distinguishes between records with and without a pending charge.

4 Observational Results: Screening Decisions

Firms that use criminal records in screening applicants face a difficult problem. Criminal records are potentially high-dimensional objects that can hold many charges, of different types and severities, at different points in time relative to the firm's screening decision. The firm has to turn all of this into a single, binary decision: whether to allow an applicant through to the next stage (in this case on the marketplace) or not.

In this section, we ask what features of a criminal record this screening decision appears to be based on in practice. When an adjudicator looks at a visible record, which features lead them to exclude the applicant? We answer this descriptively, using the status quo sample—new entrants who were required to do mandatory background checks on sign up and assigned to the control group. This gives us an idea of how the firm was screening under the status quo. In our setting we have access to applicants' full criminal histories as well as the visibility rules applied to the record and the adjudicator's decision— hence we can see what information the adjudicator knew and how they acted on it.

Table 4 reports first the share of the sample that is allowed on the platform by existence of criminal record and its visibility. Unsurprisingly basically all applicants with no legally reportable criminal records are allowed on. 66% of applicants with a legally reportable criminal record are allowed on, though this comes from a combination of applicants with no visible charges who are always allowed and 53.6% of applicants with at least one least one

visible charge who are allowed. Thus, exclusion is not an automatic consequence of having a criminal record, or even of having charges visible to an adjudicator. Among applicants with visible charges, adjudicators allow slightly more than half onto the marketplace. We next examine which record characteristics distinguish those allowed from those excluded.

We come at this question in multiple ways. We start with regressions of allowance on visible criminal record characteristics and demographics, to get the direction and size of the association between record characteristics and allowance decisions. We then decompose the explained variation in allowance to see how much of the decision each criminal record feature actually accounts for. Finally, we ask how well the visible record predicts allowance out of sample while also allowing flexible interactions between features as well as non-nonlinearities. Throughout we are describing what adjudicators *do*, not whether they are right to do it. Whether the applicants they screen out would have in fact performed worse is a separate question we return to in Sections 6 and 7. One thing to keep in mind is that the visible record is itself a product of the status quo visibility rules already applied—some charges are hidden before the adjudicator ever sees the file. So what we describe is how adjudicators use and react to the record they are shown.

Table 5 shows regressions of allowance on various visible criminal record and demographic characteristics, among those applicants in the status quo sample with a visible criminal record. These criminal characteristics explain about 52% of the variation in allowance decisions of adjudicators (column 1). Adding demographic characteristics (column 2) or information about the adjudicator who made the decisions (column 4) only increases the R^2 by 0.001.

Applicants with a visible felony are about 40 percentage points less likely to be allowed, and a visible property or violent charge lowers the chance of allowance by about 26 and 21 percentage points. This is mostly what we would expect: more serious charges make allowance less likely. Having an older record is associated with a higher probability of allowance; more charges with a lower probability of allowance. A couple of signs are less intuitive. Conditional on the rest of the record, a visible traffic or drug charge is associated with a slightly higher chance of being allowed. We read this as relative severity rather than adjudicators favoring these charges, that is, once the more serious charges on a record are accounted for, a traffic or drug charge is a comparatively mild signal. In Column (2) we add in demographic variables as described in Section 3. While adjudicators rely heavily on the visible record characteristics, conditional on these record characteristics predicted race, gender, and age add almost no incremental explanatory power to the visible criminal

record.[12] In Column (3) we add process variables such as which of 3 adjudicators screened the criminal background check; the count of applicants processed by the same adjudicator on the same day; and year-month fixed effects. These variables similarly do not increase explanatory power (and none are significant, regression coefficients are suppressed).

The coefficients on these record characteristics are hard to compare on their own, though, because many of these features move together: applicants with felonies also tend to have violent charges, more charges, and a higher share of felonies. To get at how much each feature contributes to explaining the decisions, we use a Shapley decomposition, which splits the explained variation across features and shares credit fairly among correlated ones (Shapley, 1953; Shorrocks et al., 2013; Israeli, 2007). Because allowance is binary, our main version decomposes the pseudo-R^2 of a logistic regression; the resulting dominance weights are Shapley values (Azen and Traxel, 2009). Figure 3 shows the result. Severity and charge type explain most of the variation in screening decisions. Having a felony is the single largest contributor (about 26%), followed by the share of charges that are felonies (18%), a property charge (16%), a violent charge (13%), and a traffic charge (12%). The number of charges matters somewhat (about 8%), while recency, pending status, demographics, and the adjudicator each contribute very little. These shares measure how much each characteristic helps distinguish allowance decisions, not whether it raises or lowers allowance. Traffic charges, for example, are important because conditional on other record features they identify relatively less serious records and predict allowance. This is a different way of showing the same point: adjudicators are screening on how serious the record looks, but not on other characteristics of the applicant or process.[13]

The regressions so far are mainly linear and additive, but adjudicators might instead pay attention to interactions of these variable or non-linearities (e.g. a more recent felony may matter more; or going from 0 to 1 charge may matter more than 4 to 5 charges). To check this, we add squares of the continuous features and a set of interactions among the main ones (felony with recency, number of charges, and pending status; recency with number; and so on) to the logit. The fit barely moves: the pseudo-R^2 rises from 0.46 to 0.47. Because this is in-sample, where adding terms can only improve fit, even this small gain overstates how much the interactions matter. Adjudicators appear to be applying a simple, roughly

[12] The only statistically significant demographic coefficient is on predicted female: applicants predicted to be female are approximately 2 percentage points less likely to be allowed than applicants predicted to be male, conditional on visible record characteristics. This association does not establish differential treatment by gender, given that gender is inferred and record characteristics may differ in ways not captured by the specification.

[13] We get nearly identical shares decomposing the linear probability model instead (Figure B.1).

additive rule, in which each part of the record contributes largely on its own.

Finally, because R^2 is hard to interpret for a binary decision, we ask directly how well the visible record predicts allowance out of sample. Training a model on part of the sample and evaluating it on the rest, the visible record predicts allowance with an area under the ROC curve of 0.92 (Table 9). Allowing for nonlinearities and interactions barely improves fit, either in sample or out of sample. The adjudication rule therefore appears highly systematic but approximately additive: a small number of record dimensions account for nearly all of its predictive content.

Overall, adjudicators are clearly not making allowance decisions randomly. Their decisions are strongly associated with the characteristics of the visible criminal record, and particularly with measures of record severity, charge type, and extensiveness of the record. At the same time, these patterns do not tell us whether adjudicators are placing weight on the characteristics that are most informative about subsequent performance on the marketplace. We turn to those questions next.

5 Observational Results: Outcomes

We now switch to the question of how criminal records are associated with outcomes on the marketplace. The purpose is no longer to reconstruct what adjudicators saw; it is to ask whether actual criminal histories predict performance on the marketplace. So, for the outcome analysis, we use the corresponding characteristics of the applicant's full legally reportable record, rather than only the charges visible to adjudicators. This gives the prediction exercise more information than the adjudicator had and asks whether criminal history could predict outcomes even if the firm used the record more fully. Note, of course, that we can only see outcomes for applicants who were actually allowed on the marketplace.

We start with the simple comparison between applicants with and without any criminal record. Table 6 shows that, among control-group applicants allowed onto the marketplace, applicants with a criminal record are less likely to make any contact with a client (17.8pp, or 25%), make fewer contacts on average (6.4, or 23%), and are more likely to have an incident reported to the marketplace (.6 pp or 33%, off a base rate of 1.8%). These gaps remain similar after adjusting for predicted race, predicted gender, age, and stated main service provided. Thus, at least in the cross section, applicants with records appear to engage less and experience somewhat higher incidence rates.

This does not yet tell us whether the detailed distinctions adjudicators draw among appli-

cants with records are useful. As shown in Section 4, adjudicators rely heavily on characteristics such as whether the applicant has a felony, the severity and type of the visible charges, and the number of charges or episodes. We next ask whether these same underlying criminal-record characteristics are associated with outcomes among applicants with records who were allowed under the status quo.

Table 7 runs the same kind of "kitchen-sink regressions" for three marketplace outcomes: whether the applicant makes any contact, the number of contacts, and whether they have any incident.[14] The contrast with the allowance regressions is striking. Although several record characteristics are individually associated with engagement, the record as a whole explains little of the variation in realized outcomes. Table 8 makes this concrete: record features explain about 1.5% of the variation in any contact, 0.4% in the number of contacts, and 0.2% in whether there is any incident, compared with about 52% for the allowance decision itself.[15]

The coefficients reveal somewhat more structure for engagement than the low R^2 values alone suggest. Applicants with pending, violent, or property charges are less likely to contact clients, and applicants with more charges, pending charges, or property charges make fewer contacts. But the record dimensions most heavily weighted in adjudication do not map cleanly into safety outcomes: felony, violent, and property charges are not associated with a higher probability of an incident. Thus, detailed records contain some information about subsequent engagement, but very little information about the safety outcome that is one central motivation for exclusion.[16]

To make the comparison sharper, we ask how well the record predicts these outcomes out of sample, using the same approach we used for the allowance decision (Table 9). The contrast is stark. The same record dimensions predict allowance extremely well out of sample, with an AUC of 0.92. They predict client contact only slightly better than chance, with an AUC of 0.54, and have essentially no out-of-sample power for incidents or contact volume. Allowing

[14] Full tables that progressively add characteristics, mirroring Table 5, can be found in Appendix Tables B.7–B.9.

[15] Appendix Figures B.2, B.3, and B.4 decompose the explained variation for each outcome into the contribution of each block of characteristics. Because the record explains so little to begin with, these are decompositions of a very small total; consistent with the regressions, the criminal-record features account for little of it, and what is explained comes mostly from occupation, demographics, and the timing of entry.

[16] The number of contacts associations should be read with caution: the number of contacts is heavily skewed, with most allowed applicants making only a handful and a few making hundreds, so individual coefficients are sensitive to a small number of high-volume providers. The property-charge association, for instance, reflects a few more high-volume providers among applicants without property charges rather than any difference for the typical applicant—the median is unchanged, and the association disappears under median or rank-based methods.

for nonlinearities and interactions through an elastic net does not change this. The record that closely predicts who is allowed carries little information about realized engagement or incidents among those allowed.

These estimates do not validate or invalidate the screening policy, because outcomes are observed only after the policy has selected applicants onto the marketplace. If adjudicators successfully exclude the highest risk applicants within each record category, the remaining allowed applicants may look relatively similar. Incidents are also rare, making them difficult to predict. The estimates instead show that, within the portion of the record distribution allowed under the status quo, detailed criminal-history characteristics provide little additional rank-ordering of realized outcomes.

The missing object is the performance of applicants whom the existing policy excludes—particularly those near the firm's allowance threshold. Applicants with records engage less and have somewhat higher incident rates than applicants without records on average, but the fine-grained record characteristics that drive screening decisions do not strongly predict outcomes among those allowed. To learn what would happen if the firm relaxed its policy and allowed some applicants it would otherwise reject, we turn to the experiment. We later use the incumbent sample to similarly tell us more about outcomes for applicants with records that would not have been allowed on under the status quo.

6 Experiment

6.1 Experimental Variation and First Stage

We now use the experiment to study the consequences of relaxing the firm's criminal-record screening policy. As described in Section 2, background checks conducted during the sample period were randomly assigned either to the status quo control ruleset or to a treatment ruleset that hid additional categories of charges.

The treatment ruleset changes the visibility of applicants' criminal records in heterogeneous ways. We classify applicants with charges according to how the treatment ruleset changes the record shown to adjudicators relative to the control ruleset:

1. Potentially Impacted By Treatment

    - **Newly Clear**: the applicant has at least one visible charge under the control ruleset, but the record is completely clear under treatment. For example, an applicant whose only visible charge is a misdemeanor speeding charge from two years

ago is sent to a human adjudicator under the control ruleset, but is automatically allowed under the treatment ruleset, which hides the charge.

- **Reduced Visibility**: the applicant has visible charges under the control ruleset, and treatment hides some but not all of them. For example, an applicant with a felony drug possession charge and a felony assault charge, both from five years ago, has two visible charges under the control ruleset; the treatment ruleset hides the drug charge but not the assault charge, so a human adjudicator still reviews the record.

2. Record Unchanged (Placebo)

   - **Always Clear**: the applicant's record is already completely clear under the control ruleset and remains clear under treatment.
   - **Same Visibility**: the visible record is the same under the treatment and control matrices.

Because we know the rules of both matrices, each applicant's treatment response type can be determined regardless of actual assignment.

Randomization was implemented by computer at the background-check company at the individual level, without stratification. Human adjudicators were not informed that the experiment was taking place, and no indicator on the reports they reviewed revealed treatment assignment. Applicants with no criminal charges were also randomly assigned, although assignment could not affect the information shown for these applicants.

The experiment hides some charges from adjudicators, potentially allowing applicants with records who would otherwise have been excluded to access the marketplace. Let Z_i indicate assignment to the treatment ruleset and D_i indicate being allowed following the background check. We estimate the first-stage relationship

$$D_i = \delta + \pi Z_i + \nu_i.$$

Table 10 reports how allowance rates differed for the treatment group compared to the control group. Among new entrants to the marketplace with at least one legally reportable charge ($N = 27,039$), treatment increases the probability of being allowed by 3.5 percentage points from a control-group mean of 65.8 percent, an increase of approximately 5.3%. This implies approximately 950 additional applicants allowed onto the marketplace by the experiment.

The robust first-stage F is 38.91.

The remaining columns separate applicants by treatment response type. As expected, the first stage is concentrated among applicants whose visible records are changed by treatment — the newly clear and reduced visibility groups, which we refer to together as *movers*. Among movers, treatment increases allowance by 4.7 percentage points over a control mean of 68.6%, a 6.9% increase in access to the marketplace.

For newly-clear applicants, the control ruleset sends the record to a human adjudicator, whereas treatment hides all charges and results in essentially automatic clearance. Allowance therefore rises from 94.6% under control to 100% under treatment, a 5.3 percentage-point increase ($F = 225$). The high control allowance rate implies that adjudicators already allowed most applicants with these records; the experiment changes access for the relatively small share whom human review would otherwise exclude.

Among reduced-visibility applicants, a human adjudicator reviews the record under both assignments; treatment does not bypass that review but changes the decision by removing some of the charges the adjudicator sees. Allowance rises by 3.5 percentage points from a control mean of 23.7%, a 15% increase. The first stage is weaker in this group ($F = 7.93$), consistent with treatment changing decisions for only a subset of applicants whose records remain subject to human review.

As expected, treatment has essentially no effect among applicants whose records are unchanged by treatment or those with no charges. These placebo first stages provide evidence against treatment spillovers affecting applicants whose displayed records are unchanged.

6.2 Reduced Forms

We estimate reduced forms of being assigned to the treatment ruleset on outcomes on the marketplace via

$$Y_i = \kappa + \rho Z_i + \eta_i$$

Where Y_i are various marketplace outcomes, such as making any contact with clients or any incident reported to the marketplace. Z is random assignment to the treatment ruleset as defined previously. These results combine potential impacts of treatment on being allowed on the marketplace and usage conditional on being allowed.

Table 11 displays these results for unconditional outcomes: any contact, review, or incident,

as well as number of contacts and how many days someone was on the marketplace (measured from date the background check was completed to date of the last client contact we have recorded). Among applicants with a charge, assignment to treatment increases the probability of contacting at least one client by 1.6 percentage points (4.5%) and leaves the incident rate unchanged (a point estimate of −0.2 percentage points on a control mean of 1.5%). The treatment group also tends to stay somewhat longer in the marketplace. These effects on engagement are concentrated in movers (the newly clear and reduced visibility groups). We see no statistically significant differences between treatment and control in the placebo groups. Appendix Table B.10 reports these results separately for the newly clear and reduced visibility groups, along with intensity of marketplace usage and other conditional outcomes. Because the latter outcomes are defined conditional on engagement, random assignment no longer guarantees balance in the sample used for estimation; these estimates should therefore be interpreted as descriptive. Conditional on engagement, we see no systematic differences between treatment and control in the number of contacts, reviews or incidents per contact, or average ratings; the few individually significant differences do not survive adjustment for multiple hypothesis testing.

Overall, we interpret these results as suggesting that relaxing the screen does not increase incidents and, if anything, modestly increases engagement on the marketplace relative to the status quo.

6.3 Identifying Outcomes for Marginal Applicants

The reduced form results suggest that assignment to treatment does not increase incidents and, if anything, modestly increases engagement on the marketplace. However, these estimates average over all applicants, including a relatively small share of compliers (4.7% among movers). In addition, these comparisons include both applicants who are allowed on the marketplace and those who are not, for whom outcomes are mechanically equal to zero.

To isolate the outcomes of interest for the compliers, we adopt the LATE framework of Imbens and Angrist (1994). This framework allows us to characterize the applicants at the margin of the screening rule—those who would be denied access under the control policy but are allowed under treatment—and to estimate their outcomes when allowed. A key question is whether these marginal applicants perform differently from always takers, who are allowed under both policies. In our setting, compliers correspond to applicants at the margin of the firm's screening threshold.

Table 12 summarizes the standard compliance types for a binary instrument Z (treatment assignment) and binary treatment D (marketplace access). Given random assignment, the characteristics of always takers and never takers are directly observed:

$$E[X \mid Z = 0, D = 1] \quad \text{and} \quad E[X \mid Z = 1, D = 0].$$

Complier characteristics can be recovered using the first stage and standard IV weighting arguments (equivalently, via 2SLS regressions of $X_i \times \mathbb{1}(D_i = 1)$ on Z_i; see Abadie 2003).

The same approach identifies outcomes for compliers. Letting $D(z)$ and $Y(d)$ be standard potential outcomes notation for treatment (when $Z = 0$ or $Z = 1$) or outcomes (when $D = 0$ or $D = 1$):

$$E[Y(1) \mid D(1) > D(0)].$$

In our setting, this corresponds directly to the 2SLS estimate from

$$Y_i = \alpha + \beta D_i + \varepsilon_i,$$

instrumenting D_i with Z_i. Because $Y_i(0) = 0$ by construction for applicants not allowed on the marketplace, the IV estimand recovers the average outcome for compliers when allowed.

Among applicants who are allowed onto the marketplace, those in the control group are all always takers, while those in the treatment group are a mix of always takers and compliers. Because assignment is random, always takers in the two groups have the same expected outcomes, so any difference between the two allowed pools reflects the compliers. Writing π_{AT} and π_C for the shares of always takers and compliers,

$$E[Y \mid Z = 1, D = 1] = \frac{\pi_{AT}}{\pi_{AT} + \pi_C} E[Y(1) \mid \text{always taker}] + \frac{\pi_C}{\pi_{AT} + \pi_C} E[Y(1) \mid \text{complier}],$$

where $E[Y(1) \mid \text{always taker}] = E[Y \mid Z = 0, D = 1]$ and the first stage identifies the shares ($\pi_{AT} = E[D \mid Z = 0]$ and $\pi_C = E[D \mid Z = 1] - E[D \mid Z = 0]$). The complier mean is the only unknown. This decomposition also shows why the complier estimates are imprecise: compliers are a small share of the allowed treatment pool (about 6% among movers), so small differences between the two pool means translate into large differences in the implied complier mean.

Interpreting these estimates requires the standard LATE assumptions of independence, exclusion, relevance, and monotonicity. Appendix C discusses each in our setting and reports

tests of monotonicity for the reduced visibility group, where it does not hold mechanically.

6.4 Who are the Compliers?

The experiment allows some applicants who would have been denied access under the control ruleset to instead be allowed to the marketplace. These compliers are of interest in their own right: they are the applicants adjudicators were excluding at the margin under the status quo, and their criminal records show where the firm's screening threshold binds. In the spirit of Gruber et al. (1999), who characterize the "marginal child," we first ask who the marginal applicant is before asking how they perform.

Table 13 reports characteristics of compliers, always takers, and (where relevant) never takers, using the approach described above. We present results for the pooled sample of movers — the newly clear and reduced visibility groups, whose visible records treatment can change — as well as separately for each group.[17]

Compliers have more severe records than always takers, but less severe records than never takers. Among movers, never takers have on average 9.7 charges, compared to 5.8 for compliers and 3.4 for always takers. The same ordering holds for felony, drug, and pending charges: 53% of compliers have a felony, compared to 7% of always takers and 72% of never takers. The exception is violent and property charges, where compliers look similar to always takers (5% and 7% of compliers, compared to 9% and 8% of always takers) and very different from never takers, more than half of whom have each type of charge.

There are also meaningful differences across treatment response types. In the newly clear group, compliers are substantially more likely to have felony or drug charges, suggesting that these charges play an important role in triggering exclusion by humans under the control ruleset. When these charges are fully hidden under treatment, these applicants are automatically allowed. In contrast, compliers in the reduced visibility group tend to have more extensive criminal histories—including more charges and more felonies overall—suggesting that adjudicators respond to the cumulative number and severity of records. When some, but not all, charges are hidden, these applicants move across the allowance threshold.

Comparing compliers across these two groups is also informative. "Newly clear" compliers are automatic compliers and have fewer charges, fewer felonies, are less likely to have violent

[17] There are no never takers in the newly clear group: treatment hides the entire visible record, so these applicants are automatically allowed.

offenses, and have older records than compliers in the reduced visibility group (where a human makes the decision). In contrast, "reduced visibility" compliers tend to have more recent and more extensive records. This variation provides additional leverage for understanding which dimensions of criminal history drive screening decisions and how these dimensions relate to subsequent outcomes.

Table 14 and Figure 5 show how the experiment generates these differences, reporting compliers' *visible* records under the control and treatment matrices. Under the control ruleset, compliers among movers have on average 3.3 visible charges, and 48% have a visible felony; under the treatment ruleset, they have 1.0 visible charge, and 11% have a visible felony. For newly clear compliers the change is mechanical, since treatment hides the entire visible record, and what it hides is mostly drug and traffic charges: 68% have a visible drug charge and 58% a visible traffic charge under control. For reduced visibility compliers, treatment removes roughly 40% of visible charges (from 6.2 to 3.7) but leaves the most serious ones in place: the share with a visible felony falls only from 49% to 39%, and the share with a visible violent charge is essentially unchanged.

Overall, the complier analysis suggests that adjudicators are not making arbitrary decisions. Criminal record characteristics clearly matter for screening: applicants with more severe records are less likely to be allowed. At the same time, the experiment induces variation precisely along these dimensions, allowing some applicants with more severe records than the average allowed worker to enter the marketplace. This variation provides leverage to study how and whether these criminal records map into realized outcomes.

6.5 Average Outcomes for Marginally Allowed Applicants

We now turn to outcomes for compliers — the marginal applicants whose allowance status changes under treatment. Table 15 reports results for the pooled sample of movers. Column (1) reports mean outcomes for always takers (control-group applicants who are allowed), column (2) for all allowed applicants in the treatment group (always takers plus compliers), and column (3) the implied complier mean.

Adding the marginally allowed applicants changes little about the pool of providers on the marketplace. Among allowed movers, 55.6% of the control group contact at least one client, compared to 54.8% of the treatment group; 20.8% and 20.6% make more than ten contacts; and 2.5% and 2.1% have an incident.

The implied complier means point the same way but are much less precise. We estimate

that 41% of compliers contact at least one client, compared to 56% of always takers, and we cannot reject that the two are equal ($p = 0.34$). The same holds for making more than ten contacts (16% versus 21%, $p = 0.72$). For incidents, the complier point estimate lies below the always-taker mean of 2.5%, and again we cannot reject equality ($p = 0.18$).[18] In other words, we find no detectable difference between marginally allowed applicants and those the marketplace was already allowing, and no evidence of an increase in incidents, although the estimates for engagement are imprecise enough that we cannot rule out meaningful shortfalls.

Appendix Table B.11 splits these results by treatment response type. Among newly clear applicants, 50% of compliers make any contact, compared to 57% of always takers ($p = 0.74$). Among reduced visibility applicants the gap in point estimates is larger (15% versus 47%) but is not statistically significant ($p = 0.17$), and it rests on a small sample and a weak first stage ($F = 7.93$). In both groups the complier point estimate for incidents lies below the always-taker mean, and in neither can we reject equality.

Appendix Table B.12 reports outcomes conditional on marketplace activity. Because these outcomes are defined conditional on engagement, they should be interpreted as descriptive. Among movers who contact at least one client, complier point estimates for the number of contacts, reviews per contact, and average rating are at or above those of always takers, and incidents per contact are no higher; none of these differences is statistically significant.

These comparisons are relative to applicants with records whom the marketplace was already allowing, not to the typical provider. Appendix Table B.13 adds applicants with no criminal record as a benchmark. In line with Section 5, allowed applicants with records engage less than those without: 56% of always takers among movers contact at least one client, compared to 72% of applicants with no record, and 2.5% have an incident, compared to 1.8%. Allowing the marginal applicants leaves these gaps essentially unchanged. In other words, relaxing the screening rule does not change how allowed applicants with records compare to the rest of the marketplace, but it does not close the gap either.

Taken together, these results suggest that applicants allowed at the margin of the firm's screening rule are not detectably different, on measured engagement or incidents, from those the marketplace was already allowing. The complier estimates are imprecise, so this conclusion rests on the absence of a detectable difference rather than on a tightly estimated

[18] The point estimate for the complier incident rate is negative. Because the first stage is modest in levels (about 5 percentage points), small differences in pool means translate into noisy complier estimates, which can fall outside the natural range of the outcome.

zero.

7 Incumbents

In addition to new entrants, over 56,000 incumbent service providers completed a background check during the sample period. As described in Section 2, incumbents joined the marketplace before a check was required for their cohort upon signup. By construction, they had not completed a background check before the one we observe, and had not been removed from the marketplace for a substantiated incident before that check. Of the 56,347 incumbents, 7,300 (13.0%) have at least one legally reportable charge.

Table 16 and Table 17 compare charged incumbents' and charged entrants' criminal records. Entrants are more likely to have any charge at all (14.3% versus 13.0%) and, conditional on being charged, more likely to have a felony, violent, property, drug, theft, or burglary charge (all $p < 0.001$); incumbents skew toward more minor and administrative charges, and their most recent charge is older.

Table 18 compares excluded and allowed incumbents' pre-background-check performance directly. The two groups look statistically similar in whether they had any contact with potential clients and any incident; excluded incumbents show slightly *fewer* pre-background-check contacts.

Table 19, alongside entrants' Table 13, compares the records of always takers and compliers across the two samples. Always takers look nearly identical on the full criminal record. Compliers diverge: incumbent compliers are far more likely to carry a violent charge (44% versus 5%) or property charge (25% versus 7%) than entrant compliers. In other words, the incumbents at the margin of the screening decision have more serious records than the entrants at that margin. Table 20 splits this comparison by treatment response type and shows that the divergence is concentrated in the reduced visibility group (70% violent, 38% property among compliers) rather than spread evenly across both: newly clear compliers look closer to, though still somewhat elevated relative to, entrant compliers (26% violent, 15% property).

Table 21 compares post-background check outcomes across compliance types. Incumbent compliers are statistically indistinguishable from always takers on nearly every outcome; the exception is reviews per contact, where compliers show a higher rate.

Table 22 shows that simply having a charge predicts a higher likelihood of any pre-background

check contact and a higher incident rate (both $p < 0.01$), surviving tenure and demographic controls, though it does not significantly predict the raw number of contacts. This is the opposite sign from what we find for entrants (Section 5, Table 6), where a record predicts *less* engagement; the incident-rate result, by contrast, points the same direction in both samples.

As is true for entrants, however, the specific content of the record does not predict these outcomes: in Figure 6, coefficients on individual record features are indistinguishable from zero for both outcomes. Figure 7 decomposes this directly – occupation alone accounts for roughly 50% of the explained variation, more than every record-feature category combined (≈14%).

Figure 8 shows the same null result holds for allowed incumbents over a 12-month post-background-check window. Figure 9 again finds occupation and adjudication timing, not record content, doing the explanatory work (together over 70%, versus single digits for every record category).

Taken together, these results suggest that in both samples the applicants at the screening margin have more serious records than those allowed under either ruleset, but that record content does not predict marketplace activity, before or after the background check

8 Discussion

We study the effects of hiding certain criminal records from human adjudicators for randomly selected applicants in a large online marketplace. Under the status quo rules, adjudicators screen systematically: the visible record predicts allowance decisions closely, with felony status, charge type, and the number of charges doing most of the work. Yet among applicants allowed onto the marketplace, the same record features carry little information about engagement and almost none about incidents.

Our experiment increases access to the marketplace among applicants with criminal records. Despite this increase in allowance, we find no increase in incidents among those assigned to treatment, and treated applicants are, if anything, more engaged on the marketplace. Using a LATE framework, we find that marginally allowed applicants — those who are excluded under the control policy but allowed under treatment — are not detectably different, in incidents or engagement, from applicants with records whom the marketplace was already allowing, although the estimates for this group are imprecise.

These findings contrast with cross-sectional comparisons, where applicants with criminal records perform worse on average. Taken together, the results suggest that while criminal records may be correlated with engagement and incidents, the information used to exclude marginal applicants is not strongly predictive of realized performance in this setting. If adjudicators were optimally setting screening thresholds based on correct beliefs about expected benefits and costs, marginal applicants should have lower net value. Instead, we find no increase in incidents and no detectable difference in engagement among marginally allowed service providers. Because the marketplace faces no capacity constraint, this suggests that screening thresholds may be overly strict at the margin.

These patterns are consistent with two interpretations, which are not mutually exclusive. The first is that the screen is stricter than the measured outcomes warrant: adjudicators may weight salient features of criminal records, such as charge type, severity, or recency, more heavily than their predictive content justifies. Criminal records are high-dimensional and heterogeneous signals, and adjudicators receive little feedback on how allowed service providers perform and none on excluded applicants, which limits their ability to learn about performance at the margin and recalibrate. In this setting, even record features that are informative in cross-sectional comparisons may be weakly predictive at the margin relevant for decision-making. The second interpretation is that the firm screens for costs we do not observe well, such as reputational risk or client preferences, including discriminatory preferences.

At the same time, our results do not imply that criminal records are uninformative or that screening is unwarranted. Applicants with criminal records perform worse on average, which may partially reflect correlated unobservables. The key distinction is between average differences and marginal decision-making: the applicants excluded at the margin are not detectably different, in terms of realized outcomes, from applicants with records who are allowed.

Our estimates focus on average outcomes. Firms may rationally place weight on tail risks—rare but high-cost events—that our incident measures capture poorly, and the experiment may lack the statistical power to detect changes in such events. To the extent that tail risks or the unobserved costs discussed above are important, optimal screening thresholds may be stricter than average realized outcomes alone would imply. Even so, our results suggest that modest expansions in access would not measurably increase incidents.

While our results are specific to one online marketplace, they show how a firm uses criminal record information in practice and provide causal evidence on how relaxing that screen affects

realized outcomes. Future work should examine how these patterns generalize across settings and explore how firms can better align screening decisions with realized productivity and risk.

References

Abadie, A. (2003): "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Econometrica*, 71, 265–292.

Agan, A. and S. Starr (2018): "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment," *The Quarterly Journal of Economics*, 133, 191–235.

Agan, A. Y., A. Garin, D. K. Koustas, A. Mas, and C. Yang (Forthcoming): "Can you Erase the Mark of a Criminal Record? Labor Market Impacts of Criminal Record Remediation," *American Economic Journal: Economic Policy*.

Athey, S. and G. W. Imbens (2017): "The econometrics of randomized experiments," in *Handbook of economic field experiments*, Elsevier, vol. 1, 73–140.

Autor, D. H. and D. Scarborough (2008): "Does job testing harm minority workers? Evidence from retail establishments," *The Quarterly Journal of Economics*, 123, 219–277.

Azen, R. and N. Traxel (2009): "Using Dominance Analysis to Determine Predictor Importance in Logistic Regression," *Journal of Educational and Behavioral Statistics*, 34, 319–347.

Blumstein, A. and K. Nakamura (2009): "'Redemption' in an Era of Widespread Criminal Background Checks: (516022010-003)," Institution: American Psychological Association.

Brown, C. (1982): "Estimating the determinants of employee performance," *Journal of Human Resources*, 178–194.

Bushway, S. (2024): "Resetting the Record: The Facts on Hiring People with Criminal Histories," Research Brief RB-A2968-1, RAND Corporation.

Bushway, S. D., G. Sweeten, and P. Nieuwbeerta (2009): "Measuring Long Term Individual Trajectories of Offending Using Multiple Methods," *Journal of Quantitative Criminology*, 25, 259–286.

Cullen, Z., W. Dobbie, and M. Hoffman (2022): "Increasing the Demand for Workers with a Criminal Record," *The Quarterly Journal of Economics*, 138, 103–150, publisher: Oxford Academic.

Denver, M., G. Siwach, and S. D. Bushway (2017): "A new look at the employment and recidivism relationship through the lens of a criminal background check," *Criminology*, 55, 174–204.

Douglas, R. M. and L. Maltby (2023): "Second Chance Employment Report," Tech. rep., Legal Action Center and National Workrights Institute.

Frandsen, B., L. Lefgren, and E. Leslie (2023): "Judging judge fixed effects," *American Economic Review*, 113, 253–277.

Goggins, B. R. and D. A. DeBacco (2024): "Survey of State Criminal History Information Systems, 2022," Tech. Rep. NCJ 309360, Bureau of Justice Statistics.

Gruber, J., P. Levine, and D. Staiger (1999): "Abortion legalization and child living circumstances: Who is the "marginal child"?" *The Quarterly Journal of Economics*, 114, 263–291.

Heckman, J. J. and E. Vytlacil (2005): "Structural equations, treatment effects, and econometric policy evaluation 1," *Econometrica*, 73, 669–738.

Hoffman, M., L. B. Kahn, and D. Li (2018): "Discretion in hiring," *The Quarterly Journal of Economics*, 133, 765–800.

Holzer, H., S. Raphael, and M. Stoll (2006): "Perceived Criminality, Criminal Background Checks, and the Racial Hiring Practices of Employers," *The Journal of Law & Economics*, 49, 451–480, publisher: [The University of Chicago Press, The Booth School of Business, University of Chicago, The University of Chicago Law School].

Imai, K., K. Khanna, B. Bertelsen, S. Olivella, E. Rosenman, and A. Rossell Hayes (2024): "wru: Who are You? Bayesian Prediction of Racial Category Using Surname, First Name, Middle Name, and Geolocation," .

Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

Israeli, O. (2007): "A Shapley-Based Decomposition of the R-Square of a Linear Regression," *Journal of Economic Inequality*, 5, 199–212.

Kitagawa, T. (2015): "A test for instrument validity," *Econometrica*, 83, 2043–2063.

Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan (2018): "Human decisions and machine predictions," *The quarterly journal of economics*, 133, 237–293.

Kurlychek, M., S. Bushway, and M. Denver (2019): "Defensible decisions: Balancing employer and prospective employee rights in an era of criminal background checks," *Equality, Diversity and Inclusion: An International Journal*, 38, 529–546.

Kurlychek, M. C., S. D. Bushway, and R. Brame (2012): "Long-Term Crime Desistance and Recidivism Patterns—Evidence from the Essex County Convicted Felon Study," *Criminology*, 50, 71–103, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1745-9125.2011.00259.x.

Lageson, S. and R. Stewart (2023): "The problem with criminal records: Discrepancies between state reports and private-sector background checks," *Criminology*, 1–30, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1745-9125.12359.

Lakkaraju, H., J. Kleinberg, J. Leskovec, J. Ludwig, and S. Mullainathan (2017): "The selective labels problem: Evaluating algorithmic predictions in the presence of unobservables," in *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 275–284.

Lundquist, J. H., D. Pager, and E. Strader (2018): "Does a criminal past predict worker performance? Evidence from one of America's largest employers," *Social Forces*, 96, 1039–1068.

Minor, D., N. Persico, and D. M. Weiss (2018): "Criminal background and job performance," *IZA Journal of Labor Policy*, 7, 8.

Mueller-Smith, M. and K. Schnepel (2021): "Diversion in the criminal justice system," *The Review of Economic Studies*, 88, 883–936.

Mullen, L. (2021): "gender: Predict Gender from Names Using Historical Data," .

Müller, U. K. (2025): "A more robust t-test," *Review of Economics and Statistics*, 107, 786–802.

Pager, D. (2003): "The Mark of a Criminal Record," *American Journal of Sociology*, 108, 39.

Pallais, A. (2014): "Inefficient hiring in entry-level labor markets," *American Economic Review*, 104, 3565–3599.

Professional Background Screening Association (2021): "Background Screening: Trends in the U.S. and Abroad," Tech. rep.

Pyle, B. (2023): "Negligent Hiring: Recidivism and Employment with a Criminal Record," *Working Paper*.

Rose, E. K. (2021): "Does banning the box help ex-offenders get jobs? Evaluating the effects of a prominent example," *Journal of Labor Economics*, 39, 79–113.

Schnepel, K. T. (2018): "Good jobs and recidivism," *The Economic Journal*, 128, 447–469.

Shannon, S. K. S., C. Uggen, J. Schnittker, M. Thompson, S. Wakefield, and M. Massoglia (2017): "The Growth, Scope, and Spatial Distribution of People With Felony Records in the United States, 1948–2010," *Demography*, 54, 1795–1818.

Shapley, L. S. (1953): "A Value for n-Person Games," in *Contributions to the Theory of Games (AM-28), Volume II*, ed. by H. W. Kuhn and A. W. Tucker, Princeton University Press, vol. 28 of *Annals of Mathematics Studies*, 307–317.

Shorrocks, A. F. et al. (2013): "Decomposition procedures for distributional analysis: a unified framework based on the Shapley value," *Journal of Economic Inequality*, 11, 99–126.

Uggen, C., M. Vuolo, S. Lageson, E. Ruhland, and H. K. WHITHAM (2014): "The edge of stigma: An experimental audit of the effects of low-level criminal records on employment," *Criminology*, 52, 627–654.

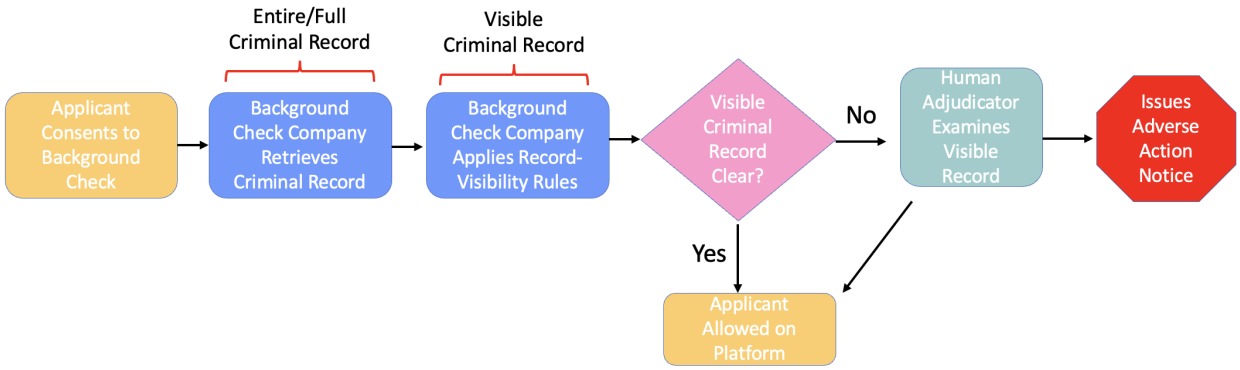
Weiss, D. M., W. Dawson, R. McKinley, and L. Webster (2025): "Do ban-the-box policies increase the hiring of applicants with criminal records?" *Plos one*, 20, e0320736.

Yang, C. S. (2017): "Local labor markets and criminal recidivism," *Journal of Public Economics*, 147, 16–29.

Young, A. (2019): "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results," *The Quarterly Journal of Economics*, 134, 557–598.

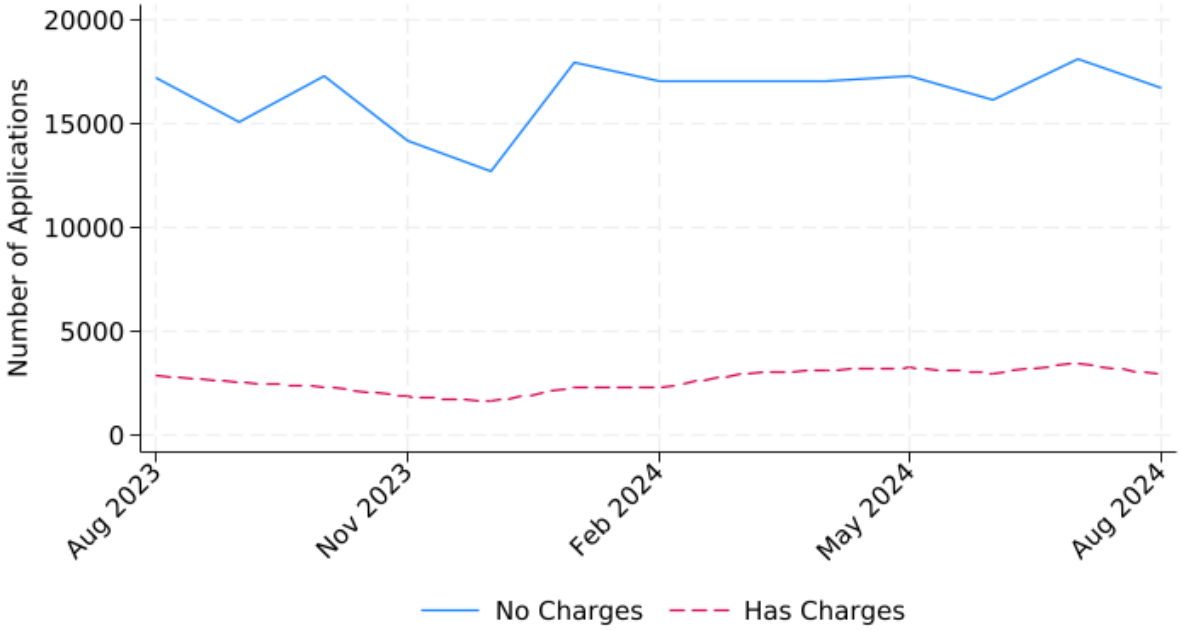
Figures and Tables

Figure 1: Criminal Background Check Process



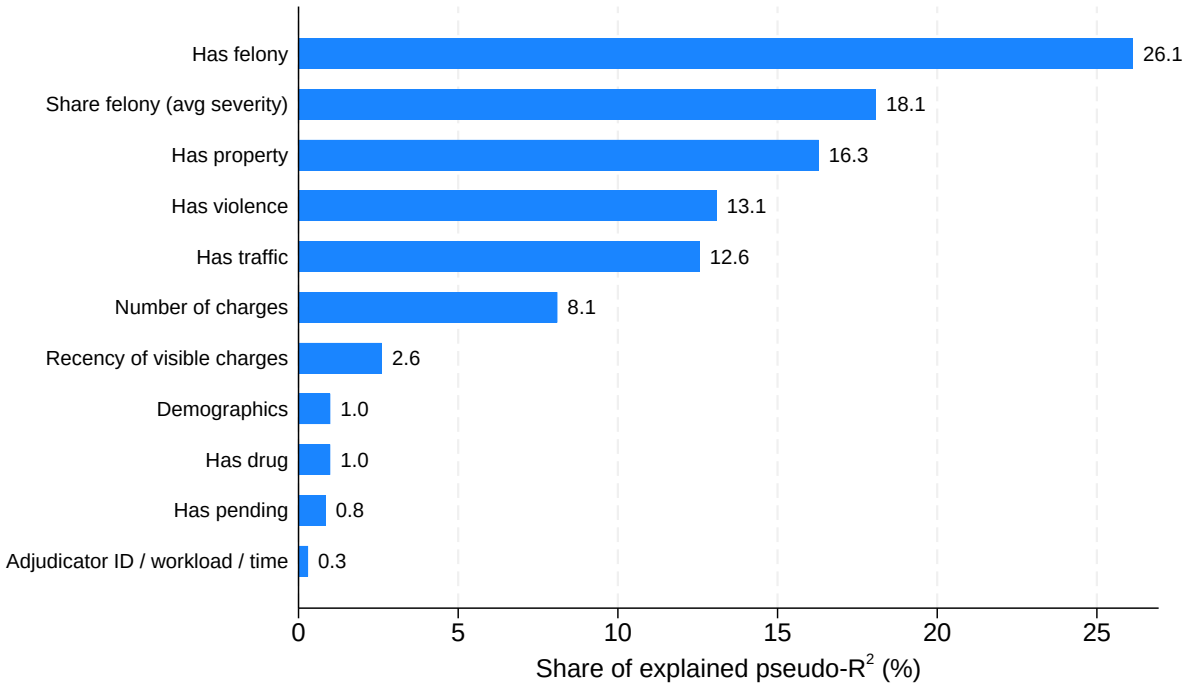
Note: This figure shows a simplified flowchart detailing the criminal background check process.

Figure 2: Application Volume over Time by Charge Status



Note: The figure shows the monthly volume of background check reports run for applicants to the platform over the course of the experimental period, broken out by whether the applicant has a legally reportable criminal charge on their record.

Figure 3: Logit Dominance Decomposition of Allowance Decisions

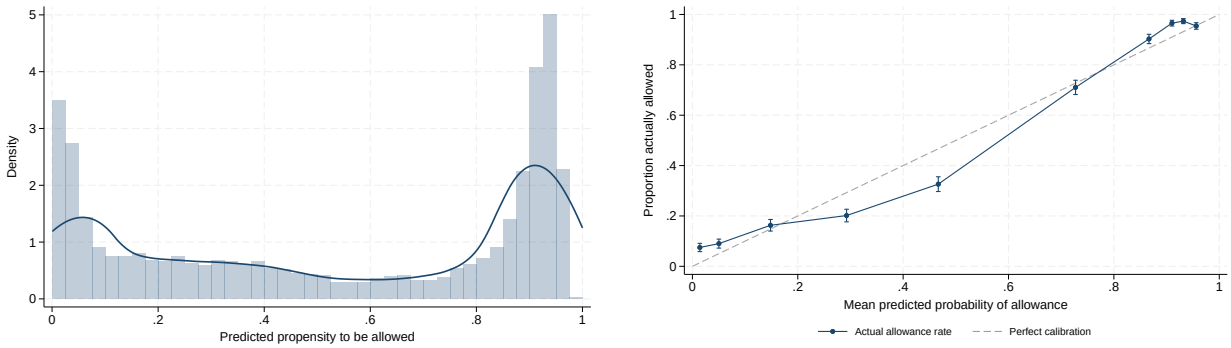


Note: Bars show each block's general-dominance share (Shapley value) of the McFadden pseudo-R^2 from a logit model of the allowance decision, based on the specification in Column (6) of Table 5 (pseudo-$R^2 \approx 0.46$). Demographics include predicted race and gender dummies and age; adjudicator ID / workload includes individual adjudicator dummies, the number of applicants on the same day per adjudicator, and year-month fixed effects. Sample is the control group with a visible charge.

Figure 4: Predicted Probability of Allowance and Observed Allowance Rates

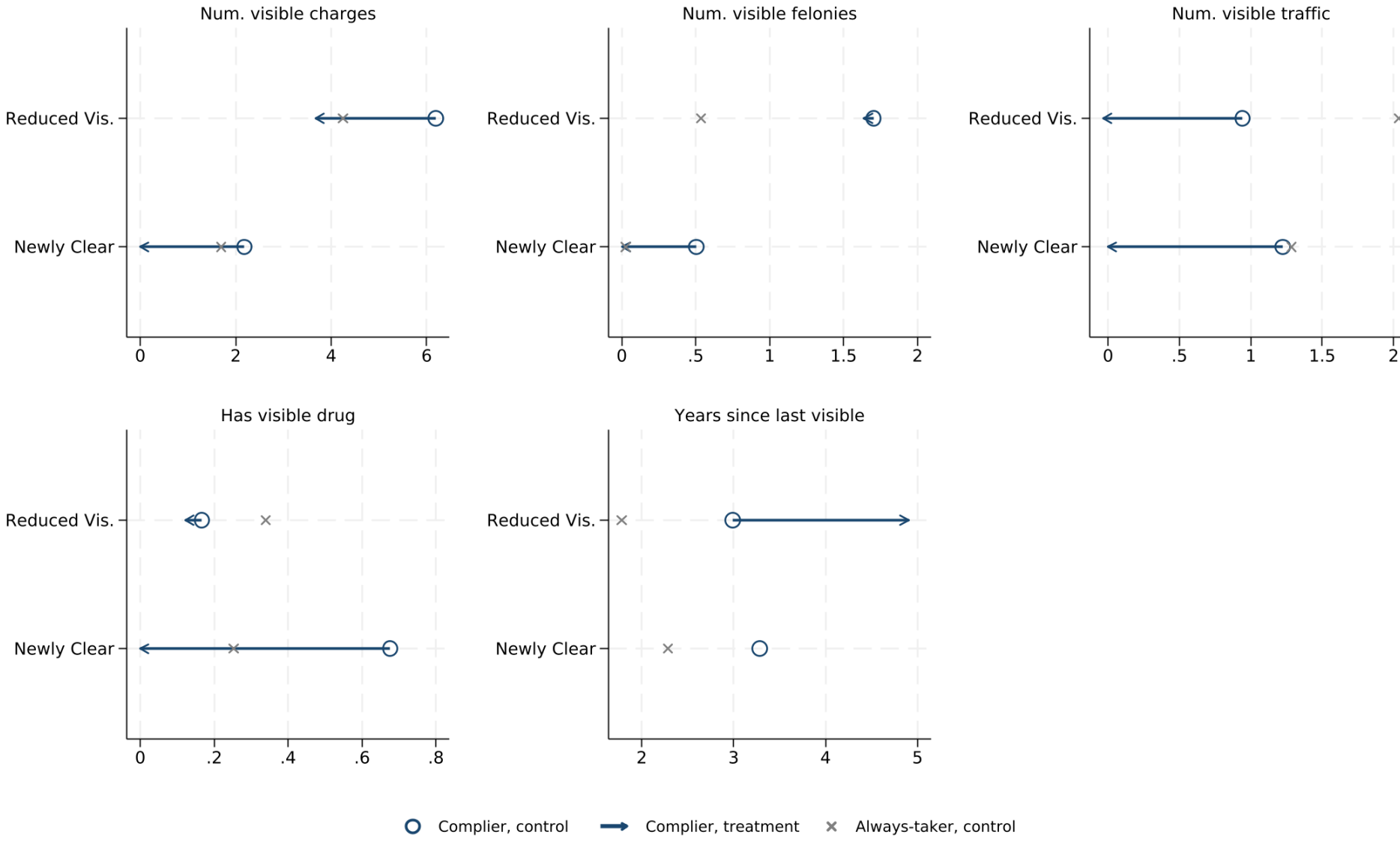
(a) Distribution of predicted allowance probabilities

(b) Observed allowance by predicted probability decile



Note: The sample consists of control group applicants with at least one charge visible to the adjudicator. Predicted probabilities are ten-fold cross-fitted estimates from a logistic regression of allowance on characteristics of the visible record: years since the most recent charge and its square, number of charges and its square, any pending charge, any felony, number and share of felony charges, and indicators for drug, violent, property, and traffic charges. Panel A shows the distribution of predicted probabilities. Panel B groups applicants into deciles of the predicted-probability distribution and plots the observed allowance rate against the mean predicted probability in each decile. Vertical bars show 95 percent confidence intervals; the dashed line denotes perfect calibration.

Figure 5: What Treatment Removes from Compliers' Visible Records



Note: For compliers, each panel shows the mean of one visible-record dimension under the control matrix (hollow circle) and the treatment matrix (arrowhead); the gray × is the always-taker mean under control. Complier means are estimated as in Section 6.3; newly clear compliers' records are fully hidden under treatment (arrows to zero) while reduced visibility compliers' are only partially reduced (fewer charges, but felonies largely retained). Axis scales differ across panels. Full results for all visible-record characteristics are in Table 14.

Table 1: Marginal Changes to Treatment Rule-Based Visibility Matrix

Charge Type	**Subtype**	**Level**	**Threshold**		**# Charges**
			C	**T**	**Newly Hidden**
Vehicles & Traffic	License & Registration	Misd	3	0	11,730
Vehicles & Traffic	Unsafe Operation	Misd	7	0	7,848
Drugs & Alcohol	Driving Under the Influence (DUI)	Misd	7	0	5,743
Vehicles & Traffic	Speeding	Misd	3	0	3,496
Vehicles & Traffic	Vehicle Equipment	Misd	3	0	1,707
Criminal Intent	Court Orders	Misd	7	3	1,626
Drugs & Alcohol	Drugs - Possession/Use	Felony	7	3	1,555
Criminal Intent	Obstruction	Misd	7	3	1,321
Drugs & Alcohol	Driving Under the Influence (DUI)	Felony	7	0	476
Theft & Property	Vandalism & Mischief	Misd	7	3	394
Violence	Disorderly Behavior	Misd	7	3	185
Drugs & Alcohol	Marijuana Possession/Use	Felony	7	3	129
Criminal Intent	Conspiracy	Misd	7	3	62
Criminal Intent	Criminal Tools	Misd	7	3	48
Criminal Intent	Accessory	Misd	7	3	22
Criminal Intent	Organized Crime	Misd	7	3	11
Drugs & Alcohol	Alcohol & Tobacco	Felony	7	3	1

Note: Categories and subcategories of charges whose thresholds differ between the control and treatment matrix are listed above. The threshold is measured in years since the charges adjudication and indicates how long a certain type of charges remains visible, e.g. 7 means charges 7 years and younger at the time of the background check are shown, but charges older than 7 years are not. They are sorted by the number of charges affected during the experimental period. The full control rule-based visibility matrix can be found in Appendix A.

Table 2: Balance and Sample Criminal Record Characteristics

	Mean (T)	Mean (C)	p-value
Full Sample			
Any criminal charge	0.143	0.142	0.312
Number of charges	0.584	0.583	0.968
Has Felony	0.044	0.045	0.147
Always Clear	0.039	0.037	0.142
Newly Clear	0.046	0.044	0.092
Reduced Visibility	0.025	0.026	0.161
Same Visibility	0.034	0.034	0.750
(pred.) White	0.431	0.430	0.833
(pred.) Black	0.140	0.137	0.054
(pred.) Hispanic	0.211	0.214	0.062
(pred.) API/AIAN	0.023	0.023	0.624
(pred.) Race unknown	0.195	0.195	0.841
(pred.) Female	0.291	0.291	0.953
(pred.) Male	0.607	0.608	0.699
(pred.) Gender unknown	0.102	0.101	0.474
Candidate age (years)	35.433	35.434	0.976
Observations	94,953	94,566	
Joint p-value			0.279
Conditional on Any Charge			
Number of charges	4.070	4.113	0.489
Number of felonies	0.871	0.897	0.339
Number violent	0.439	0.452	0.334
Number property	0.522	0.519	0.898
Number drug	0.820	0.865	0.022
Number traffic	1.536	1.502	0.252
Years since last charge	4.598	4.654	0.408
Avg Charge Age (Years)	6.422	6.509	0.240
Always Clear	0.270	0.264	0.264
Newly Clear	0.321	0.313	0.175
Reduced Visibility	0.173	0.182	0.050
Same Visibility	0.237	0.241	0.377
(pred.) White	0.534	0.528	0.289
(pred.) Black	0.227	0.226	0.946
(pred.) Hispanic	0.137	0.139	0.688
(pred.) API/AIAN	0.009	0.009	0.554
(pred.) Race unknown	0.093	0.098	0.107
(pred.) Female	0.210	0.208	0.682
(pred.) Male	0.728	0.732	0.515
(pred.) Gender unknown	0.062	0.061	0.609
Candidate age (years)	35.965	35.917	0.684
Observations	13,624	13,415	
Joint p-value			0.239

Note: Each variable is based on the full criminal record (whether visible or not). Panel A includes 0s for anyone without a criminal record. The joint p-values report F-tests of the null that all listed variables are jointly equal across treatment and control groups within each panel, based on regressions of treatment assignment on the variables with robust standard errors. A subfelony is either a misdemeanor or petty

Table 3: Selected Examples

Record type	Most recent charge	N	Any contact	Any review	Any incident
1 drug felony	6+ years ago	72	0.444	0.208	0.000
1 drug misdemeanor	< 1 year ago	209	0.507	0.225	0.029
1 drug misdemeanor	6+ years ago	430	0.542	0.193	0.016
1 property misdemeanor	6+ years ago	112	0.464	0.188	0.045
1 traffic misdemeanor	3–< 6 years ago	957	0.582	0.193	0.021
1 violent misdemeanor	3–< 6 years ago	167	0.587	0.204	0.012
3–5 charges including a drug felony	3–< 6 years ago	73	0.479	0.137	0.041
6–7 mixed charges	< 1 year ago	156	0.519	0.128	0.019

Note: The table reports 12-month outcomes for applicants allowed onto the platform with selected criminal record characteristics. Record type summarizes the number and primary type of charges in the record; records with multiple charges may also contain other types of charges. Recency refers to the most recent charge in the record. Outcome columns report the share of experiencing each outcome within 12 months. Each row contains at least 50 approved applicants.

Table 4: Allowance by Criminal Record and Visibility: Status Quo Entrant Sample

Applicant group	N	Share allowed
No legally reportable charge	81,151	0.999
Any legally reportable charge	13,415	0.658
No visible charge	3,540	0.999
At least one visible charge	9,875	0.536

Note: The table reports allowance rates among applicants assigned to the control visibility rules. The first two rows distinguish applicants with and without a legally reportable criminal charge. Among applicants with a charge, the indented rows distinguish those with no visible charge from those with at least one visible charge. Applicants with at least one visible charge are subject to human adjudication.

Table 5: Allowed on Platform by Visible Criminal Record Characteristics

	(1) Visible record	(2) + Demographics	(3) + Process variables
Years since last visible charge	0.026***	0.025***	0.025***
	(0.003)	(0.003)	(0.003)
Years since last visible charge squared	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)
Number of visible charges	-0.020***	-0.021***	-0.021***
	(0.003)	(0.003)	(0.003)
Number of visible charges squared	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)
Has visible pending charge	-0.020+	-0.018+	-0.018+
	(0.010)	(0.010)	(0.010)
Has visible felony	-0.401***	-0.403***	-0.401***
	(0.021)	(0.021)	(0.021)
Share of visible felonies	-0.047*	-0.047*	-0.049*
	(0.024)	(0.024)	(0.024)
Has visible violent	-0.207***	-0.206***	-0.206***
	(0.011)	(0.011)	(0.011)
Has visible property	-0.259***	-0.256***	-0.256***
	(0.012)	(0.012)	(0.012)
Has visible drug	0.054***	0.052***	0.052***
	(0.010)	(0.010)	(0.010)
Has visible traffic	0.203***	0.204***	0.203***
	(0.011)	(0.010)	(0.011)
(pred.) Black		-0.012	-0.011
		(0.009)	(0.009)
(pred.) Hispanic		0.008	0.008
		(0.011)	(0.011)
(pred.) API/AIAN		0.015	0.013
		(0.040)	(0.040)
(pred.) Race unknown		0.016	0.018
		(0.012)	(0.012)
(pred.) Female		-0.021*	-0.022*
		(0.009)	(0.009)
(pred.) Gender unknown		0.003	0.003
		(0.015)	(0.015)
Candidate age (years)		-0.000	-0.000
		(0.002)	(0.002)
Candidate age squared		0.000	0.000
		(0.000)	(0.000)
Constant	0.685***	0.664***	0.689***
	(0.012)	(0.043)	(0.046)
Mean allowance	0.536	0.536	0.536
Observations	9874	9874	9874
Adjusted R-squared	0.516	0.517	0.518
Process controls	No	No	Yes

Note: Dependent variable is an indicator for being allowed on the platform. Moving across columns, the specification adds criminal-record characteristics, then predicted demographics, then adjudicator identity and timing controls; all record characteristics are of the *visible* record the adjudicator sees. Sample is the control group with at least one visible charge. Robust standard errors in parentheses. Left-out race is predicted white; left-out gender is predicted male. Process variables include the ID numberes for the 3 adjudicators that handle most of background checks (+ "other" for a handful of others who seem to step in sometimes); the count of applicants processed by the same adjudicator on the same day; and year-month fixed effects.

Table 6: Association between Criminal Records (Any) and Platform Outcomes

	Any Contact			Num Contacts			Any Incident		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Any Criminal Record	-0.178***	-0.173***	-0.178***	-6.028***	-5.598***	-6.441***	0.006***	0.006**	0.004*
	(0.006)	(0.006)	(0.006)	(0.872)	(0.875)	(0.873)	(0.002)	(0.002)	(0.002)
(pred.) Black		-0.019***	-0.018***		-6.416***	-6.384***		0.004*	0.004**
		(0.005)	(0.005)		(0.777)	(0.787)		(0.001)	(0.001)
(pred.) Hispanic		0.019***	0.012**		-0.946	-0.910		-0.002+	-0.002*
		(0.004)	(0.004)		(0.706)	(0.721)		(0.001)	(0.001)
(pred.) API/AIAN		0.016	0.028**		4.253+	6.534**		-0.002	0.000
		(0.010)	(0.010)		(2.285)	(2.288)		(0.003)	(0.003)
(pred.) Race unknown		0.052***	0.050***		11.730***	11.423***		0.007***	0.006***
		(0.004)	(0.004)		(1.059)	(1.054)		(0.001)	(0.001)
(pred.) Female		-0.035***	-0.009*		-9.985***	-2.286**		-0.006***	0.004***
		(0.003)	(0.004)		(0.658)	(0.798)		(0.001)	(0.001)
(pred.) Gender unknown		0.005	0.017**		0.545	3.817**		-0.001	0.003+
		(0.005)	(0.005)		(1.177)	(1.200)		(0.002)	(0.002)
Candidate age (years)		0.006***	0.006***		1.224***	1.135***		-0.001**	-0.001**
		(0.001)	(0.001)		(0.138)	(0.138)		(0.000)	(0.000)
Candidate age squared		-0.000***	-0.000***		-0.014***	-0.013***		0.000*	0.000**
		(0.000)	(0.000)		(0.002)	(0.002)		(0.000)	(0.000)
Constant	0.722***	0.630***	0.616***	25.583***	3.030	2.048	0.018***	0.031***	0.030***
	(0.002)	(0.015)	(0.015)	(0.339)	(2.662)	(2.673)	(0.000)	(0.005)	(0.005)
Occupation FE	No	No	Yes	No	No	Yes	No	No	Yes
Observations	89891	89891	89891	89891	89891	89891	89891	89891	89891
R-squared	0.013	0.019	0.030	0.000	0.007	0.016	0.000	0.001	0.007

Note: Sample is the control group admitted to the platform. Each column is a regression of the platform outcome on an indicator for having any legally reportable criminal record (whether or not it is visible to the adjudicator); within each outcome, successive columns add predicted demographics and then occupation fixed effects. Left-out race is predicted white; left-out gender is predicted male.

Table 7: Platform Outcomes and Detailed Criminal Record Characteristics

	(1)	(2)	(3)
Years since last charge	0.00*	0.22	-0.00
	(0.00)	(0.42)	(0.00)
Years since last charge squared	-0.00	-0.00	0.00
	(0.00)	(0.02)	(0.00)
Number of charges	-0.00	-1.12***	0.00
	(0.00)	(0.29)	(0.00)
Number of charges squared	-0.00	0.01**	-0.00
	(0.00)	(0.01)	(0.00)
Has Pending Charge	-0.07***	-4.31**	0.01
	(0.02)	(2.07)	(0.01)
Has Felony	-0.06	-1.89	0.00
	(0.03)	(4.43)	(0.01)
Share of charges that are felonies	-0.07	1.01	-0.00
	(0.05)	(7.29)	(0.02)
Has Violent	-0.03**	-0.23	-0.00
	(0.02)	(2.74)	(0.01)
Has Property	-0.06***	-4.78**	0.00
	(0.02)	(2.31)	(0.01)
Has Drug	0.00	-0.56	-0.01**
	(0.01)	(1.94)	(0.00)
At least one traffic	0.04***	0.94	0.00
	(0.01)	(2.23)	(0.00)
(pred.) Black	-0.04***	0.14	-0.00
	(0.01)	(2.39)	(0.00)
(pred.) Hispanic	-0.00	-2.38	-0.01
	(0.02)	(1.90)	(0.00)
(pred.) API/AIAN	0.10*	4.64	0.00
	(0.05)	(6.50)	(0.02)
(pred.) Race unknown	0.05***	4.70	-0.00
	(0.02)	(2.93)	(0.01)
(pred.) Female	-0.06***	-2.54	0.00
	(0.01)	(2.21)	(0.00)
(pred.) Gender unknown	-0.03	4.00	0.02*
	(0.02)	(4.17)	(0.01)
Candidate age (years)	0.01	0.43	-0.00
	(0.00)	(0.51)	(0.00)
Candidate age squared	-0.00*	-0.00	0.00
	(0.00)	(0.01)	(0.00)
Constant	0.42***	12.17	0.04*
	(0.07)	(9.97)	(0.02)
Year-Month FE	Yes	Yes	Yes
Adjudicator Vars	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes
Observations	8831	8831	8831
R-squared	0.055	0.013	0.007

Note: Sample is the control group with any legally reportable charge, admitted to the platform. The three columns are separate regressions of, respectively, an indicator for any contact, the number of contacts, and an indicator for any incident, on the full set of criminal-record characteristics (visible or not), together with predicted demographics, adjudicator and year-month fixed effects, and occupation fixed effects. Robust standard errors in parentheses. Left-out race is predicted black; left-out gender is predicted male.

Table 8: Share of Platform-Outcome Variation Explained (R^2)

	Any Contact	Num Contacts	Any Incident
Record features	0.015	0.004	0.002
+ Demographics	0.028	0.008	0.003
+ Adjudicator & time	0.045	0.011	0.004
+ Occupation FE	0.065	0.024	0.017

Note: Each cell reports the R^2 from a regression of the column outcome on the row's covariates, added cumulatively down the rows. "Record" is the full legally-reportable criminal record; occupation fixed effects are for the applicant's stated main occupation, which the adjudicator does not observe. Sample is the control group with any charge, admitted to the platform.

Table 9: Prediction Summary Exercise

	Baseline	Additive	Flexible
Allowed (AUC)	0.500	0.922	0.922
Any contact (AUC)	0.500	0.540	0.541
Number of contacts (OOS R^2)	0.000	0.002	0.003
Any incident (AUC)	0.500	0.463	0.501

Note: Each row reports how well a model predicts the target out of sample, using a 70/30 train-test split. For the adjudication decision (Allowed) we predict from the visible criminal record among control applicants with a visible charge; for the platform outcomes we predict from the full criminal record among admitted control applicants with a record. The Additive model uses the record characteristics linearly (with squares for years since last charge and number of charges); the Flexible model is an elastic net that adds interactions among them. For the binary targets we report the area under the ROC curve (AUC), where 0.5 is the no-information baseline; for number of contacts we report the out-of-sample R^2, where 0 is the baseline (predicting the training-sample mean).

Table 10: Impact of Treatment on Access to Marketplace (First Stages)

		Record Changed by Treatment			Placebo	
	(1) Any Charge	(2) All	(3) Newly Clear	(4) Reduced Vis.	(5) Record Unchanged	(6) No Charges
Treatment	0.035***	0.047***	0.053***	0.035**	0.005	0.000
	(0.006)	(0.005)	(0.004)	(0.013)	(0.005)	(0.000)
Control mean	0.658	0.686	0.946	0.237	0.632	0.999
Robust F	38.91	85.31	225.03	7.93	0.87	0.48
N	27,039	13,364	8,572	4,792	13,675	162,480

Note: Each column reports the coefficient on treatment assignment from a regression of being allowed on the platform (not issued a post-adverse-action) on treatment. Column (1) includes anyone with a charge (visible to adjudicators or not). Columns (2)–(4) are the complier-bearing groups: (2) all "movers" (newly clear ∪ reduced visibility), (3) newly clear, and (4) reduced visibility. Columns (5)–(6) are placebo groups with no first stage: (5) applicants whose visible record is unchanged by treatment (always clear ∪ same visibility) and (6) applicants with no charges. The pooled columns (All Movers, Record Unchanged) absorb a stratum fixed effect; their reported Robust F is the test of the excluded instrument (treatment). Control mean is the control-group allowance rate.
+ p<0.10, * p<0.05, ** p<0.01, *** p < 0.001

Table 11: Reduced Form Effects

	(1) Any Charge	(2) All Movers	(3) Placebo
Any Contact	0.016***	0.019**	0.001
	(0.006)	(0.008)	(0.007)
Control mean	0.358	0.381	0.336
Any Review	0.005	0.010*	-0.005
	(0.004)	(0.006)	(0.005)
Control mean	0.118	0.126	0.109
Any Incident	-0.002	-0.002	-0.002
	(0.001)	(0.002)	(0.002)
Control mean	0.015	0.017	0.013
Number of Contacts	0.023	0.618	-1.016
	(0.746)	(1.072)	(1.020)
Control mean	12.876	13.612	12.154
Days BGC to Last Contact	2.411*	3.578*	-0.553
	(1.384)	(1.955)	(1.787)
Control mean	54.536	58.825	50.332
N	27,039	13,364	13,675

Note: Each row is a separate regression of the outcome on treatment assignment; the control mean is the regression constant. Column (1) includes all applicants with a charge; column (2) restricts to "movers" (newly clear ∪ reduced visibility), the groups whose visible record treatment can change; column (3) is a placebo of applicants whose visible record is unchanged by treatment (always clear ∪ same visibility). Outcomes are measured over the 12 months after the background check and are not conditioned on admission: they are set to 0 for applicants not allowed on the platform.
* p<0.10, ** p<0.05, *** p < 0.01

Table 12: Compliance Types by Treatment Assignment and Platform Access

	Treatment Assignment	
Platform Access	$Z = 0$ (Control)	$Z = 1$ (Treatment)
$D = 0$ (Not Allowed)	Never Taker / Complier	Never Taker
$D = 1$ (Allowed)	Always Taker	Always Taker / Complier

Table 13: Characteristics of (Full) Criminal Records by Compliance Type

	Overall (Movers)			Newly Clear		Reduced Visibility		
	(1) AT	(2) C	(3) NT	(4) AT	(5) C	(6) AT	(7) C	(8) NT
Number of charges	3.37	5.75	9.72	2.82	3.22	7.17	12.53	9.73
Charge types (any)								
Has Felony	0.07	0.53	0.72	0.04	0.52	0.28	0.55	0.73
Has Misdemeanor	0.77	0.63	0.95	0.75	0.60	0.92	0.70	0.95
Has Petty Offense	0.37	0.08	0.26	0.37	0.07	0.37	0.13	0.26
Has Violent	0.09	0.05	0.53	0.04	-0.09	0.45	0.43	0.54
Has Property	0.08	0.07	0.52	0.05	-0.03	0.27	0.35	0.52
Has Drug	0.35	0.61	0.68	0.33	0.68	0.49	0.43	0.68
Has Traffic	0.82	0.68	0.75	0.82	0.60	0.84	0.90	0.75
Has Pending Charge	0.17	0.35	0.37	0.15	0.34	0.29	0.37	0.37
Charge types (number)								
Number of felonies	0.18	1.26	2.59	0.07	0.87	0.87	2.28	2.60
Number violent	0.12	0.17	1.15	0.05	-0.12	0.66	0.94	1.15
Number property	0.14	0.69	1.55	0.08	0.19	0.53	2.04	1.56
Number drug	0.66	1.31	2.08	0.58	1.31	1.21	1.29	2.08
Number traffic	2.08	2.34	2.95	1.90	1.74	3.30	3.94	2.95
Recency								
Years since last charge	2.15	3.09	1.54	2.21	3.32	1.74	2.47	1.54
Demographics								
(pred.) White	0.52	0.60	0.59	0.52	0.50	0.52	0.85	0.59
(pred.) Black	0.20	0.28	0.23	0.20	0.34	0.26	0.14	0.23
(pred.) Hispanic	0.14	0.17	0.12	0.14	0.22	0.11	0.03	0.12
(pred.) Male	0.72	0.56	0.76	0.72	0.49	0.76	0.75	0.76
Candidate age (years)	33.98	34.48	34.59	33.93	33.23	34.35	37.83	34.58

Note: Sample is applicants with charges. AT = always-takers (control, admitted; $Z = 0$, $D = 1$); C = compliers; NT = never-takers (treatment, not admitted; $Z = 1$, $D = 0$). The Overall columns pool the two complier-bearing groups ("movers" = newly clear ∪ reduced visibility), and the Overall complier column is estimated with a stratum fixed effect. Complier means are Abadie (2003) κ-weighted (Section 6.3) and can fall outside a variable's natural range for rare characteristics in small complier subgroups.

Table 14: Compliers' Visible Records under Control and Treatment

	Overall (Movers)		Newly Clear		Reduced Visibility	
	(1) AT	(2) C	(3) AT	(4) C	(5) AT	(6) C
Has Visible Charge **Control**	1.00	1.00	1.00	1.00	1.00	1.00
Has Visible Charge **Treatment**	0.13	0.27	0.00	0.00	1.00	1.00
Num. Visible Charges **Control**	2.02	3.28	1.69	2.18	4.25	6.20
Num. Visible Charges **Treatment**	0.20	1.00	0.00	0.00	1.58	3.67
Charge types (any)						
Has Visible Felony **Control**	0.05	0.48	0.02	0.47	0.25	0.49
Has Visible Felony **Treatment**	0.03	0.11	0.00	0.00	0.23	0.39
Has Visible Misdemeanor **Control**	0.72	0.58	0.70	0.52	0.87	0.75
Has Visible Misdemeanor **Treatment**	0.09	0.20	0.00	0.00	0.69	0.74
Has Visible Petty Offense **Control**	0.34	0.18	0.34	0.18	0.34	0.18
Has Visible Petty Offense **Treatment**	0.02	0.03	0.00	0.00	0.13	0.10
Has Visible Violent **Control**	0.06	0.06	0.01	-0.05	0.40	0.34
Has Visible Violent **Treatment**	0.05	0.08	0.00	0.00	0.40	0.30
Has Visible Property **Control**	0.03	0.08	0.01	-0.01	0.16	0.31
Has Visible Property **Treatment**	0.02	0.06	0.00	0.00	0.13	0.21
Has Visible Drug **Control**	0.26	0.54	0.25	0.68	0.34	0.17
Has Visible Drug **Treatment**	0.01	0.03	0.00	0.00	0.06	0.12
Has Visible Traffic **Control**	0.77	0.57	0.77	0.58	0.76	0.56
Has Visible Traffic **Treatment**	0.01	-0.01	0.00	0.00	0.04	-0.03
Has Visible Pending Charge **Control**	0.16	0.29	0.14	0.29	0.28	0.30
Has Visible Pending Charge **Treatment**	0.02	0.09	0.00	0.00	0.16	0.34
Charge types (number)						
Num. Visible Felonies **Control**	0.09	0.83	0.03	0.50	0.54	1.70
Num. Visible Felonies **Treatment**	0.05	0.45	0.00	0.00	0.42	1.64
Num. Visible Violent Crimes **Control**	0.07	0.15	0.01	-0.03	0.51	0.62
Num. Visible Violent Crimes **Treatment**	0.06	0.13	0.00	0.00	0.50	0.49
Num. Visible Traffic Offenses **Control**	1.38	1.15	1.28	1.22	2.04	0.94
Num. Visible Traffic Offenses **Treatment**	0.01	-0.01	0.00	0.00	0.06	-0.03
Recency						
Years Since Last Visible **Control**	2.22	3.20	2.28	3.28	1.78	2.99
Years Since Last Visible **Treatment**	3.87	4.90			3.87	4.90

Note: Sample is applicants with charges. For each characteristic the table reports the potential *visible* record under the control matrix (Control) and the treatment matrix (Treatment), by compliance type. AT = always-takers (control, admitted); C = compliers. The Overall columns pool the movers (newly clear ∪ reduced visibility), with a stratum fixed effect on the complier column. Never-takers are omitted: their visible records are not informative about treatment's effect on the visible margin. Complier means are estimated as in Section 6.3 and can fall outside a variable's natural range for rare characteristics.

Table 15: Average Outcomes By Treatment Response

Among movers:	(1) Control (AT)	(2) Treatment (AT+C)	(3) Implied (C Mean)	(4) p (AT=C)
Any Contact	0.556	0.548	0.407	0.340
	(0.007)	(0.007)	(0.156)	
More than 10 Contacts	0.208	0.206	0.163	0.722
	(0.006)	(0.006)	(0.127)	
Any Incident	0.025	0.021	-0.039	0.178
	(0.002)	(0.002)	(0.048)	
N	4,552	5,007		

Note: Sample is "movers" (applicants with charges in the newly clear or reduced visibility groups). Column (1), AT, is the always-taker mean (control, admitted; $Z = 0$, $D = 1$). Column (2) is the mean among all treatment-admitted applicants ($Z = 1$, $D = 1$) — always-takers plus compliers, i.e. the observed pool under the relaxed policy. Column (3), C, is the complier mean (2SLS local average treatment effect, estimated with a stratum fixed effect; Section 6.3). Column (4) is the p-value for the test that the always-taker and complier means are equal. Outcomes are measured over the 12 months after the background check and are set to 0 for applicants not allowed on the platform. Complier means are imprecise where the first stage is modest in levels and can fall outside a variable's natural range. N is the number of admitted applicants in each arm. Robust standard errors in parentheses.
+ p<0.10, * p<0.05, ** p<0.01, *** p < 0.001

Table 16: Incumbent vs. Entrant Criminal Record Characteristics: Binary Shares

	(1) Incumbent	(2) Entrant	(3) p-value
Has Felony	0.260	0.313	0.000
Has Misdemeanor	0.793	0.802	0.073
Has Petty Offense	0.229	0.210	0.001
Has Violent	0.222	0.254	0.000
Has Property	0.188	0.235	0.000
Has Drug	0.366	0.389	0.000
Has Traffic	0.630	0.597	0.000
Has Sexual	0.013	0.015	0.098
Has Fraud	0.089	0.097	0.042
Has DUI	0.188	0.178	0.067
Has Theft	0.083	0.115	0.000
Has Burglary	0.031	0.044	0.000
Has Pending Charge	0.149	0.165	0.001

Note: Pooled comparison (no AT/C/NT or treatment split), both samples restricted to having at least one legally reportable charge. Incumbent column uses the incumbent sample; Entrant column uses the main entrant sample. p-value tests equality of means with robust standard errors, independent samples. `Has Assault` is omitted: the underlying charge category has zero variance in both samples.

Table 17: Incumbent vs. Entrant Criminal Record Characteristics: Counts and Severity

	(1) Incumbent	(2) Entrant	(3) p-value
Number of Charges	3.669	4.091	0.000
Number of Felonies	0.668	0.884	0.000
Number of Misdemeanors	2.418	2.646	0.000
Number of Petty Offenses	0.562	0.538	0.313
Number Violent	0.371	0.445	0.000
Number Property	0.367	0.521	0.000
Number Drug	0.732	0.843	0.000
Number Traffic	1.558	1.519	0.242
Number Sexual	0.023	0.026	0.448
Number Fraud	0.160	0.183	0.032
Number DUI	0.275	0.260	0.105
Number Theft	0.135	0.198	0.000
Number Pending	0.408	0.453	0.040
Years Since Last Charge	5.253	4.626	0.000
Share of Charges Felony	0.145	0.174	0.000

Note: Same sample and restrictions as Table 16. `Number of Assaults` is omitted for the same reason as `Has Assault` in that table.

Table 18: Pre-BGC Platform Performance, Removed vs. Retained Incumbents

	Any Contact (Pre)		Num Contacts (Pre)		Any Incident (Pre)	
	(1)	(2)	(3)	(4)	(5)	(6)
Removed	-0.014	-0.020	-3.289	-6.587*	0.003	-0.000
	(0.016)	(0.021)	(2.168)	(3.221)	(0.003)	(0.005)
Observations	3777	3777	3777	3777	3777	3777
Adjusted R^2	0.004	0.013	0.000	0.007	-0.000	-0.001

Note: Pooled sample (both experimental arms), final disposition (removed vs. retained) regardless of which arm produced it; visible-charge population. Odd columns include no controls beyond tenure; even columns add the full record and demographic controls. Robust standard errors in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p < 0.001

Table 19: Incumbent Criminal Record Characteristics by Compliance Type

	(1) Always Taker	(2) Complier	(3) Never Taker
Has Felony	0.06	0.49	0.68
Has Violent	0.05	0.44	0.59
Has Property	0.05	0.25	0.49
Has Drug	0.34	0.58	0.65
Has Traffic	0.82	0.83	0.77
Has Pending Charge	0.18	-0.09	0.39
Number of Charges	3.10	5.37	9.09
Years Since Last Charge	2.18	3.90	1.57
Share of Charges Felony	0.02	0.20	0.26

Note: Direct incumbent analog of Table 13, for comparison against that table's entrant AT/C/NT columns. Recovered via Abadie's (2003) kappa-weighting estimator, pooled with a stratum fixed effect. Uses the full criminal record (visible or not), matching the construction of Table 13. Sample is incumbent movers – charged incumbents whose visible record changes under the treatment matrix – who number 3,608 of the 56,347 providers in the incumbent sample (7,300, or 13.0%, have any legally reportable charge at all). The complier estimate for Has Pending Charge is outside the logical [0,1] range and reflects small-sample estimator instability rather than a real value; see main text.

Table 20: Incumbent Criminal Record Characteristics by Compliance Type, by Mover Subtype

	Overall (Movers)			Newly Clear		Reduced Visibility		
	(1) AT	(2) C	(3) NT	(4) AT	(5) C	(6) AT	(7) C	(8) NT
Has Felony	0.06	0.49	0.68	0.04	0.69	0.25	0.20	0.68
Has Violent	0.05	0.44	0.59	0.02	0.26	0.34	0.70	0.59
Has Property	0.05	0.25	0.49	0.04	0.15	0.17	0.38	0.49
Has Drug	0.34	0.58	0.65	0.32	0.70	0.49	0.41	0.65
Has Traffic	0.82	0.83	0.77	0.82	0.69	0.83	1.03	0.77
Has Pending Charge	0.18	-0.09	0.39	0.17	-0.18	0.32	0.04	0.39
Number of Charges	3.10	5.37	9.09	2.70	5.48	7.27	5.23	9.09
Years Since Last Charge	2.18	3.90	1.57	2.23	5.11	1.70	2.25	1.57
Share of Charges Felony	0.02	0.20	0.26	0.01	0.35	0.09	-0.01	0.26

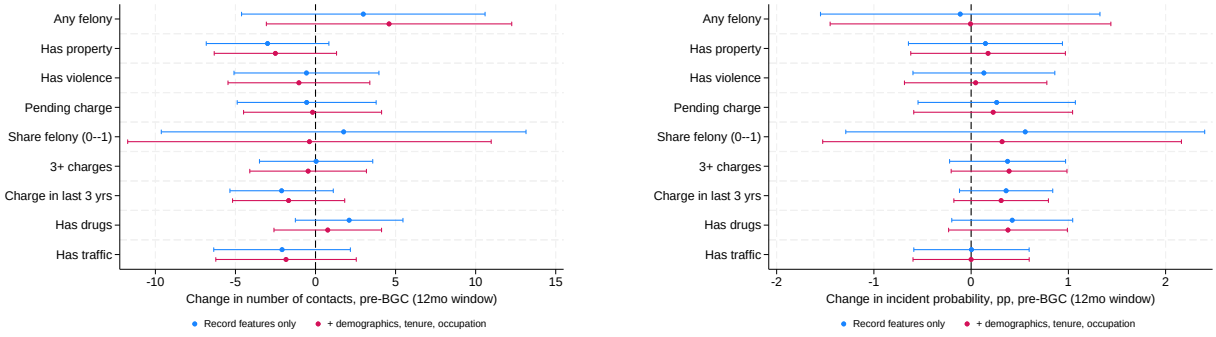
Note: Same as Table 19, additionally split by mover subtype. Newly Clear has no never-taker column since treatment mechanically clears these applicants. Reduced Visibility retains human review under both arms and so has a never-taker column.

Table 21: Incumbent Post-BGC Platform Outcomes by Compliance Type

	(1) Always Taker	(2) Treatment (AT+C)	(3) Complier (LATE)	(4) p (AT=C)
Any Contact	0.566	0.594	1.021	0.136
	(0.014)	(0.013)	(0.305)	
Any Incident	0.032	0.027	-0.046	0.441
	(0.005)	(0.004)	(0.101)	
Log(Contacts)	2.297	2.406	3.313	0.197
	(0.063)	(0.060)	(0.784)	
Reviews per Contact	0.042	0.059	0.206	0.004
	(0.003)	(0.005)	(0.057)	
Incidents per Contact	0.001	0.001	0.001	0.996
	(0.000)	(0.000)	(0.005)	
Average Rating	4.546	4.507	4.350	0.552
	(0.055)	(0.051)	(0.325)	

Note: Direct incumbent analog of Table 15 (unconditional rows) and its conditional-outcomes companion, for comparison against entrants' post-BGC outcomes – since entrants have no pre-BGC platform history, post-BGC outcomes are the only outcome comparison available on equal footing across the two samples. Always Taker and Treatment (AT+C) are direct sample means among movers, allowed and not allowed pooled correctly via zero-filling; Complier (LATE) is recovered via `ivregress 2sls` of the (zero-filled) outcome on allowed, instrumented by treatment, with a Newly Clear stratum fixed effect. Complier estimates for Any Contact and Any Incident fall outside the logical range for these binary/rate outcomes in both the incumbent and entrant samples alike, consistent with small-sample estimator instability rather than a real difference; see main text.

Figure 6: Record Features Do Not Predict Pre-BGC Platform Outcomes



(a) Num. Contacts, Pre-BGC (12mo)

(b) Any Incident, Pre-BGC (12mo)

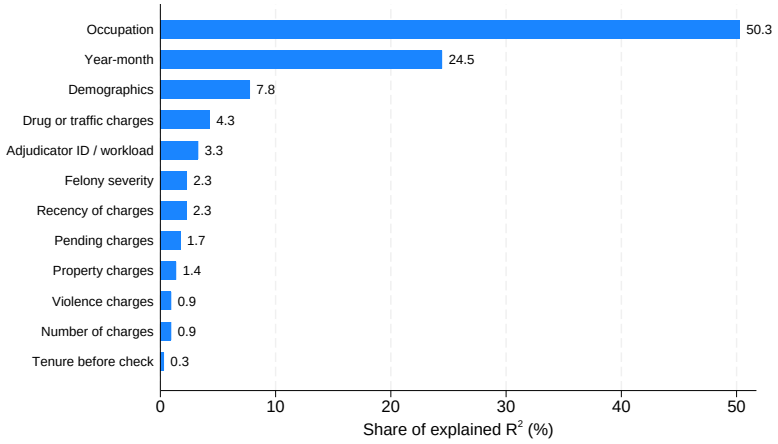
Note: Full charged incumbent population (allowed and removed alike), 12-month pre-BGC window. Points show coefficients from separate regressions of the outcome on each individual record feature; confidence intervals cross zero throughout, record-only or with full controls.

Table 22: Does Having Any Criminal Charge Predict Pre-BGC Platform Activity?

	Any Contact (Pre)		Num Contacts (Pre)		Ar
	(1)	(2)	(3)	(4)	
Has criminal charges (any legally reportable charge)	0.227***	0.225***	0.855	1.051	0.0
	(0.006)	(0.006)	(0.834)	(0.856)	(0
Tenure Before Check (Days)		-0.000***		-0.000	
		(0.000)		(0.000)	
(pred.) Black		-0.031***		-2.011*	
		(0.005)		(0.808)	
(pred.) Hispanic		0.083***		10.624***	
		(0.005)		(0.924)	
(pred.) API/AIAN		-0.039***		-1.451	
		(0.010)		(1.354)	
(pred.) Race unknown		0.025***		3.819***	
		(0.005)		(0.801)	
(pred.) Female		-0.097***		-8.693***	
		(0.004)		(0.573)	
(pred.) Gender unknown		-0.048***		-2.906**	
		(0.007)		(1.039)	
Candidate age (years)		0.006***		1.012***	
		(0.001)		(0.131)	
Candidate age squared		-0.000***		-0.010***	
		(0.000)		(0.001)	
Constant	0.238***	0.109***	14.268***	-8.328**	0.0
	(0.002)	(0.019)	(0.314)	(2.598)	(0
Observations	56347	56347	56347	56347	5
Adjusted R^2	0.030	0.051	-0.000	0.009	0

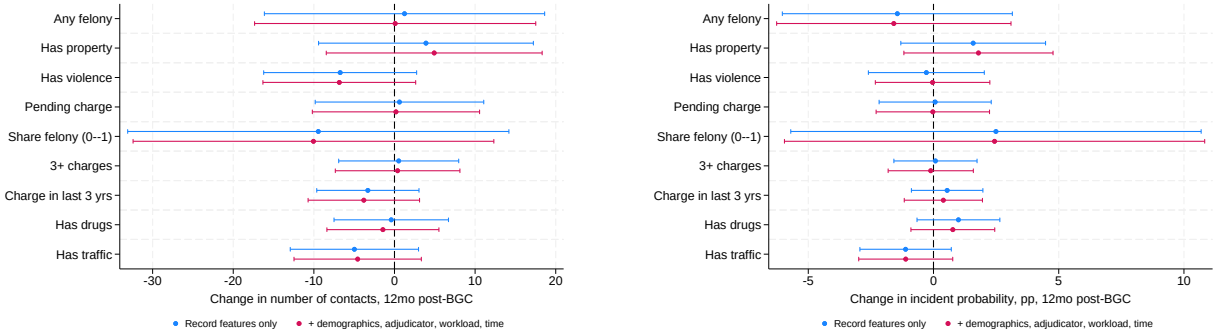
Note: Full incumbent sample. Outcomes measured over the 12 months before the background check. Odd columns regress the outcome on simply having any legally reportable charge; even columns add tenure and demographic controls. Robust standard errors in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p < 0.001

Figure 7: Dominance Decomposition, Any Incident (Pre-BGC, 12mo)



Note: Bars show each block's general-dominance share of R^2 from a linear probability model of the outcome, full charged population, 12-month pre-BGC window. Occupation alone accounts for roughly 50% of the explained variation; every record-feature category combined accounts for roughly 14%.

Figure 8: The Pattern Holds After Reinstatement, Too

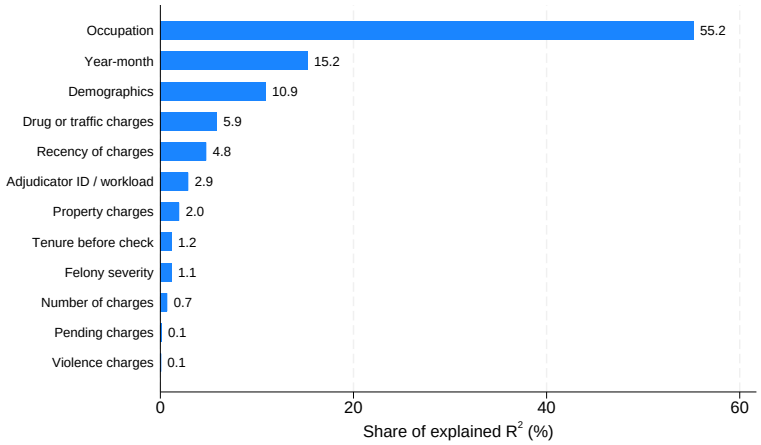


(a) Num. Contacts, 12mo Post-BGC

(b) Any Incident, 12mo Post-BGC

Note: Retained incumbents only (treatment==0 & allowed==1), 12-month post-BGC window. Same specification as Figure 6; confidence intervals again cross zero throughout.

Figure 9: Dominance Decomposition, Any Incident (12mo Post-BGC)



Note: Same decomposition as Figure 7, retained-only 12-month post-BGC sample. Occupation and adjudication year-month together account for more than 70% of the explained variation; every record-feature category remains in the single digits.

Appendix

A Description of Rule-Based Visibility Matrices

The company's pre-experiment and control list of hidden charges is below. The columns show the year thresholds at or above which a charge with a particular severity would be hidden. For example, an animal cruelty petty offense would be hidden if it had been 7 or more years since that charge at the time of the background check. Where there is no number (such as under violent felonies), these offenses would never be hidden, regardless of the time since the charge. For example, homicide charges are never hidden.

Table A.1: Pre-Experiment and Control List of Hidden Charges

Charge Category	Charge Subcategory	Year Threshold		
		Petty Offense	**Misdemeanor Offense**	**Felony Offense**
Violence	Animal Cruelty	7	7	—
	Disorderly Behavior	7	7	—
	Weapons & Endangerment	7	7	—
	Abduction & Restraint	7	7	—
	Harassment & Threats	7	7	—
	Child & Elder Abuse	7	7	—
	Assault & Battery	7	7	—
Vehicles & Traffic	Unsafe Operation	7	7	7
	Parking	0	0	7
	Vehicle Equipment	3	3	7
	License & Registration	3	3	7
	Speeding	3	3	7
Theft & Property	Trespassing	0	0	7
	Vandalism & Mischief	7	7	7
	Possession of Stolen Property	7	7	7
	Petty Theft	7	7	7
	Theft	7	7	7
	Arson	7	7	—
	Burglary	7	7	7
	Robbery	7	7	—
Statutory	Miscellaneous Citations & Violations	0	0	7
	Gambling	0	0	7
	Animal Ordinances	0	0	—
	Fish & Game	0	0	7
	Safety & Zoning	0	0	7
	Public Nuisance	0	0	7
	Custody & Support	0	0	7
Sexual	Prostitution	7	7	—
	Sexual Abuse	7	7	—
	Lewd Behavior	7	7	—

Charge Category	Charge Subcategory	Year Threshold		
		Petty Offense	**Misdemeanor Offense**	**Felony Offense**
Security	Treason	7	7	—
	Terrorism	7	7	—
	Immigration	7	7	—
Homicide	Unintentional Killing	—	—	—
	Attempted Homicide	—	—	—
	Intentional Homicide	—	—	—
Fraud & Deception	Business & Tax	7	7	7
	Worthless Check	7	7	7
	Forgery	7	7	7
	Identity Theft & Impersonation	7	7	7
	Bribery & Corruption	7	7	7
	Cyber Crimes	7	7	7
	Embezzlement	7	7	7
	Fraud	7	7	7
Drugs & Alcohol	Driving under the Influence (DUI)	7	7	7
	Marijuana Possession/Use	0	0	7
	Alcohol & Tobacco	0	0	7
	Drugs-Sale & Manufacture	0	0	7
	Drugs-Possession/Use	0	0	7
Criminal Intent	Court Orders	7	7	7
	Obstruction	7	7	—
	Organized Crime	7	7	—
	Criminal Tools	7	7	7
	Accessory	7	7	—
	Conspiracy	7	7	—

B Additional Figures and Tables

Table B.1: Examples

Record type	Most recent charge	Any pending	N	Any contact	Any review	Any incident
1 drug felony	3–< 6 years ago	No	57	0.439	0.246	0.018
1 drug felony	6+ years ago	No	72	0.444	0.208	0.000
1 drug misdemeanor	< 1 year ago	No	209	0.507	0.225	0.029
1 drug misdemeanor	1–< 3 years ago	No	379	0.536	0.214	0.013
1 drug misdemeanor	3–< 6 years ago	No	564	0.603	0.199	0.018
1 drug misdemeanor	6+ years ago	No	430	0.542	0.193	0.016
1 older misdemeanor	6+ years ago	No	195	0.554	0.195	0.015
1 property misdemeanor	6+ years ago	No	112	0.464	0.188	0.045
1 recent misdemeanor	1–< 3 years ago	No	104	0.510	0.183	0.019
1 traffic misdemeanor	< 1 year ago	Yes	197	0.609	0.183	0.025
1 traffic misdemeanor	1–< 3 years ago	No	1045	0.589	0.198	0.014
1 traffic misdemeanor	3–< 6 years ago	No	957	0.582	0.193	0.021
1 violent misdemeanor	3–< 6 years ago	No	167	0.587	0.204	0.012
2 charges including a drug misdemeanor	< 1 year ago	No	124	0.629	0.290	0.024
2 charges including a drug misdemeanor	1–< 3 years ago	No	213	0.493	0.183	0.033
2 charges including a property misdemeanor	6+ years ago	No	60	0.500	0.150	0.033
2 charges including a violent misdemeanor	1–< 3 years ago	No	51	0.471	0.118	0.020
2 charges including a violent misdemeanor	6+ years ago	No	94	0.585	0.149	0.000
3–5 charges including a drug felony	3–< 6 years ago	No	73	0.479	0.137	0.041
3–5 charges including a drug felony	6+ years ago	No	110	0.500	0.173	0.009
3–5 charges including a drug misdemeanor	< 1 year ago	Yes	125	0.512	0.200	0.008
3–5 charges including a drug misdemeanor	3–< 6 years ago	No	251	0.566	0.187	0.032
3–5 charges including a property misdemeanor	1–< 3 years ago	No	54	0.463	0.167	0.000
3–5 charges including a property misdemeanor	6+ years ago	No	76	0.447	0.171	0.013
3–5 charges including a violent misdemeanor	< 1 year ago	No	65	0.708	0.277	0.046
3–5 charges including a violent misdemeanor	1–< 3 years ago	No	115	0.417	0.130	0.035
3–5 charges including a violent misdemeanor	3–< 6 years ago	No	145	0.648	0.269	0.041
6–7 mixed charges	< 1 year ago	No	156	0.519	0.128	0.019
6–7 mixed charges	< 1 year ago	Yes	125	0.504	0.184	0.024
6–7 mixed charges	1–< 3 years ago	No	212	0.453	0.151	0.019
6–7 mixed charges	3–< 6 years ago	No	173	0.520	0.145	0.023
6–7 mixed charges	6+ years ago	No	173	0.607	0.145	0.046

Note: The table reports 12-month outcomes for applicants allowed onto the platform with selected crimina record characteristics. Record type

Table B.2: Balance By Treatment Response Type: Always Clear

	Mean (T)	Mean (C)	p-value
Number of charges	2.355	2.342	0.809
At least one felony	0.117	0.115	0.767
At least one misdemeanor	0.844	0.859	0.044
At least one petty	0.147	0.136	0.113
At least one pending charge	0.034	0.032	0.512
At least one violent	0.102	0.096	0.284
At least one property	0.142	0.139	0.658
At least one drug	0.326	0.328	0.811
At least one traffic	0.590	0.575	0.148
Number of felonies	0.246	0.252	0.727
Number of convictions	2.189	2.190	0.996
Number violent	0.130	0.126	0.671
Number property	0.231	0.221	0.545
Number drug	0.535	0.536	0.967
Years first to last charge	2.440	2.308	0.184
Years since last charge	9.859	10.008	0.300
Avg Charge Age (Years)	11.134	11.208	0.623
Num distinct charge dates	1.872	1.866	0.866
Observations	4,809	4,638	
Joint p-value			0.401

Notes: Each variable is based on the full criminal record (whether visible or not). The joint p-values report F-tests of the null that all listed variables are jointly equal across treatment and control groups within each panel, based on regressions of treatment assignment on the variables with robust standard errors.

Table B.3: Balance By Treatment Response Type: Newly Clear

	Mean (T)	Mean (C)	p-value
Number of charges	2.837	2.844	0.910
At least one felony	0.065	0.063	0.649
At least one misdemeanor	0.741	0.751	0.217
At least one petty	0.354	0.356	0.852
At least one pending charge	0.161	0.155	0.376
At least one violent	0.030	0.034	0.234
At least one property	0.047	0.053	0.145
At least one drug	0.349	0.349	0.963
At least one traffic	0.813	0.812	0.929
Number of felonies	0.115	0.118	0.806
Number of convictions	2.228	2.279	0.377
Number violent	0.037	0.044	0.184
Number property	0.083	0.081	0.780
Number drug	0.608	0.623	0.508
Years first to last charge	2.786	2.824	0.704
Years since last charge	2.284	2.282	0.959
Avg Charge Age (Years)	3.686	3.684	0.975
Num distinct charge dates	2.193	2.220	0.496
Observations	5,662	5,472	
Joint p-value			0.141

Notes: Each variable is based on the full criminal record (whether visible or not). The joint p-values report F-tests of the null that all listed variables are jointly equal across treatment and control groups within each panel, based on regressions of treatment assignment on the variables with robust standard errors.

Table B.4: Balance By Treatment Response Type: Reduced Visibility

	Mean (T)	Mean (C)	p-value
Number of charges	9.047	8.909	0.495
At least one felony	0.604	0.616	0.339
At least one misdemeanor	0.938	0.940	0.736
At least one petty	0.286	0.272	0.221
At least one pending charge	0.347	0.331	0.213
At least one violent	0.515	0.523	0.546
At least one property	0.449	0.455	0.623
At least one drug	0.622	0.643	0.088
At least one traffic	0.782	0.768	0.197
Number of felonies	2.092	2.038	0.539
Number of convictions	7.054	6.932	0.507
Number violent	1.019	1.039	0.671
Number property	1.259	1.165	0.187
Number drug	1.814	1.884	0.276
Years first to last charge	7.162	6.985	0.324
Years since last charge	1.594	1.651	0.265
Avg Charge Age (Years)	4.844	4.770	0.421
Num distinct charge dates	5.537	5.479	0.598
Observations	2,929	2,998	
Joint p-value			0.593

Notes: Each variable is based on the full criminal record (whether visible or not). The joint p-values report F-tests of the null that all listed variables are jointly equal across treatment and control groups within each panel, based on regressions of treatment assignment on the variables with robust standard errors.

Table B.5: Balance By Treatment Response Type: Same Visibility

	Mean (T)	Mean (C)	p-value
Number of charges	3.909	3.967	0.561
At least one felony	0.628	0.640	0.258
At least one misdemeanor	0.716	0.708	0.442
At least one petty	0.053	0.055	0.746
At least one pending charge	0.200	0.182	0.038
At least one violent	0.525	0.525	0.967
At least one property	0.408	0.401	0.541
At least one drug	0.305	0.330	0.020
At least one traffic	0.210	0.213	0.737
Number of felonies	1.646	1.678	0.580
Number of convictions	3.061	3.168	0.259
Number violent	0.898	0.891	0.807
Number property	0.827	0.850	0.558
Number drug	0.656	0.701	0.151
Years first to last charge	4.044	4.308	0.074
Years since last charge	4.334	4.383	0.687
Avg Charge Age (Years)	6.386	6.597	0.140
Num distinct charge dates	2.517	2.548	0.579
Observations	4,009	4,036	
Joint p-value			0.431

Notes: Each variable is based on the full criminal record (whether visible or not). The joint p-values report F-tests of the null that all listed variables are jointly equal across treatment and control groups within each panel, based on regressions of treatment assignment on the variables with robust standard errors.

Table B.6: Allowed on Platform by Visible Criminal Record Characteristics: Alternative Specs

	(1) Main specification	(2) + Category counts	(3) + Record span
Years since last visible charge	0.025***	0.026***	0.026***
	(0.003)	(0.003)	(0.003)
Years since last visible charge squared	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)
Visible criminal record span in years			0.004**
			(0.001)
Number of visible charges	-0.021***	-0.062***	-0.063***
	(0.003)	(0.008)	(0.008)
Number of visible charges squared	0.000***	0.000**	0.000**
	(0.000)	(0.000)	(0.000)
Has visible felony	-0.401***	-0.345***	-0.353***
	(0.021)	(0.021)	(0.021)
Number of visible felonies		0.044***	0.043***
		(0.006)	(0.006)
Share of visible felonies	-0.049*	-0.218***	-0.212***
	(0.024)	(0.029)	(0.029)
Number of visible misdemeanors		-0.010**	-0.010**
		(0.004)	(0.004)
Has visible violent	-0.206***	-0.213***	-0.216***
	(0.011)	(0.013)	(0.013)
Number of visible violence charges		0.025***	0.024***
		(0.006)	(0.006)
Has visible property	-0.256***	-0.260***	-0.264***
	(0.012)	(0.014)	(0.014)
Number of visible property charges		0.021***	0.022***
		(0.006)	(0.006)
Has visible drug	0.052***	0.082***	0.079***
	(0.010)	(0.013)	(0.013)
Number of visible drugs charges		-0.001	0.001
		(0.008)	(0.008)
Has visible traffic	0.203***	0.148***	0.146***
	(0.011)	(0.012)	(0.012)
Number of visible traffic charges		0.051***	0.051***
		(0.006)	(0.006)
Has visible pending charge	-0.018+	-0.023	-0.024+
	(0.010)	(0.014)	(0.014)
Number of visible pending charges		0.010+	0.011+
		(0.006)	(0.006)
Number of visible convictions		0.010*	0.009*
		(0.004)	(0.004)
(pred.) Black	-0.011	-0.011	-0.012
	(0.009)	(0.009)	(0.009)
(pred.) Hispanic	0.008	0.010	0.011
	(0.011)	(0.011)	(0.011)
(pred.) API/AIAN	0.013	0.014	0.014
	(0.040)	(0.040)	(0.040)
(pred.) Race unknown	0.018	0.016	0.017
	(0.012)	(0.011)	(0.011)
(pred.) Female	-0.022*	-0.023**	-0.021*

Table B.7: Any Contact for those Allowed on Platform by Record Chars (Control Only)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Years since last charge	0.00	-0.00	0.00	0.00*	0.00**	0.00*	0.00*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Number of charges	-0.01***	-0.01*	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Num distinct charge dates		0.00	0.00	0.00	0.00	0.00	0.00
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Has Pending Charge		-0.07***	-0.08***	-0.08***	-0.07***	-0.07***	-0.07***
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Has Felony			-0.07*	-0.04	-0.04	-0.05	-0.05
			(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Share of charges that are felonies			-0.10+	-0.08+	-0.09+	-0.08	-0.07
			(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Has Violent				-0.04*	-0.04*	-0.04*	-0.03*
				(0.02)	(0.02)	(0.02)	(0.02)
Has Property				-0.07***	-0.06***	-0.06***	-0.06**
				(0.02)	(0.02)	(0.02)	(0.02)
Has Drug				0.01	0.00	0.00	0.00
				(0.01)	(0.01)	(0.01)	(0.01)
At least one traffic				0.04**	0.04**	0.04**	0.04**
				(0.01)	(0.01)	(0.01)	(0.01)
(pred.) Black					-0.04**	-0.04***	-0.04**
					(0.01)	(0.01)	(0.01)
(pred.) Hispanic					-0.00	0.00	-0.00
					(0.02)	(0.02)	(0.02)
(pred.) API/AIAN					0.08	0.08	0.10+
					(0.05)	(0.05)	(0.05)
(pred.) Race unknown					0.05**	0.05**	0.05**
					(0.02)	(0.02)	(0.02)
(pred.) Female					-0.11***	-0.11***	-0.06***
					(0.01)	(0.01)	(0.01)
(pred.) Gender unknown					-0.04+	-0.05*	-0.03
					(0.02)	(0.02)	(0.02)
Candidate age (years)					0.01+	0.01+	0.01+
					(0.00)	(0.00)	(0.00)
Candidate age squared					-0.00*	-0.00*	-0.00*
					(0.00)	(0.00)	(0.00)
Adjudicator 1						0.02	0.03
						(0.02)	(0.02)
Adjudicator 2						-0.01	-0.00
						(0.01)	(0.01)
Other Adjudicator						-0.01	-0.01
						(0.04)	(0.04)
Num Applicants Same Day/Adj						-0.00	-0.00
						(0.00)	(0.00)
Constant	0.56***	0.57***	0.57***	0.54***	0.48***	0.43***	0.42***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.06)	(0.07)	(0.07)
Occupation FE	No	No	No	No	No	No	Yes
Observations	8832	8832	8832	8832	8832	8831	8831
R-squared	0.002	0.004	0.010	0.014	0.026	0.041	0.054

Note: Sample is control group with any (legally reportable) charge that is allowed on platform. Dependent variable is number of contacts (incl 0 if no contacts). Robust standard errors in parentheses. The Final column also includes FE for the year-month

Table B.8: Num Contacts for those Allowed on Platform by Record Chars (Control Only)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Years since last charge	0.28	0.23	0.26	0.29	0.24	0.22	0.22
	(0.19)	(0.19)	(0.20)	(0.20)	(0.21)	(0.21)	(0.21)
Number of charges	-0.94***	-1.19***	-1.08***	-1.00***	-1.02***	-1.06***	-1.07***
	(0.14)	(0.31)	(0.31)	(0.30)	(0.30)	(0.30)	(0.30)
Num distinct charge dates		0.55	0.56	0.70	0.67	0.71	0.69
		(0.51)	(0.50)	(0.52)	(0.52)	(0.52)	(0.52)
Has Pending Charge		-4.51*	-4.55*	-4.63*	-4.43*	-4.22*	-4.56*
		(1.98)	(1.97)	(1.99)	(1.99)	(1.99)	(2.01)
Has Felony			-4.49	-2.15	-2.61	-2.75	-2.97
			(4.09)	(4.25)	(4.27)	(4.32)	(4.31)
Share of charges that are felonies			2.46	2.00	1.85	2.20	2.32
			(6.96)	(7.12)	(7.17)	(7.24)	(7.21)
Has Violent				-0.84	-1.25	-0.95	-0.96
				(2.68)	(2.70)	(2.74)	(2.75)
Has Property				-5.90**	-5.70*	-5.44*	-5.60*
				(2.27)	(2.29)	(2.32)	(2.33)
Has Drug				-0.86	-1.17	-1.29	-1.30
				(1.87)	(1.87)	(1.89)	(1.91)
At least one traffic				0.15	-0.10	-0.26	-0.01
				(2.16)	(2.18)	(2.20)	(2.20)
(pred.) Black					-0.03	-0.01	0.17
					(2.37)	(2.39)	(2.40)
(pred.) Hispanic					-2.17	-2.10	-2.23
					(1.87)	(1.86)	(1.90)
(pred.) API/AIAN					2.99	3.05	4.80
					(6.56)	(6.50)	(6.51)
(pred.) Race unknown					4.94+	4.72	4.81
					(2.96)	(2.94)	(2.94)
(pred.) Female					-9.87***	-9.80***	-2.51
					(1.78)	(1.80)	(2.22)
(pred.) Gender unknown					2.22	2.17	4.06
					(4.12)	(4.11)	(4.17)
Candidate age (years)					0.36	0.37	0.41
					(0.47)	(0.47)	(0.47)
Candidate age squared					-0.00	-0.00	-0.00
					(0.01)	(0.01)	(0.01)
Adjudicator 1						1.71	1.71
						(2.61)	(2.64)
Adjudicator 2						3.44+	3.54+
						(1.95)	(1.97)
Other Adjudicator						0.45	0.65
						(3.91)	(3.84)
Num Applicants Same Day/Adj						0.00	0.00
						(0.02)	(0.02)
Constant	20.83***	21.09***	20.98***	21.02***	15.64+	14.98	12.21
	(1.31)	(1.46)	(1.45)	(2.02)	(8.96)	(9.75)	(9.76)
Occupation FE	No	No	No	No	No	No	Yes
Observations	8832	8832	8832	8832	8832	8831	8831
R-squared	0.003	0.003	0.003	0.003	0.006	0.007	0.013

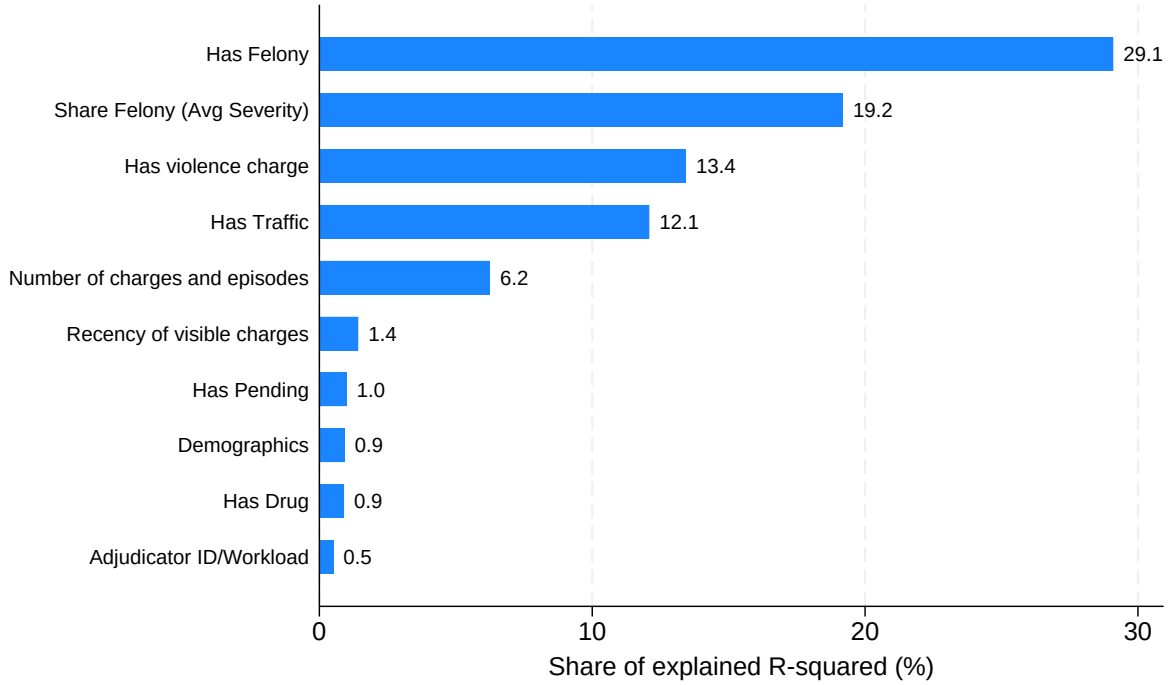
Note: Sample is control group with any (legally reportable) charge that is allowed on platform. Dependent variable is number of contacts (incl 0 if no contacts). Robust standard errors in parentheses. The Final column also includes FE for the year-month

Table B.9: Any Incident for those Allowed on Platform by Record Chars (Control Only)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Years since last charge	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Number of charges	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Num distinct charge dates		0.00^{+}	0.00^{+}	0.00^{+}	0.00^{+}	0.00^{+}	0.00^{+}
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Has Pending Charge		0.01	0.01	0.01	0.01	0.01	0.01
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Has Felony			-0.00	0.00	0.00	0.00	0.00
			(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Share of charges that are felonies			0.01	0.01	0.01	0.00	0.00
			(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Has Violent				-0.00	-0.00	-0.00	-0.00
				(0.01)	(0.01)	(0.01)	(0.01)
Has Property				0.00	0.00	0.00	0.00
				(0.01)	(0.01)	(0.01)	(0.01)
Has Drug				-0.01*	-0.01*	-0.01*	-0.01*
				(0.00)	(0.00)	(0.00)	(0.00)
At least one traffic				0.00	0.00	0.00	0.00
				(0.00)	(0.00)	(0.00)	(0.00)
(pred.) Black					-0.00	-0.00	-0.00
					(0.00)	(0.00)	(0.00)
(pred.) Hispanic					-0.00	-0.00	-0.01
					(0.00)	(0.00)	(0.00)
(pred.) API/AIAN					-0.00	-0.00	0.00
					(0.02)	(0.02)	(0.02)
(pred.) Race unknown					-0.00	-0.00	-0.00
					(0.01)	(0.01)	(0.01)
(pred.) Female					-0.01*	-0.01*	0.00
					(0.00)	(0.00)	(0.00)
(pred.) Gender unknown					0.01	0.01	0.02^{+}
					(0.01)	(0.01)	(0.01)
Candidate age (years)					-0.00	-0.00	-0.00
					(0.00)	(0.00)	(0.00)
Candidate age squared					0.00	0.00	0.00
					(0.00)	(0.00)	(0.00)
Adjudicator 1						0.00	0.00
						(0.01)	(0.01)
Adjudicator 2						-0.00	-0.00
						(0.00)	(0.00)
Other Adjudicator						0.00	0.00
						(0.01)	(0.01)
Num Applicants Same Day/Adj						-0.00	-0.00
						(0.00)	(0.00)
Constant	0.02***	0.02***	0.02***	0.02***	0.03^{+}	0.04^{+}	0.04^{+}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.02)	(0.02)
Occupation FE	No	No	No	No	No	No	Yes
Observations	8832	8832	8832	8832	8832	8831	8831
R-squared	0.000	0.001	0.000	0.001	0.001	0.000	0.007

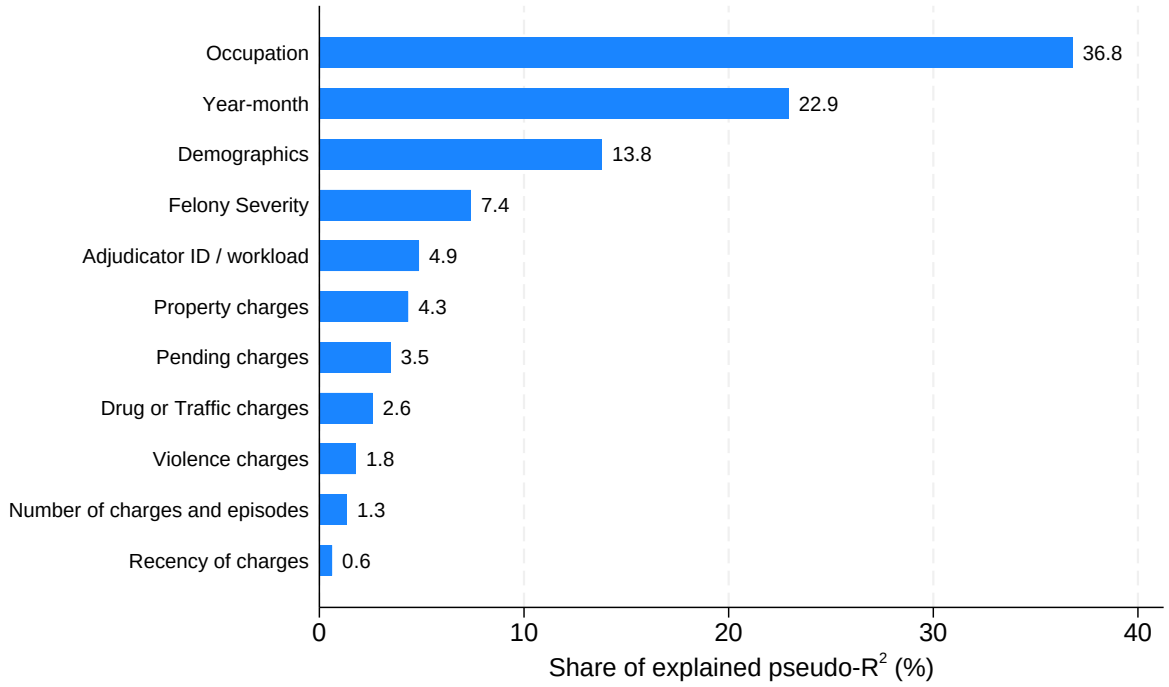
Note: Sample is control group with any (legally reportable) charge that is allowed on platform. Dependent variable is whether there was incident (binary). Robust standard errors in parentheses. The Final column also includes FE for the year-month in which the background check was adjudicated. Sample is control group, with a visible charge. Leftout race is predicted black;

Figure B.1: Decomposition of Explained Variation: Allowed on Platform



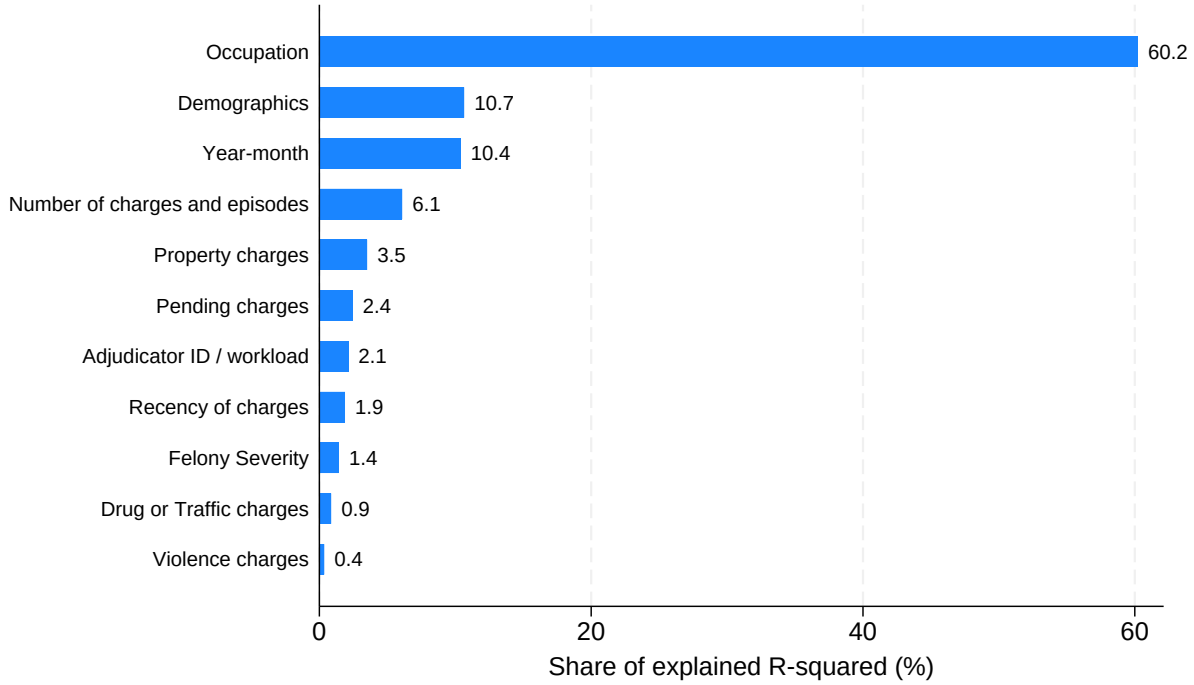
Note: Based on Column 6 from Table 5. Adds up to the *explained* variation in that regression, where Adjusted R^2=0.523. Demographics include predicted race and gender dummy variables as well as age. Adjudicator ID/workload includes individual adjudicator id dummies, number of applications on same day/adjudicator, and year-month FE.

Figure B.2: Decomposition of Explained Variation: Any Contact



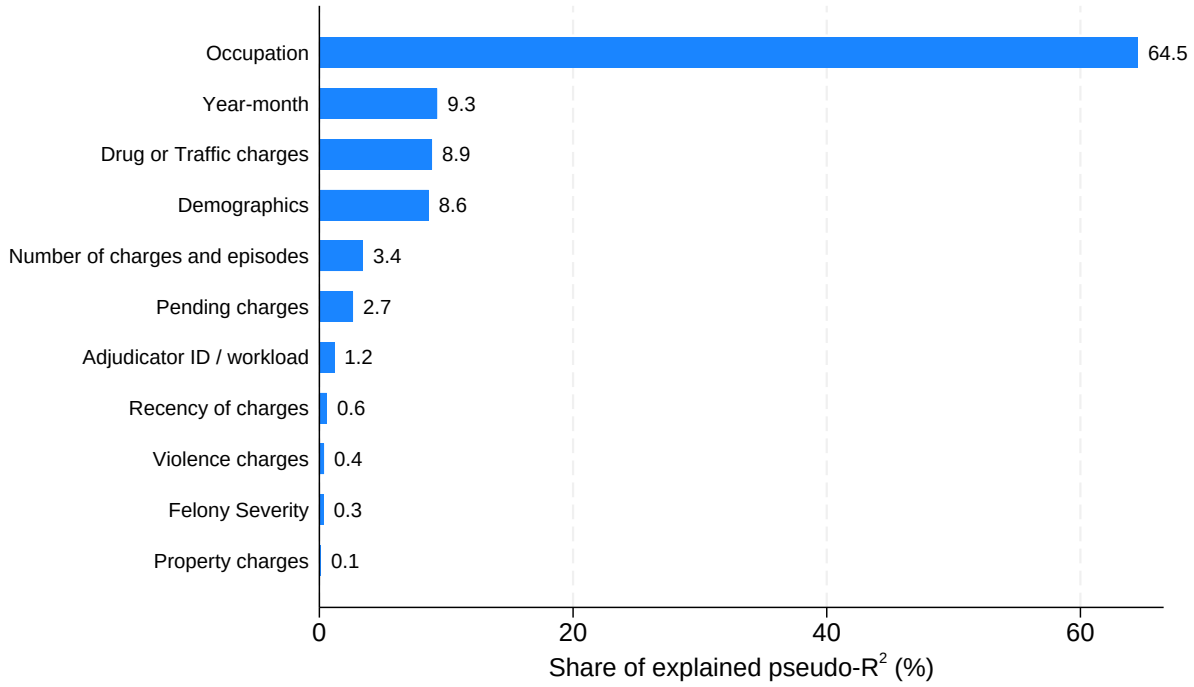
Note: Dominance decomposition of the McFadden pseudo-R^2 from a logistic regression of any contact on the listed blocks of characteristics, mimicking Column (2) of Table 7 among control applicants with a record who were allowed on the platform. General dominance weights are Shapley values (Shapley, 1953; Azen and Traxel, 2009). Bars sum to 100% of the explained pseudo-R^2; the total explained variation is small at 0.055 (Table 7).

Figure B.3: Decomposition of Explained Variation: Number of Contacts



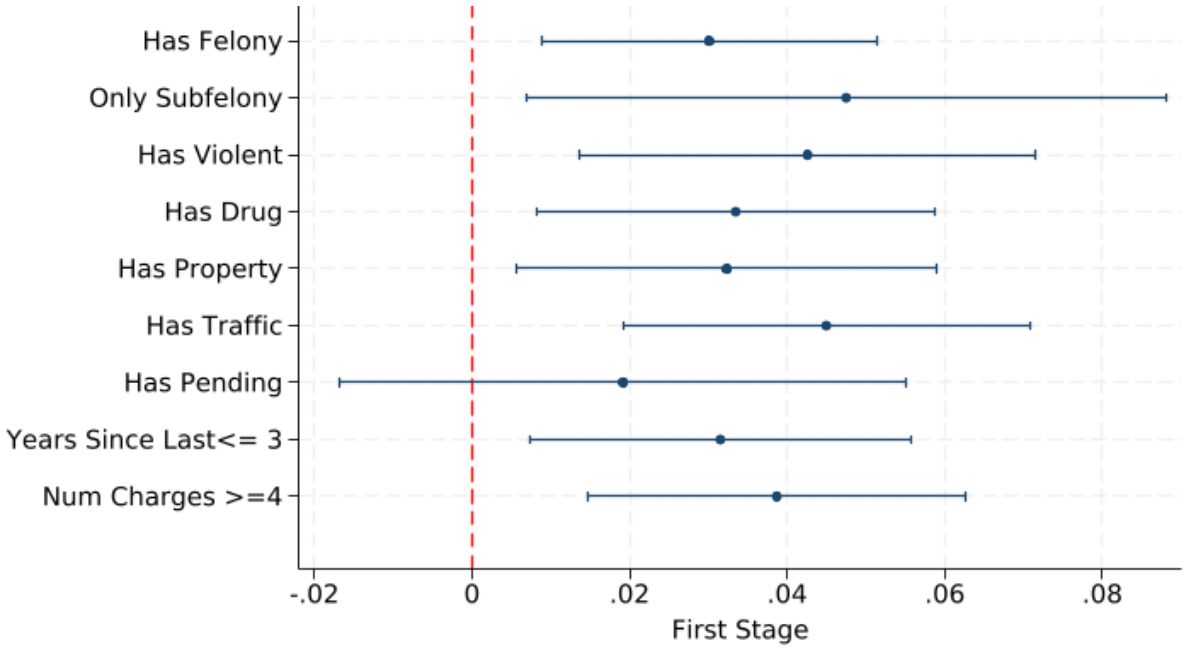
Note: Shapley decomposition of the R^2 from a linear regression of the number of contacts on the listed blocks of characteristics, same as Column (1) of Table 7 among control applicants with a record who were allowed on the platform (Shapley, 1953; Shorrocks et al., 2013). Bars sum to 100% of the explained R^2; the total explained variation is small at 0.013 (Table 7).

Figure B.4: Decomposition of Explained Variation: Any Incident



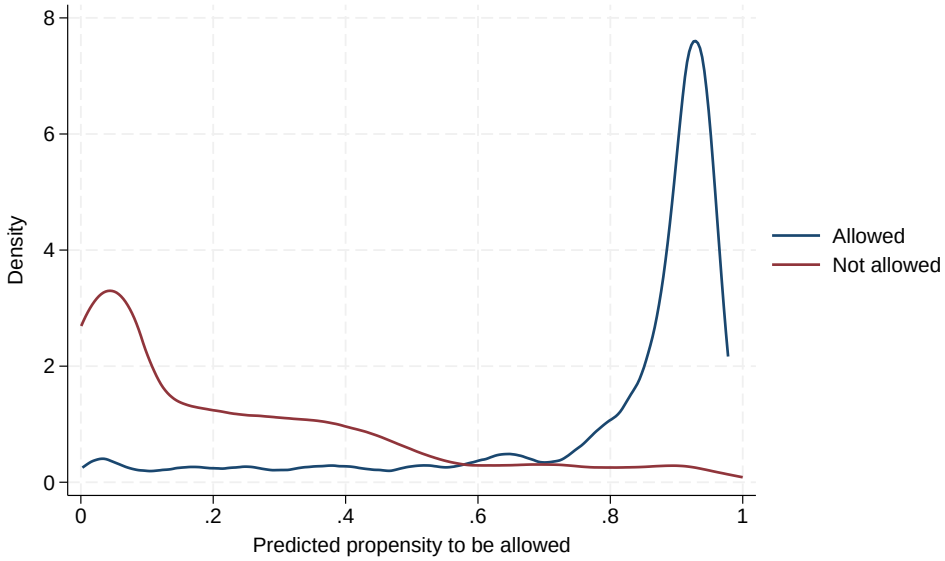
Note: Dominance decomposition of the McFadden pseudo-R^2 from a logistic regression of any contact on the listed blocks of characteristics, mimicking Column (2) of Table 7 among control applicants with a record who were allowed on the platform. General dominance weights are Shapley values (Shapley, 1953; Azen and Traxel, 2009). Bars sum to 100% of the explained pseudo-R^2; the total explained variation is small at 0.007 (Table 7).

Figure B.5: First Stages for Reduced Visibility Group in Subsamples



Notes: Each dot is a coefficient from a first stage regression of allowed on treatment in the reduced visibility group for the subgroups listed. Robust 95% CIs shown.

Figure B.6: Predicted Probability of Allowance by Adjudication Decision



Note: The figure shows kernel density estimates of the cross-fitted predicted probability of allowance, separately for applicants who were allowed and not allowed. Each density is normalized within the corresponding realized-decision group. The estimation sample and predictors are described in Figure 4.

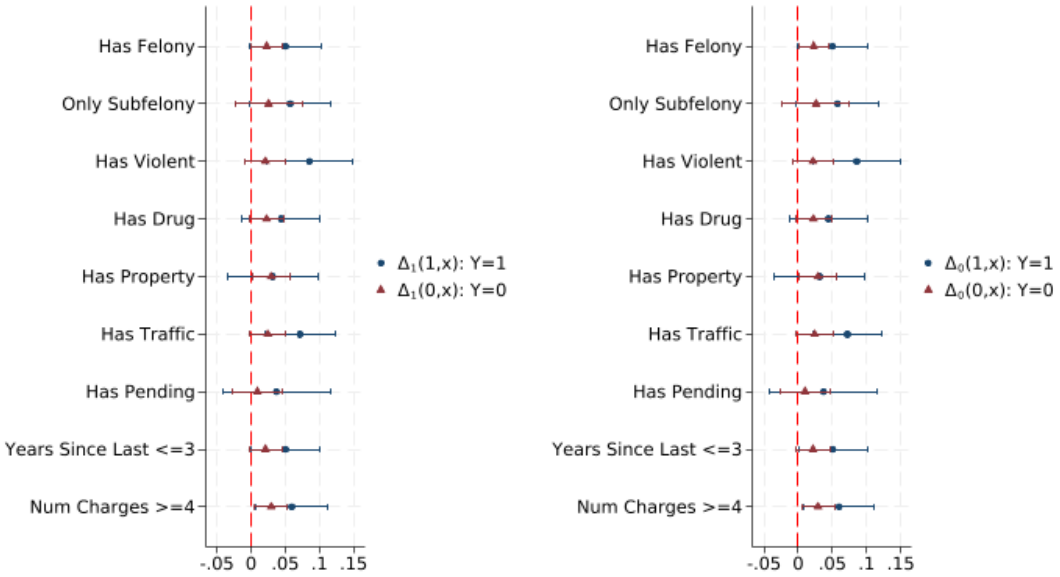
Table B.10: Reduced Form Effects: Including Conditional

	(1) All Movers	(2) Placebo	(3) Newly Clear	(4) Reduced Vis
Any Contact	0.019**	0.001	0.027**	0.005
	(0.008)	(0.007)	(0.011)	(0.009)
[RW p]	[0.064]	[0.930]	[0.048]	[0.760]
Control mean	0.381	0.336	0.538	0.110
Any Review	0.010*	-0.005	0.010	0.011*
	(0.006)	(0.005)	(0.008)	(0.006)
[RW p]	[0.224]	[0.828]	[0.507]	[0.208]
Control mean	0.126	0.109	0.179	0.035
Any Incident	-0.002	-0.002	-0.001	-0.003
	(0.002)	(0.002)	(0.003)	(0.002)
[RW p]	[0.629]	[0.828]	[0.912]	[0.557]
Control mean	0.017	0.013	0.023	0.008
Number of Contacts	0.618	-1.016	0.195	1.373*
	(1.072)	(1.020)	(1.625)	(0.706)
[RW p]	[0.629]	[0.828]	[0.912]	[0.208]
Control mean	13.612	12.154	20.113	2.416
Days BGC to Last Contact	3.578*	-0.553	5.770**	-0.343
	(1.955)	(1.787)	(2.844)	(1.961)
[RW p]	[0.224]	[0.930]	[0.134]	[0.862]
Control mean	58.825	50.332	82.932	17.303
Log(Contacts)	0.030	-0.067	0.012	0.185
	(0.046)	(0.049)	(0.049)	(0.139)
[RW p]	[0.832]	[0.457]	[0.972]	[0.487]
Control mean	2.011	2.013	2.041	1.757
Reviews per Contact	0.003	-0.001	0.001	0.021**
	(0.003)	(0.003)	(0.003)	(0.010)
[RW p]	[0.808]	[0.828]	[0.972]	[0.273]
Control mean	0.043	0.043	0.044	0.042
Incidents per Contact	-0.001	-0.001	-0.002*	0.007
	(0.001)	(0.001)	(0.001)	(0.005)
[RW p]	[0.832]	[0.766]	[0.251]	[0.487]
Control mean	0.003	0.002	0.004	0.002
Average Rating	0.025	-0.099*	0.007	0.176
	(0.051)	(0.055)	(0.053)	(0.181)
[RW p]	[0.832]	[0.261]	[0.972]	[0.487]
Control mean	4.451	4.547	4.478	4.216
N	13,364	13,675	8,572	4,792

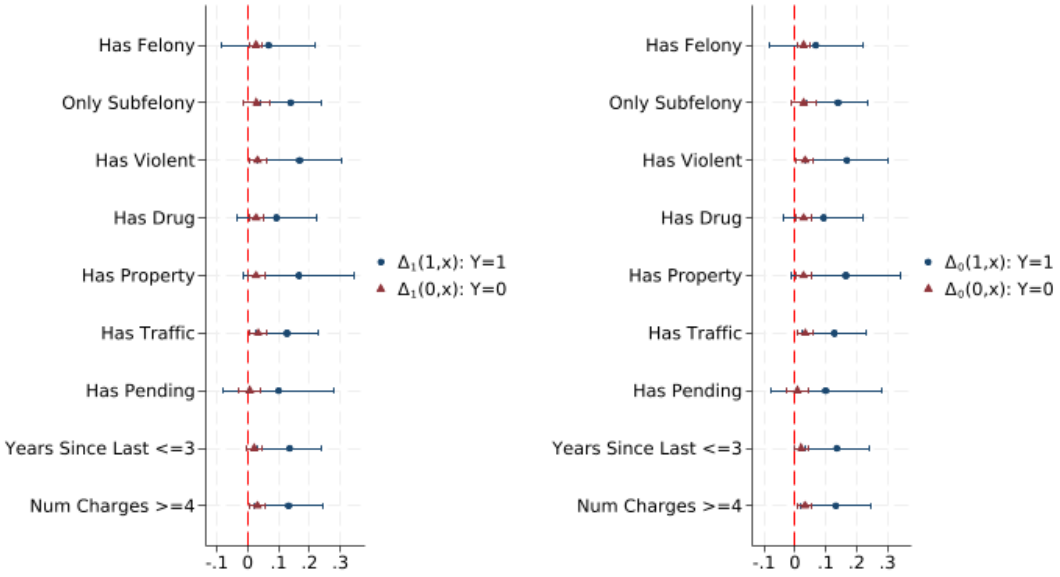
Note: Each row is a different regression; each column is a different sample: (1) all movers, (2) the placebo group whose visible record is unchanged by treatment, (3) newly clear, and (4) reduced visibility. The values represent coefficients from a regression of the outcome listed in the row on treatment; The control mean is the constant from that regression. All outcomes are measured for each individual for 12 months after the background check is completed. The first five outcomes are not conditioned on being allowed: they are set to 0 for anyone not allowed onto the platform. Log(Contacts), reviews per contact, incidents per contact, and average rating are conditional on having at least one contact (or, for ratings, at least one rating). Romano–Wolf p-values, adjusted for multiple hypothesis testing, are in brackets.
* p<0.10, ** p<0.05, *** p < 0.01

Figure B.7: Instrument Validity Test for Reduced Visibility Group

(a) Y =Any Contact



(b) Y =Any Review



Notes: All results are shown within reduced visibility treatment response group. $\Delta_z(d,x) = Pr(Y=y, D=d \mid Z=z, X=x) - Pr(Y=y, D=d \mid Z=1-z, X=x)$. If there is a first stage and the instrument is valid (exclusion, independence, and monotonicity hold), $\Delta_z(d,x) \geq 0$ (Kitagawa, 2015; Heckman and Vytlacil, 2005). Blue dots represent $Y=1$, maroon represent $Y=0$. Each X is listed in the rows. Panel (a) uses Y=Any Contact; Panel (b) uses Y =Any Review.

Table B.11: Average Outcomes by Treatment Response Type: Newly Clear and Reduced Visibility

	Newly Clear				Reduced Visibility			
	(1) Control (AT)	(2) Treatment (AT+C)	(3) Implied (C Mean)	(4) p (AT=C)	(5) AT	(6) Treatment (AT+C)	(7) C	(8) p (AT=C)
Any Contact	0.569	0.566	0.504	0.743	0.465	0.424	0.150	0.171
	(0.008)	(0.008)	(0.200)		(0.021)	(0.020)	(0.230)	
More than 10 Contacts	0.218	0.213	0.125	0.578	0.140	0.156	0.262	0.424
	(0.007)	(0.006)	(0.166)		(0.014)	(0.014)	(0.152)	
Any Incident	0.024	0.022	-0.025	0.411	0.035	0.020	-0.075	0.148
	(0.002)	(0.002)	(0.060)		(0.008)	(0.006)	(0.076)	
N	3,974	4,366			578	641		

Note: Columns (1)–(4) are restricted to the newly clear treatment response group; columns (5)–(8) to the reduced visibility group. Columns (1) and (5), AT, report the always-taker mean (control, admitted; $Z = 0$, $D = 1$). Columns (2) and (6) report the mean among all treatment-admitted applicants ($Z = 1$, $D = 1$) — always takers plus compliers. Columns (3) and (7), C, report the complier mean, estimated via 2SLS as described in Section 6.3. Columns (4) and (8) report the p-value for the test that the always-taker and complier means are equal. Outcomes are measured over the 12 months after the background check and are set to 0 for applicants not allowed on the platform. Complier means are imprecise where the first stage is modest in levels and can fall outside a variable's natural range. N is the number of admitted applicants in each arm. Robust standard errors in parentheses.

+ p<0.10, * p<0.05, ** p<0.01, *** p < 0.001

Table B.12: Average Outcomes by Treatment Response Type: Conditional Outcomes

	Movers				Newly Clear				Reduced Visibility			
	(1) Control (AT)	(2) Treatment (AT+C)	(3) Implied (C Mean)	(4) p (AT=C)	(5) AT	(6) Treatment (AT+C)	(7) C	(8) p (AT=C)	(9) AT	(10) Treatment (AT+C)	(11) C	(12) p (AT=C)
Log(Contacts)	2.011	2.042	2.348	0.501	2.041	2.053	2.172	0.802	1.757	1.942	4.108	0.398
	(0.033)	(0.032)	(0.500)		(0.035)	(0.034)	(0.520)		(0.093)	(0.103)	(2.781)	
Reviews per Contact	0.043	0.047	0.079	0.281	0.044	0.045	0.056	0.714	0.042	0.063	0.305	0.343
	(0.002)	(0.002)	(0.033)		(0.002)	(0.002)	(0.033)		(0.005)	(0.009)	(0.277)	
Incidents per Contact	0.003	0.003	-0.006	0.403	0.004	0.002	-0.016	0.065	0.002	0.009	0.094	0.402
	(0.001)	(0.001)	(0.012)		(0.001)	(0.000)	(0.011)		(0.001)	(0.005)	(0.109)	
Average Rating	4.451	4.474	4.557	0.598	4.478	4.485	4.509	0.882	4.216	4.392	4.898	0.304
	(0.037)	(0.035)	(0.197)		(0.038)	(0.037)	(0.208)		(0.147)	(0.105)	(0.647)	
N	839	935			754	828			85	107		

Note: Columns (1)-(4) pool the movers (newly clear and reduced visibility); columns (5)-(8) are restricted to the newly clear treatment response group, and (9)-(12) to the reduced visibility group. Always taker outcomes, in columns (1), (5), and (9), are estimated among applicants in the control group ($Z = 0$) who are allowed on the platform ($D = 1$). Columns (2), (6), and (10) report the average among treatment-group applicants who are allowed ($Z = 1$, $D = 1$). Columns (3), (7), and (11) estimate average outcomes for compliers via methodology in Section 6.3. Columns (4), (8), and (12) are p-values testing the null that the mean for always takers and compliers is the same. Robust standard errors in parentheses. Logs are reported conditional on $Y > 0$. Each regression is conditional on having at least one contact or review for the "per" values; for the star rating values it is conditional on at least one star rating.
+ p<0.10, * p<0.05, ** p<0.01, *** p < 0.001

Table B.13: Outcomes of Admitted Applicants With and Without a Criminal Record

		Record, admitted	
	(1) No Record	(2) Control (AT)	(3) Treatment (AT+C)
Any Contact	0.721	0.556	0.548
	(0.001)	(0.007)	(0.007)
More than 10 Contacts	0.273	0.208	0.206
	(0.001)	(0.006)	(0.006)
Any Incident	0.018	0.025	0.021
	(0.000)	(0.002)	(0.002)
N	162,303	4,552	5,007

Note: Column (1) reports mean outcomes for applicants with no legally reportable charge who are allowed on the platform. Columns (2) and (3) repeat columns (1) and (2) of Table 15: always takers among movers (control, admitted), and all treatment-admitted movers (always takers plus compliers). Outcomes are measured over the 12 months after the background check. Standard errors in parentheses.

C LATE Assumptions

Interpreting the estimates above requires the standard LATE assumptions (Imbens and Angrist, 1994):

1. **Independence** ($Z \perp (Y_i(0), Y_i(1), D_i(0), D_i(1))$): This follows from random assignment. Balance on observable criminal record characteristics is shown in Tables 2 and B.2-B.5. Appendix Tables B.2–B.5 report balance separately by treatment response type: always clear, newly clear, reduced visibility, and same visibility. We similarly find no systematic differences in pre-treatment characteristics between treatment and control within these groups.

2. **Exclusion** ($Y_i(D_i(Z_i), Z_i) = Y_i(D_i(Z_i))$): Treatment affects outcomes only through platform access. This is plausible in our setting because neither adjudicators nor applicants are aware of treatment assignment, and the intervention operates solely by altering the visibility of criminal records used in the admission decision.

3. **Relevance** ($Cov(Z_i, D_i) \neq 0$): The first stage is reported in Table 10. Treatment significantly increases the probability of platform access, particularly in the newly clear group, where the first stage is large by construction.

4. **Monotonicity** ($\forall i : D_i(1) \geq D_i(0)$): Treatment weakly reduces the set of visible charges and therefore (weakly) increases the probability of admission. In the newly clear group, monotonicity holds mechanically, as treatment clears all visible charges and bypasses adjudication. In the always clear and same visibility groups, treatment does not change the visible record, implying no change in admission decisions. In the reduced visibility group, treatment removes a subset of charges, weakening the signal observed by adjudicators and thereby (weakly) increasing the probability of admission.[19]

[19]While monotonicity is automatic for the newly clear group, one could imagine behavioral responses in the reduced visibility group that imply the existence of defiers. For example, if adjudicators *only* attend to the average severity of the visible record (e.g., felony versus misdemeanor), then if the treatment removes a misdemeanor but leaves a felony, an applicant could be rejected under treatment but admitted under control.While such behavior may reflect heuristic or salience-based processing rather than a fully rational evaluation of the record, Figures B.5 and B.7 explore this possibility for the reduced visibility group. Figure B.5 shows that the first stage is positive across several subsamples defined by criminal record characteristics within the reduced visibility group, consistent with "average monotonicity" (Frandsen et al., 2023). In addition, if the instrument is valid (exclusion, independence, and monotonicity), then $\Delta_z(d, x)$ should be ≥ 0, where $\Delta_z(d, x) = \Pr(Y = y, D = d \mid Z = z, X = x) - \Pr(Y = y, D = d \mid Z = 1 - z, X = x)$ (Heckman and Vytlacil, 2005; Kitagawa, 2015). Figure B.7 evaluates these restrictions for Y equal to any contact and any review, using the same set of X as in the average monotonicity tests. Across a wide range of pre-treatment record characteristics, we do not observe systematic violations of the IV model's testable implications. Estimates of $\Delta_z(d, x)$ are generally positive, although some confidence intervals include zero.

D Description of December 2023 Anomaly

D.1 Description of the Issue

In the last two weeks of December 2023, there are data oddities affecting applicants with records theoretically unaffected by treatment. Specifically, the share of individuals with no charges hidden under the control matrix is higher in the control group (49% of candidates with charges) than in the treatment group (42% of candidates with charges) in this two-week period.

Additionally, those with no difference in the visibility of their record between the control and treatment matrices are more likely to be issued an adverse action notice in the control group during this same period. This group of 54 applicants in the control group was issued adverse action notices at a rate of 94% versus 63% among the 49 applicants in the treatment group.

A similar phenomenon is visible, but with smaller magnitudes and only borderline significance, for individuals with some charges hidden in the control matrix and the same visible charges under the treatment matrix.

D.2 Ruling Out Potential Explanations

We confirm that there is general treatment and control group balance across covariates during this time. Visible records look qualitatively similar, showing no significant differences in number of charges or severity of charges across several metrics in Table D.1.

We check these two weeks against other placebo two-week periods in the months before and after this window, and find no other significant differences, in Figure D.1 and in Table D.2. The two-week anomaly period in late December 2023 is a rare event relative to nearby placebo windows. Again, this suggests a temporary shock confined to this period of time rather than a stable structural change differentially affecting treatment and control.

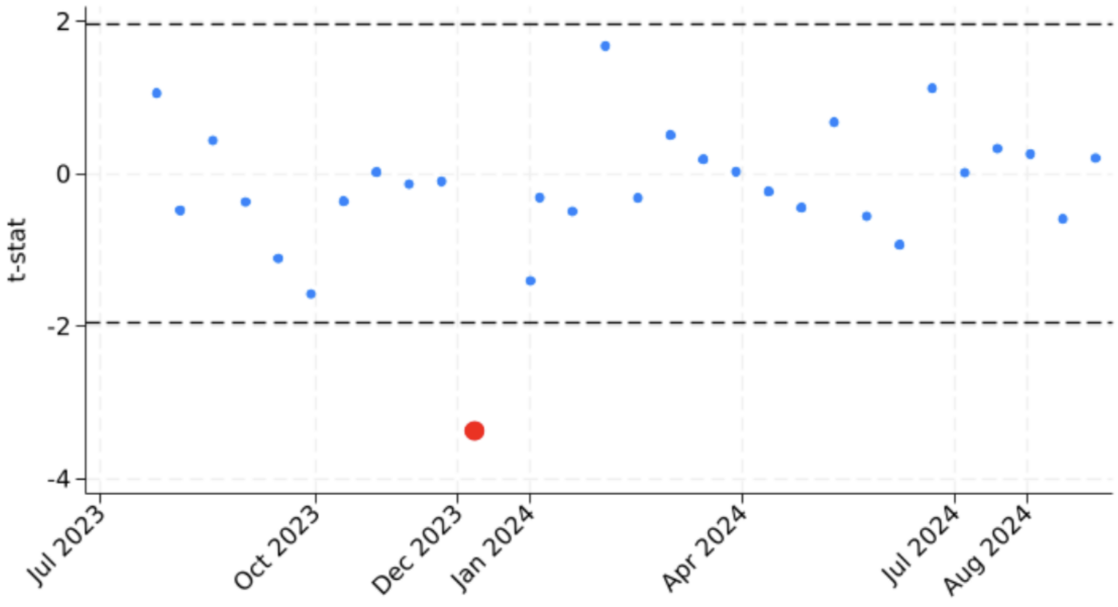
We also investigate whether this can be explained by either assignment to particular adjudicator(s) or odd behavior limited to a specific adjudicator, conditional on assignment, which also does not seem to be the case. Adjudicator assignment probabilities and adverse action probabilities conditional on adjudicator are in Table D.3. Within the adjudicators working on the holiday period reports, adverse action rates are higher for both adjudicators in the control group relative to the treatment group.

One potential explanation we cannot currently directly test for is some kind of adjudication

backlog after the holidays affecting the groups differentially (per the company, this kind of backlog is a common occurrence around winter holidays). However, in order for this to result in adjudication differences across treatment and control groups, the backlog list would have to be sorted somehow on treatment/control matrix assignment, leading to sequencing effects. This may be theoretically possible, though unlikely, due to the user interface looking exactly the same for treatment and control assigned applicant; however, on the back end, there are tags that differentiate between treatment and control assignment for the researchers' use. Given that this does not seem to have happened in any other time period, it would have to interact with the backlog to make sense as an explanation.

Since we cannot account for a singular explanation for the anomaly with the information we currently have access to, and since we have no reason to believe there were similar issues in any other period of the data, we have decided to drop this two-week period from the main analysis for all groups.

Figure D.1: Placebo 2-Week Window T-Statistics of the Treatment Effect on Platform Allowance



Notes: Each dot represents a t-statistic from a 2-week window, testing the null that P(Allowed | Treatment)=P(Allowed | Control) for the Same Visibility group.

Table D.1: Covariate Balance in Same Visibility Groups by Time Period

Group	Period	Metric	Mean C	Mean T	Diff (T-C)	P-value	N C	N T
NCH	Non-Anomaly	N reports	2105	2119	14	0.829	2105	2119
NCH	Non-Anomaly	N charges	1.955	1.954	-0.001	0.986	2105	2119
NCH	Non-Anomaly	Any felony	0.572	0.578	0.006	0.710	2105	2119
NCH	Non-Anomaly	Any conviction	0.671	0.652	-0.019	0.202	2105	2119
NCH	Non-Anomaly	Any pending	0.386	0.432	0.046	0.291	2105	2119
NCH	Non-Anomaly	Time since most recent charge	1428.382	1373.421	-54.961	0.337	2104	2118
NCH	Anomaly	N reports	54	49	-5	0.622	54	49
NCH	Anomaly	N charges	2.352	2.245	-0.107	0.810	54	49
NCH	Anomaly	Any felony	0.611	0.469	-0.142	0.152	54	49
NCH	Anomaly	Any conviction	0.685	0.673	-0.012	0.900	54	49
NCH	Anomaly	Any pending	0.870	0.347	-0.523	0.073	54	49
NCH	Anomaly	Time since most recent charge	1333.444	1189.204	-144.240	0.642	54	49
SCH	Non-Anomaly	N reports	1932	1890	-42	0.497	1932	1890
SCH	Non-Anomaly	N charges	6.157	6.101	-0.056	0.743	1932	1890
SCH	Non-Anomaly	Any felony	0.715	0.685	-0.030	0.042	1932	1890
SCH	Non-Anomaly	Any conviction	0.938	0.932	-0.007	0.403	1932	1890
SCH	Non-Anomaly	Any pending	0.497	0.590	0.094	0.154	1932	1890
SCH	Non-Anomaly	Time since most recent charge	1786.390	1815.601	29.211	0.664	1932	1890
SCH	Anomaly	N reports	52	46	-6	0.544	52	46
SCH	Anomaly	N charges	5.308	5.478	0.171	0.867	52	46
SCH	Anomaly	Any felony	0.558	0.674	0.116	0.243	52	46
SCH	Anomaly	Any conviction	0.885	0.913	0.028	0.647	52	46
SCH	Anomaly	Any pending	0.462	0.196	-0.266	0.183	52	46
SCH	Anomaly	Time since most recent charge	1582.327	1530.522	-51.805	0.887	52	46

Note: Group NCH refers to the same visibility group with No Charges Hidden in both treatment and control. Group SCH refers to the same visibility group with Some Charges Hidden in both treatment and control. Period Anomaly refers to December 15-31, 2023, and Period Non-Anomaly refers to the remaining sample.

Table D.2: Two-week placebo-window comparison around Dec 15–31 anomaly window.

Window	Group	N windows	Target N	Target diff	Target t	Tail share	Tail %	[Min t, Max t]
14-day	Same visibility (all)	392	180	0.262	4.239	0.003	0.26	[-2.597, 4.239]
14-day	Same visibility NCH	392	98	0.308	3.860	0.020	2.04	[-2.164, 4.377]
14-day	Same visibility SCH	392	82	0.209	2.148	0.064	6.38	[-2.345, 2.937]
17-day	Same visibility (all)	389	201	0.244	4.242	0.013	1.29	[-2.549, 4.562]
17-day	Same visibility NCH	389	103	0.312	4.083	0.021	2.06	[-2.317, 4.271]
17-day	Same visibility SCH	389	98	0.172	1.997	0.103	10.28	[-2.757, 3.262]

Note: Each row represents a separate rolling-placebo exercise (window length × subgroup). *N windows* is the number of placebo windows/regressions estimated when sliding that window around nearby dates. *Target diff* is the treatment–control adverse-action (AA) difference in the focal holiday window (Dec 15–31 for the 14-day spec; analogous centered 17-day target for the 17-day spec). *Target t* is the t-statistic on the treatment effect from that focal-window regression. *Tail share* is the fraction of placebo windows with $|t_{\text{placebo}}| \geq |t_{\text{target}}|$; smaller values mean the focal-window result is more unusual relative to placebo windows.

Table D.3: Adjudicator assignment balance and conditional adverse-action balance (AA | adjudicator) by Same-Visibility Group

Group	Period	Metric	Mean C	Mean T	Diff (T-C)	P-value	N C	N T
NCH	Non-Anomaly	Assignment: adj389	0.343	0.328	-0.015	0.302	2105	2119
NCH	Non-Anomaly	AA | adj389	0.843	0.827	-0.016	0.412	721	694
NCH	Non-Anomaly	Assignment: adj829	0.239	0.234	-0.005	0.682	2105	2119
NCH	Non-Anomaly	AA | adj829	0.881	0.869	-0.012	0.566	503	495
NCH	Non-Anomaly	Assignment: adj4df	0.373	0.394	0.021	0.167	2105	2119
NCH	Non-Anomaly	AA | adj4df	0.794	0.806	0.012	0.543	785	834
NCH	Non-Anomaly	Assignment: adjother	0.046	0.045	-0.000	0.963	2105	2119
NCH	Non-Anomaly	AA | adjother	0.875	0.896	0.021	0.652	96	96
NCH	Anomaly	Assignment: adj389	0.574	0.571	-0.003	0.979	54	49
NCH	Anomaly	AA | adj389	1.000	0.643	-0.357	0.000	31	28
NCH	Anomaly	Assignment: adj829	0.000	0.000	0.000		54	49
NCH	Anomaly	AA | adj829					0	0
NCH	Anomaly	Assignment: adj4df	0.426	0.429	0.003	0.979	54	49
NCH	Anomaly	AA | adj4df	0.870	0.619	-0.251	0.057	23	21
NCH	Anomaly	Assignment: adjother	0.000	0.000	0.000		54	49
NCH	Anomaly	AA | adjother					0	0
SCH	Non-Anomaly	Assignment: adj389	0.314	0.304	-0.010	0.498	1931	1890
SCH	Non-Anomaly	AA | adj389	0.807	0.794	-0.013	0.591	606	574
SCH	Non-Anomaly	Assignment: adj829	0.279	0.261	-0.018	0.217	1931	1890
SCH	Non-Anomaly	AA | adj829	0.798	0.777	-0.020	0.422	539	494
SCH	Non-Anomaly	Assignment: adj4df	0.367	0.398	0.032	0.043	1931	1890
SCH	Non-Anomaly	AA | adj4df	0.763	0.725	-0.038	0.100	708	753
SCH	Non-Anomaly	Assignment: adjother	0.040	0.037	-0.004	0.533	1931	1890
SCH	Non-Anomaly	AA | adjother	0.821	0.783	-0.038	0.567	78	69
SCH	Anomaly	Assignment: adj389	0.635	0.565	-0.069	0.489	52	46
SCH	Anomaly	AA | adj389	0.848	0.692	-0.156	0.156	33	26
SCH	Anomaly	Assignment: adj829	0.000	0.000	0.000		52	46
SCH	Anomaly	AA | adj829					0	0
SCH	Anomaly	Assignment: adj4df	0.365	0.435	0.069	0.489	52	46
SCH	Anomaly	AA | adj4df	0.842	0.650	-0.192	0.179	19	20
SCH	Anomaly	Assignment: adjother	0.000	0.000	0.000		52	46
SCH	Anomaly	AA | adjother					0	0

Note: Group NCH refers to the same visibility group with No Charges Hidden in both treatment and control. Group SCH refers to the same visibility group with Some Charges Hidden in both the treatment and control matrices. Period Anomaly refers to December 15-31, 2023, and Period Non-Anomaly refers to the remaining analysis sample.

Table D.4: Report Adjudication Delay by Treatment Arm by Same-Visibility Group.

Group	Period	Metric	Mean C	Mean T	Diff (T-C)	P-value	N C	N T
NCH	Non-Anomaly	Delay (hours)	142.2	143.8	1.6	0.694	2097	2119
NCH	Anomaly	Delay (hours)	186.1	204.5	18.3	0.461	54	49
SCH	Non-Anomaly	Delay (hours)	138.7	140.3	1.6	0.702	1928	1886
SCH	Anomaly	Delay (hours)	182.9	190.8	7.9	0.747	52	46

Notes: The outcome is the number of hours between `report_completed_at_raw` and `adjudicated_at_raw`. The sample is restricted to same-visibility NCH/SCH records with charges and human review (`nohumanreview==0`). Period N excludes Dec. 15–31, 2023; period B is Dec. 15–31, 2023. Rows with missing timestamps, negative delay, or delay above 60 days are excluded from the timing outcome; this removes one same-visibility human-reviewed row. P-values are from two-sample t-tests comparing treatment and control means within each group-period cell.

E Distribution and Outliers in Number of Contacts

Table 11 and B.10 displayed some apparent differences between those assigned treatment and those assigned control within the always clear group, in particular in ln(contacts) conditional on contacting anyone which flowed downstream to apparently fewer reviews and incidents in the treatment group. The always clear group is a placebo group in that treatment does not directly impact this group.

Figure E.2 illustrates the issue. It displays the distribution of number of contacts in the always clear group, separately for contacts < 100 (about the 95th percentile) and contacts > 100. The distribution has a lot of 0s and is also fat tailed. With contacts < 100 the treatment and control distributions lie on top of each other and it's hard to see much difference. However, with contacts > 100 there does appear to be some right skew to the contact distribution in favor of the control group. We seem to have randomly gotten some very high contact individuals in the control group. These high contacts do not seem to represent data errors as far as we can tell, the pattern tends to be high contact over time, not just a single-day spike, the most extreme cases have about 4 contacts/day on active days, with a few very high-frequency accounts reaching 12 contacts/day, consistent with active contact behavior and not data issues.

To assess whether this pattern reflects a coding error or is instead explicable given the information we observe, we conduct four additional checks: we characterize the extreme accounts directly, we formally test whether the imbalance is distinguishable from randomization noise, we test whether it is concentrated in a particular month, and we test whether it reflects differences in the mix of services offered.

E.1 Characterizing the Extreme Accounts

We first ask what distinguishes accounts with more than 100 contacts in the twelve months following their background check from the rest of the always-clear sample. Table E.2 compares the two groups directly.

Extreme accounts have substantially more contact activity in the twelve months *before* their background check (9.9 contacts on average, versus 0.8 for non-extreme accounts), and are more than twice as likely to have any pre-background-check contact activity at all (28% versus 13%). They also begin contacting clients sooner after their background check completes (5.8 days versus 24.2 days) and sustain activity over a much longer window (a 469-day average contact span versus 116 days). These patterns are consistent with extreme accounts being

Table E.1: Outlier Robustness by Visibility Group

	Always Clear	Newly Clear	Reduced Visibility	Same Visibility
Panel A: OLS				
Number of Contacts	-3.565**	1.068	2.083***	0.749
	(1.570)	(1.435)	(0.766)	(0.809)
Number of Reviews	-0.386**	0.089	0.114	-0.000
	(0.151)	(0.131)	(0.073)	(0.079)
Panel B: Rank Test				
Number of Contacts	-106.344	396.802***	145.692***	-12.454
	(68.162)	(74.833)	(49.267)	(55.122)
Number of Reviews	-106.476	225.379***	84.516***	-27.795
	(68.798)	(75.473)	(30.892)	(31.837)
Panel C: Contacts < p95				
Number of Contacts	-0.411	0.755**	0.596**	-0.290
	(0.406)	(0.376)	(0.234)	(0.183)
Number of Reviews	-0.045	0.081**	0.037	-0.037*
	(0.044)	(0.041)	(0.027)	(0.019)
Panel D: Median Regression				
Number of Contacts	0.000	1.000***	0.000	0.000
	(0.112)	(0.171)	(.)	(.)
Number of Reviews	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)
Panel E: p75 Regression				
Number of Contacts	-1.000	1.000	0.000	0.000
	(0.700)	(0.751)	(.)	(.)
Number of Reviews	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)
Rob. t-test (*Contacts*)	0.100	1.000	0.100	0.400
Rob. t-test (*Reviews*)	0.100	0.500	0.160	1.000
N	9,447	11,134	5,927	8,045

Note: The rank test is the rank test suggested by Athey and Imbens (2017). Panel C restricts observations ot those with values below the 95percentile within the column. Panels D/E are quantile regressions at the median and 75th percentile, respectively. Robust t-test rows are the p-values for rows (1) and (2) from Müller (2025) which is robust to the presence of (moderately) heavy tails in the distribution; the value represents largest value on a specified grid {....,0.09, 0.10, 0.12, 0.14,....,0.4,0.5,1.0} such that all tests of that level and smaller reject, and thus can be interpreted as an upper-bound.

Table E.2: Characteristics of Extreme vs. Non-Extreme Contact Accounts (Always Clear Group)

	Non-Extreme	Extreme	Total
Contacts before BGC	0.843	9.873	1.552
	(5.018)	(31.42)	(10.32)
Any contact before BGC	0.129	0.284	0.141
	(0.336)	(0.451)	(0.349)
Days to first contact after BGC	24.23	5.793	22.79
	(65.83)	(19.01)	(63.61)
Contact span (days)	116.3	468.5	143.9
	(177.6)	(163.1)	(200.3)
Days on platform before BGC	344.4	450.1	352.7
	(825.1)	(920.5)	(833.4)
Number of categories	6.596	9.205	6.800
	(9.569)	(9.988)	(9.628)
Number of active services	0.709	0.922	0.725
	(0.560)	(0.479)	(0.557)
Number of reviews (12mo)	0.756	19.75	2.247
	(2.160)	(24.74)	(8.849)

Note: Extreme accounts are those with more than 100 contacts in the twelve months following background check completion; non-extreme accounts have 100 or fewer. Sample restricted to the always-clear group (charges present but fully hidden under the control matrix), excluding the December 2023 problem period described in Appendix D. Mean coefficients reported, with standard deviations in parentheses. N = 9,108 non-extreme, 497 extreme. This is a descriptive comparison, not a randomization-based test, since extreme-account status is itself defined from the outcome.

established providers who were already active on the platform, rather than accounts whose activity is an artifact of how the data were constructed.[20]

We also check three sources of the imbalance that would point to a genuine data problem rather than a real behavioral difference. First, none of the extreme accounts originate from unmatched or zero-imputed merge records between the background check and marketplace data, ruling out a linkage failure as an explanation. Second, the rate of prior background check reports (candidates who applied more than once) is not meaningfully different between extreme and non-extreme accounts (1.0% versus 0.8%), which argues against the imbalance reflecting contacts miscounted across separate application episodes. Third, extreme accounts are not more likely to be flagged for a profile-issues hold (0.8% versus 0.7%). Taken together, we do not find evidence that the extreme accounts reflect a coding or data-linkage error.

E.2 Formal Tests of Randomization Consistency

We next ask whether the treatment-control imbalance in the tail is itself distinguishable from what we would expect under random assignment alone, independent of whether the accounts driving it look "real." Table E.3 reports several complementary tests.

A Fisher's exact test on whether an account exceeds 100 contacts is not significant at conventional levels ($p = 0.072$); the same test using the 99th percentile threshold is similarly not significant ($p = 0.182$). A permutation test on the raw treatment-control mean difference in contacts is significant ($p = 0.021$), closely matching the naive OLS result reported in the main text. However, permutation tests targeting the tail directly—the difference in the number of accounts exceeding 100 contacts ($p = 0.073$), and the difference in the summed contacts of the twenty most active accounts in each arm ($p = 0.449$)—are not significant.

We interpret this pattern as an instance of the phenomenon documented by Young (2019): experimental estimates that appear statistically significant under standard asymptotic inference can be driven by a small number of outlier observations, and this significance frequently does not survive randomization inference. In our setting, the mean-based estimator–rather than the underlying imbalance in who lands in the tail–appears to be what drives the marginal significance of the naive comparison. This reading is consistent with Müller (2025)'s robust t-test, which we already report in Table E.1 and which is designed to deliver valid inference in exactly this kind of heavy-tailed, non-normal setting: it does not reject the null of no treatment effect at conventional levels for either contacts or reviews in the always-clear

[20] Extreme accounts also report modestly more service categories (9.2 versus 6.6) and active services (0.9 versus 0.7), though the differences are smaller in relative terms than the contact-activity measures.

group.[21]

E.3 Temporal Robustness

A natural alternative explanation is that the imbalance is concentrated in a single month—for instance, if a data processing issue analogous to the one described in Appendix D affected a specific cohort of background checks. Figure E.4 plots the monthly share of always-clear accounts exceeding 100 contacts by treatment arm, with 95% confidence intervals, excluding the first and last months of the experimental window (July 2023 and September 2024), which contain too few observations to be informative.[22]

We do not see evidence of a single month driving the pattern; confidence intervals overlap for most months, and no month shows a stark, isolated departure.

We formalize this with two additional checks. Absorbing month fixed effects moves the treatment coefficient from -3.48 to -3.28, a modest attenuation that suggests month composition is not doing substantial work. We also re-estimate the tail-count permutation test leaving out one month at a time; Table E.4 reports the resulting statistic and p-value for each of the fifteen exclusions.

Fourteen of the fifteen exclusions leave the statistic qualitatively similar to the full-sample result ($p = 0.073$); dropping May 2024 alone pushes the tail-count difference to conventional significance ($p = 0.024$). Finding one exclusion out of fifteen that crosses the 5% threshold is itself unsurprising given the number of tests involved, and no single month's removal collapses or reverses the pattern in the opposite direction. We view this as broadly consistent with the imbalance being spread across the experimental period rather than attributable to a single problematic month or processing batch, though we note this result is not unanimous across exclusions.

[21] As a further, more informal characterization of how heavy the right tail of the contact distribution is, we estimate a Hill (1975) tail index on the upper 5 percent of the distribution and obtain $\hat{\alpha} \approx 1.57$ (Table E.3). A tail index below 2 is associated with a formally infinite-variance distribution, under which the sample mean need not converge to a normal distribution at the usual $\sqrt{n}$ rate. We view this as suggestive rather than definitive, since Hill-type estimators are sensitive to the choice of threshold and are more commonly used in the income- and firm-size-distribution literatures; we report it only as an additional illustrative quantification of tail heaviness.

[22] July 2023 reflects only the final week of that month, since the experiment began July 24, 2023; September 2024 contains 11 always-clear observations in total.

E.4 Category and Service Composition

Finally, we test whether the treatment and control groups differ in the mix of services offered within the always-clear group, which could independently predict contact volume. Table E.5 reports both a chi-squared test on the top 15 occupation categories and a joint F-test from a regression of treatment assignment on occupation-category indicators.

Neither is significant ($p = 0.344$ and $p = 0.311$, respectively). The same tests applied to the platform's primary service-category classification are marginally significant ($p = 0.036$ and $p = 0.026$). Given the number of tests reported across this appendix, one or two results crossing conventional significance thresholds is within what we would expect from chance alone,[23] but because this result is close to conventional significance, we probe it further with three additional checks.

Table E.6 reports the results of three additional checks.

First, we test each of the fifteen service categories individually against the rest of the always-clear sample. Two are significant on their own—Central Air Conditioning Repair or Maintenance ($p = 0.031$) and Pressure Washing ($p = 0.013$), both underrepresented in the treatment arm—and two are marginal—General Contracting ($p = 0.055$) and Lawn Mowing and Trimming ($p = 0.061$), both overrepresented in the treatment arm. Finding two clearly significant results out of fifteen categories tested is somewhat more than the roughly one we would expect by chance, though not dramatically so.

Second, we ask whether these categories plausibly explain the underlying contact-count pattern by comparing their average contact volume to the sample as a whole. The two categories overrepresented in treatment—General Contracting and Lawn Mowing and Trimming—have contact volumes well above the sample average (29.8 and 40.6 contacts, respectively, against a sample mean of 21.1), while the two categories overrepresented in control have contact volume close to the sample average. This is the wrong direction to explain the observed pattern: if these categories were driving the result, having relatively more high-contact categories in the treatment arm should push the treatment group's average contacts up, not down. Consistent with this, adding category fixed effects to the specification in Table 4 leaves the treatment coefficient on contacts effectively unchanged, moving from -3.48 to -3.73, and the coefficient on the extreme-tail indicator moves from -0.0083 to -0.0088. If category composition were confounding the contacts result, we would expect these coefficients to shrink toward zero once we adjust for it; instead, they are stable or move marginally in

[23] Naively, with roughly 20 independent tests conducted across Tables E.1–E.5 at the 5 percent level, we would expect approximately one false positive.

the opposite direction.

Third, we test whether the same category imbalance appears in the full experimental sample, where we have roughly twenty-six times as many observations and correspondingly more power to detect small imbalances. It does not: a chi-squared test on the same fifteen-category grouping in the full sample of 253,799 candidates yields $p = 0.492$, and the corresponding joint F-test similarly returns $p = 0.492$. We interpret the always-clear-specific result as consistent with the kind of imbalance a single subgroup-specific test can produce by chance out of the many reported in this appendix, rather than a systematic pattern in how service categories were randomized, and we do not find evidence that it explains the contact-count imbalance documented above.

Taken together, these results suggest that the treatment-control gap in contact counts within the always-clear group reflects sampling variation in a heavy-tailed outcome rather than a coding or data-linkage error. The accounts driving the imbalance are established, verifiably active providers; the imbalance in the propensity to be in the tail is not itself distinguishable from randomization noise; it is not concentrated in a single month; and it is not explained by differences in the services offered. We note, however, that this exercise cannot rule out every possible confound.

E.5 Robustness in the New-User, No-Prior-Contact Subsample

We repeat this exercise on two review-based outcomes – average rating and cumulative verified reviews – restricting further to new-user accounts with no contact activity before their background check completed (`new_user_exp==1 & any_before_created==0`), and, as with Tables E.2–E.6 above, excluding the December 2023 problem period described in Appendix D. This restriction sharpens the test: these accounts have no criminal record to disclose and little pre-existing platform history, so treatment assignment should be uncorrelated with downstream ratings and reviews if the treatment matrix is randomizing as intended. We define the low-star group as accounts with average rating ≤ 3 stars, and the extreme verified-reviews group as accounts above the 95th percentile (7.0 verified reviews).

E.5.1 Characteristics of the Tail Groups

Table E.7 compares each tail group to the rest of the subsample. Low-star accounts average somewhat more contacts before their background check (4.2 versus 3.0) and are more likely to have any pre-background-check contact at all (62.7% versus 49.9%). The verified-reviews tail group looks more different still: nearly four times as many contacts before background check

completion (4.3 versus 1.1), more service categories (9.2 versus 5.7), and, most strikingly, close to thirty times as many contacts in the twelve months after background check completion (241.3 versus 8.4) – consistent with these being accounts that became unusually active providers after onboarding, similar in kind to the extreme-contact accounts characterized in Table E.2 above.

E.5.2 Formal Tests of Randomization Consistency

Table E.8 reports the randomization tests for both outcomes side by side. For average rating, several are significant – the low-star Fisher's exact test ($p = 0.030$) and the permutation test on the mean difference ($p = 0.012$) both reject at conventional levels, though the bottom-p5 Fisher test does not ($p = 0.133$). Because average rating is a bounded outcome, we substitute a review-count mediation check for the Hill tail index; the treatment coefficient on rating moves only modestly when controlling for review count (from -0.155 to -0.136), so review volume does not appear to explain the gap.

Verified reviews produce the sharpest result in the appendix. Both the 95th-percentile Fisher's exact test ($p = 0.007$) and the tail-count permutation test ($p = 0.002$) are significant, as is the mean-difference permutation test ($p = 0.018$). This stands in direct contrast to the corresponding contacts-based tests in Table E.3, where the analogous tail-count permutation test was not significant ($p = 0.073$) and we attributed the marginally significant mean-difference test to the heavy-tailed-outcome phenomenon documented by Young (2019). Here, the tail-specific tests are themselves significant – materially stronger evidence against pure randomization noise. The Hill tail index ($\hat{\alpha} = 1.856$) indicates a heavy right tail, consistent with the concentration of activity in a small number of accounts noted in Table E.7.

E.5.3 Temporal Robustness

Table E.9 reports the leave-one-month-out results for both outcomes, and Figure E.5 plots the monthly rate by treatment arm. The average-rating pattern is significant in 8 of 15 monthly exclusions, with several of the remainder sitting just above the 5% threshold (0.056, 0.056, 0.068, 0.060) rather than falling apart – a real but noisier effect. The verified-reviews pattern is far more uniform: unlike the contacts-based result, where fourteen of fifteen exclusions left the statistic unchanged and only one crossed into significance, here every single monthly exclusion is significant at $p < 0.05$, several well below (April 2024 at $p = 0.000$). Neither pattern is being carried by a single cohort or seasonal spike.

E.5.4 Category and Service Composition

Table E.10 checks category and occupation composition for both outcomes. Neither is significant for average rating ($p = 0.391$ and $p = 0.211$ for occupation; $p = 0.292$ and $p = 0.217$ for category). For verified reviews, occupation composition is balanced ($p = 0.317$ and $p = 0.286$), but the platform's primary service-category classification is marginally imbalanced ($p = 0.026$ and $p = 0.022$) – close to the same magnitude that prompted the category deep dive in Table E.6 above. We probe this the same way, in Table E.11.

Two categories are individually significant – General Contracting ($p = 0.011$) and Pressure Washing ($p = 0.022$) – with the remaining fourteen balanced. The confound check in Panel B is the important part: adding category fixed effects leaves both the treatment coefficient on verified reviews (-0.393 to -0.390) and the coefficient on the extreme-tail indicator (-0.0139 to -0.0137) essentially unchanged, and the full-experimental-sample category balance test is not significant ($p = 0.502$ on both the chi-squared and joint F-test). As in Table E.6, we do not find evidence that category composition explains the imbalance.

E.5.5 Adjudication Timing and Assessment Pathway

Having ruled out composition, we check two administrative channels specific to how always-clear cases are processed. First, the 12-month verified-reviews window is anchored to the adjudication timestamp rather than report completion; if that timestamp lagged report completion differentially by treatment, the count would be biased mechanically rather than substantively. It does not: adjudication lag averages 4.8 days in control and 5.0 days in treatment, a difference of -0.156 days (SE 0.224, $p = 0.485$).

Second, we check whether treatment predicts routing through the platform's rare non-standard assessment pathway (`assessment_value==REVIEW` or `adjudication==POST_ADVERSE_ACTION`), rather than the default automated pathway that covers 99.8% of always-clear cases. Only 17 accounts in the entire subsample go through this pathway (4 control, 13 treatment), a nominally significant imbalance (Fisher's exact $p = 0.050$ two-sided, $p = 0.029$ one-sided). Given the small N, this cannot plausibly account for the tail imbalance documented in Tables E.8 and E.9, which involves hundreds of accounts; we report it for completeness rather than as an explanation.

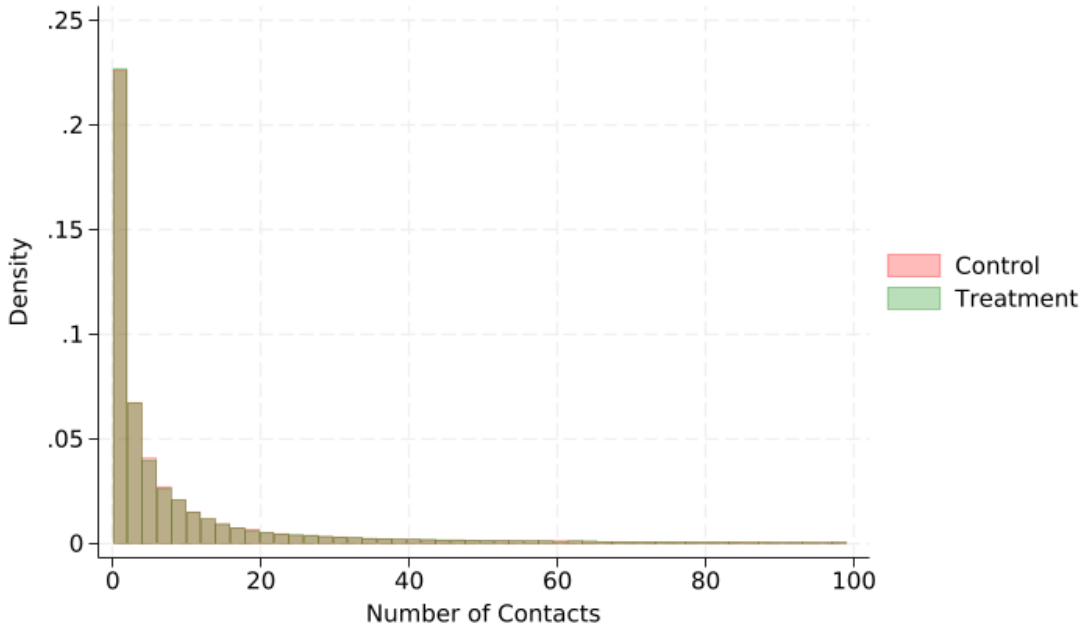
We separately confirmed that the 12-month window has fully elapsed for all but 2 of the 7,195 accounts as of the most recent data pull, so right-censoring of the review-count window is not a live explanation either: the treatment coefficient among accounts with a fully elapsed window (-0.394, SE 0.172) is essentially identical to the full-sample estimate in Table E.8

(−0.393, SE 0.172).

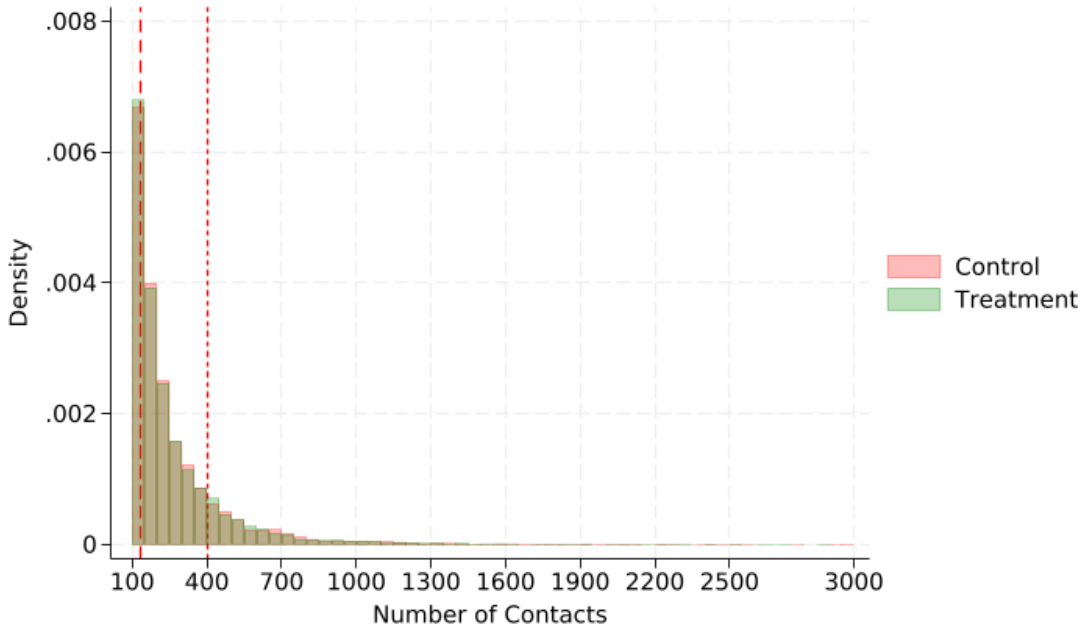
E.5.6 Interpretation

In the less restrictive always-clear sample above, the analogous contacts-based tail tests were not distinguishable from randomization noise, and we read that result as consistent with sampling variation in a heavy-tailed outcome. In this narrower subsample of newly joining pros, restricted to accounts with no criminal record to disclose and little pre-existing platform history, the verified-reviews tail imbalance is significant on two independent tests, is robust to excluding any single month, and is not explained by service-category composition, adjudication timing, or window-censoring. Because this is a placebo group by construction, there is no channel through which the treatment matrix should legitimately affect these accounts' downstream reviews. The average-rating result points in the same direction but is noisier and does not clear the bar on every cut.

Figure E.1: Distribution of Number of Contacts Overall Allowed Group by Treatment



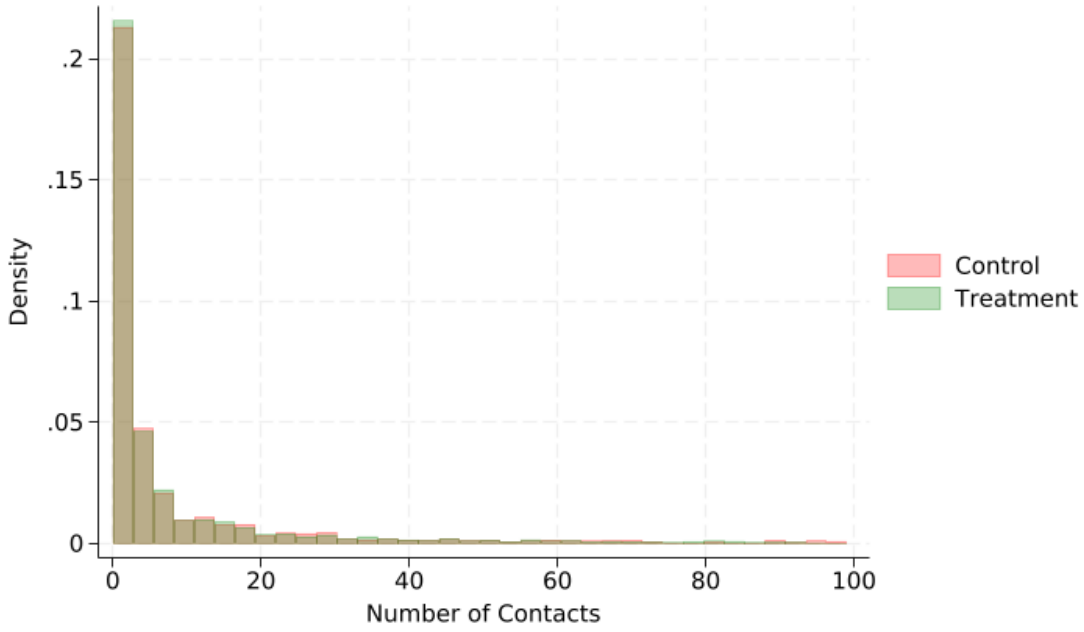
(a) Conditional on less than 100 Contacts



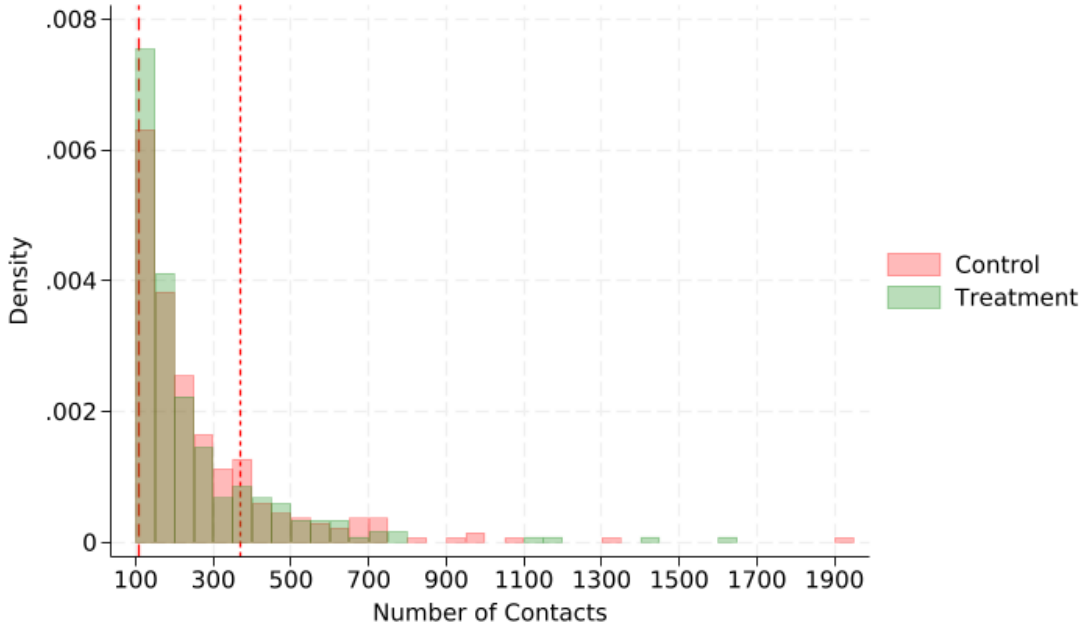
(b) Conditional on More than 100 Contacts

Note: Vertical dashed lines represent the 95th and 99th percentile in the contact distribution respectively among the allowed group.

Figure E.2: Distribution of Number of Contacts in Always Clear Group by Treatment



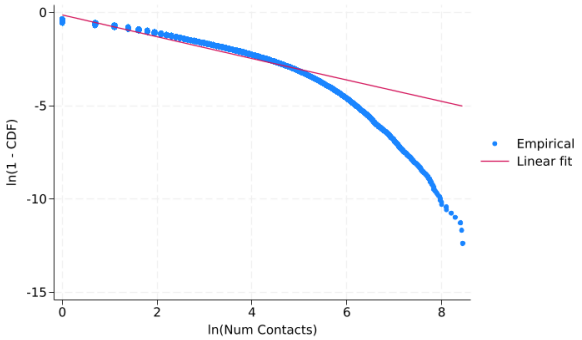
(a) Conditional on less than 100 Contacts



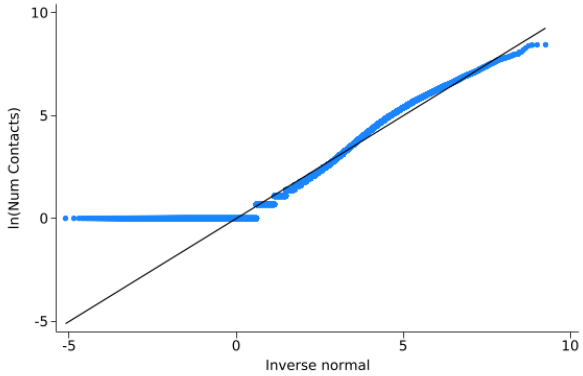
(b) Conditional on More than 100 Contacts

Note: Vertical dashed lines represent the 95th and 99th percentile in the contact distribution respectively among the always clear group.

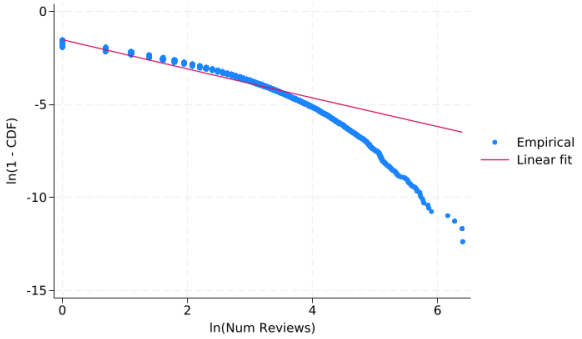
Figure E.3: Distribution Checks for Number of Contacts and Reviews



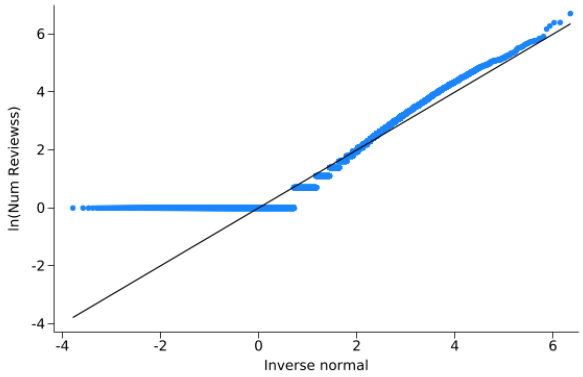
(a) Pareto: Contacts



(b) Log Normal: Contacts



(c) Pareto: Reviews



(d) Log Normal: Reviews

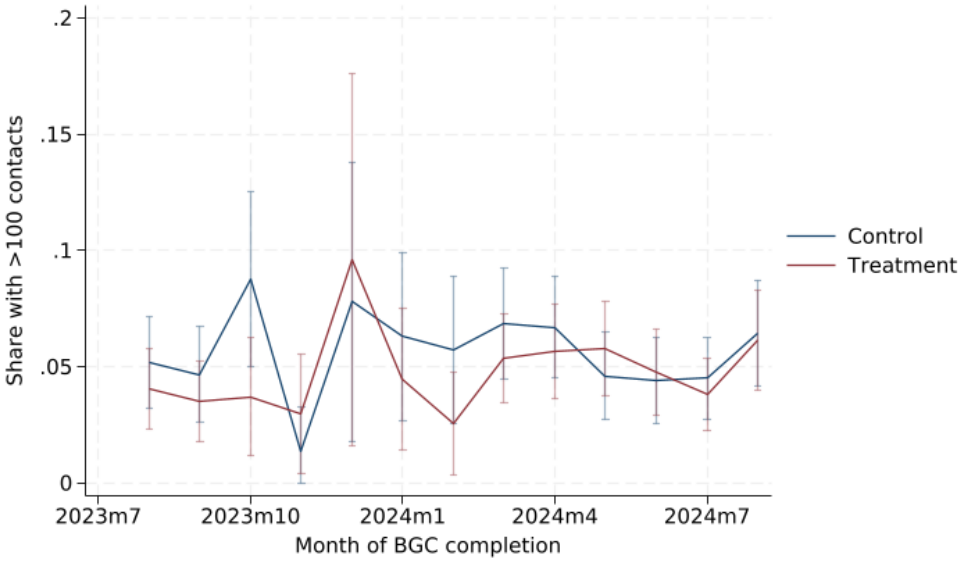
Note: This figure presents standard distribution checks for whether the distributions of number of contacts (top row) and number of reviews (bottom row) are Pareto or log normal. The top row shows the whole sample; the bottom row is conditional on being allowed on the platform.

Table E.3: Formal Tests of Randomization Consistency (Always Clear Group)

Test	Statistic	p-value (two-sided)
Fisher's exact, >100 contacts	–	0.072
Fisher's exact, p99 threshold	–	0.182
Permutation: mean difference	-3.484	0.021
Permutation: tail-count difference	-31	0.073
Permutation: top-20 sum difference	-3223	0.449
Hill tail index (top 5%)	1.565	–

Note: All tests restricted to the always-clear group (December 2023 problem period excluded). Fisher's exact tests report the exact probability of observing a treatment/control split in the tail-membership indicator (exceeding 100 contacts, or exceeding the sample's 99th percentile) at least as extreme as observed, under the null that treatment assignment is independent of tail membership. Permutation tests reshuffle treatment assignment 1,000 times, respecting the observed group sizes, and report the p-value as the share of reshuffled draws whose statistic is at least as extreme as the one computed on the actual data, following Young (2019). The tail-count difference is the number of treatment-arm accounts exceeding 100 contacts minus the number of control-arm accounts exceeding 100 contacts. The top-20 sum difference sums total contacts across the 20 highest-contact accounts in the sample, computed separately by treatment arm and differenced, and asks whether the very largest individual outliers are disproportionately concentrated in one arm. The Hill (1975) tail index is estimated on the upper 5 percent of the contact distribution; values below 2 are associated with a formally infinite-variance distribution.

Figure E.4: Extreme Contact Rate by Month and Treatment (Always Clear Group)



Note: Sample restricted to the always-clear group, excluding the December 2023 problem period, July 2023 (a partial month, since the experiment began July 24, 2023), and September 2024 (11 always-clear observations). Lines show the share of accounts exceeding 100 contacts in each month, by treatment arm, with 95 percent Wald confidence intervals ($\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n}$).

Table E.4: Leave-One-Month-Out Permutation Results (Always Clear Group)

Month excluded	Tail-count statistic	p-value (two-sided)
Jul 2023	-31	0.056
Aug 2023	-26	0.108
Sep 2023	-27	0.132
Oct 2023	-20	0.196
Nov 2023	-34	0.064
Dec 2023	-30	0.076
Jan 2024	-28	0.080
Feb 2024	-24	0.108
Mar 2024	-29	0.128
Apr 2024	-25	0.072
May 2024	-38	0.024
Jun 2024	-34	0.056
Jul 2024	-29	0.080
Aug 2024	-31	0.072
Sep 2024	-28	0.080

Note: Each row excludes one calendar month of background check completions and re-estimates the permutation test on the tail-count difference (defined in the note to Table E.3) using the remaining always-clear sample, with 500 permutations per exclusion.

Table E.5: Category/Service Balance by Treatment (Always Clear Group)

Grouping	Chi2 p-value	Joint F-test p-value
Occupation (top 15)	0.344	0.311
Most-quoted category (top 15)	0.036	0.026

Note: Chi-squared tests compare the distribution of the top-15 most common occupation or service-category groupings, with all remaining categories collapsed into an 'Other' bucket, across treatment and control. The joint F-test is from a regression of treatment assignment on the category-group indicators plus the number of categories and active services, testing whether the category dummies jointly predict treatment assignment. Sample restricted to the always-clear group.

Table E.6: Category Deep Dive: Per-Category Balance and Confound Checks (Always Clear Group)

Panel A: Per-category balance, sorted by Fisher's exact p-value

Category	N	Control	Treatment	Fisher's p	Mean contacts	Extreme rate
Pressure Washing	139	83	56	0.013	20.9	0.072
Air Conditioning Repair	87	53	34	0.031	21.2	0.057
General Contracting	122	49	73	0.055	29.8	0.074
Lawn Mowing and Trimming	263	114	149	0.061	40.6	0.091
Handyman	1084	508	576	0.122	30.5	0.080
Furniture Assembly	95	41	54	0.258	13.6	0.032
Furniture Moving	105	46	59	0.282	34.1	0.067
House Cleaning	952	482	470	0.339	13.3	0.026
Interior Painting	330	171	159	0.341	24.2	0.052
Electrical Repair	214	110	104	0.534	47.7	0.112
Junk Removal	279	142	137	0.585	27.8	0.075
Tree Trimming	81	38	43	0.738	21.0	0.037
Local Moving	256	128	128	0.800	37.4	0.098
Missing	2484	1223	1261	0.907	0.0	0.000
Appliance Repair	86	43	43	0.914	75.8	0.116
Other (rank >15)	3028	1487	1541	1.000	29.5	0.075

Panel B: Confound checks

Treatment coef. on contacts, without category FE	−3.484 (SE 1.545)
Treatment coef. on contacts, with category FE	−3.727 (SE 1.535)
Treatment coef. on extreme-tail indicator, without category FE	−0.0083 (SE 0.0045)
Treatment coef. on extreme-tail indicator, with category FE	−0.0088 (SE 0.0045)
Full-sample category balance	χ^2 $p = 0.492$; joint F $p = 0.492$

Note: Panel A reports Fisher's exact tests comparing each category against all others combined, sorted by p-value; mean contacts and the extreme-tail rate (share exceeding 100 contacts) are reported for context and are not part of the balance test itself. Panel B reports the treatment coefficient on total contacts and on the extreme-tail indicator, with and without category fixed effects, to test whether category composition confounds the contacts result; and the chi-squared and joint F-test statistics from applying the same category-balance test (Table E.5) to the full experimental sample of 253,799 candidates rather than the always-clear subgroup alone.

Table E.7: Characteristics of Tail Accounts (Always Clear, New-User–No-Prior-Contact Sample)

Panel A: Average Rating

	Rest (>3 stars)	Low-Star (≤3 stars)	Total
Contacts before BGC	3.023	4.179	3.120
	(6.415)	(6.048)	(6.389)
Any contact before BGC	0.499	0.627	0.509
	(0.500)	(0.487)	(0.500)
Days on platform before BGC	12.63	9.224	12.34
	(34.03)	(24.34)	(33.32)
Number of categories	8.300	6.746	8.169
	(9.188)	(8.010)	(9.101)
Number of active services	0.811	0.612	0.794
	(0.491)	(0.521)	(0.497)
Number of reviews (all)	17.48	4.269	16.37
	(27.53)	(5.887)	(26.65)
Number of verified reviews (12mo)	11.30	1.687	10.49
	(18.91)	(1.270)	(18.29)

Panel B: Verified Reviews

	Non-Extreme	Extreme (>p95)	Total
Contacts before BGC	1.088	4.289	1.243
	(3.105)	(8.349)	(3.608)
Any contact before BGC	0.352	0.550	0.362
	(0.478)	(0.498)	(0.481)
Days on platform before BGC	16.53	11.89	16.31
	(42.69)	(31.16)	(42.21)
Number of categories	5.729	9.160	5.895
	(7.069)	(10.13)	(7.283)
Number of active services	0.650	0.860	0.660
	(0.577)	(0.449)	(0.573)
Contacts (12mo after)	8.383	241.3	19.68
	(26.23)	(245.8)	(77.99)

Note: Mean coefficients reported, with standard deviations in parentheses. Panel A: N = 730 rest, 67 low-star (≤3 stars, verified average rating). Panel B: N = 6,846 non-extreme, 349 extreme (>p95 = 7.0 verified reviews). Sample restricted to always_clear==1 & new_user_exp==1 & any_before_created==0, excluding the December 2023 problem period.

Table E.8: Formal Tests of Randomization Consistency (Always Clear, New-User–No-Prior-Contact Sample)

	Average Rating		Verified Reviews	
Test	Statistic	p-value	Statistic	p-value
Fisher's exact, primary threshold	–	0.133	–	0.007
Fisher's exact, secondary threshold	–	0.030	–	0.341
Permutation: mean difference	−0.155	0.012	−0.393	0.018
Permutation: tail-count difference	17	0.036	−43	0.002
Permutation: top/bottom-20 sum difference	8.000	0.102	−714	0.120
Average rating: review-count mediation check (outcome is bounded, no Hill index)				
Treatment coef., unconditional	−0.155 (SE 0.066)			
Treatment coef., controlling for num_reviews	−0.136 (SE 0.065)			
Verified reviews: tail-heaviness check				
Hill tail index (top 5%)			$\hat{\alpha} = 1.856$	

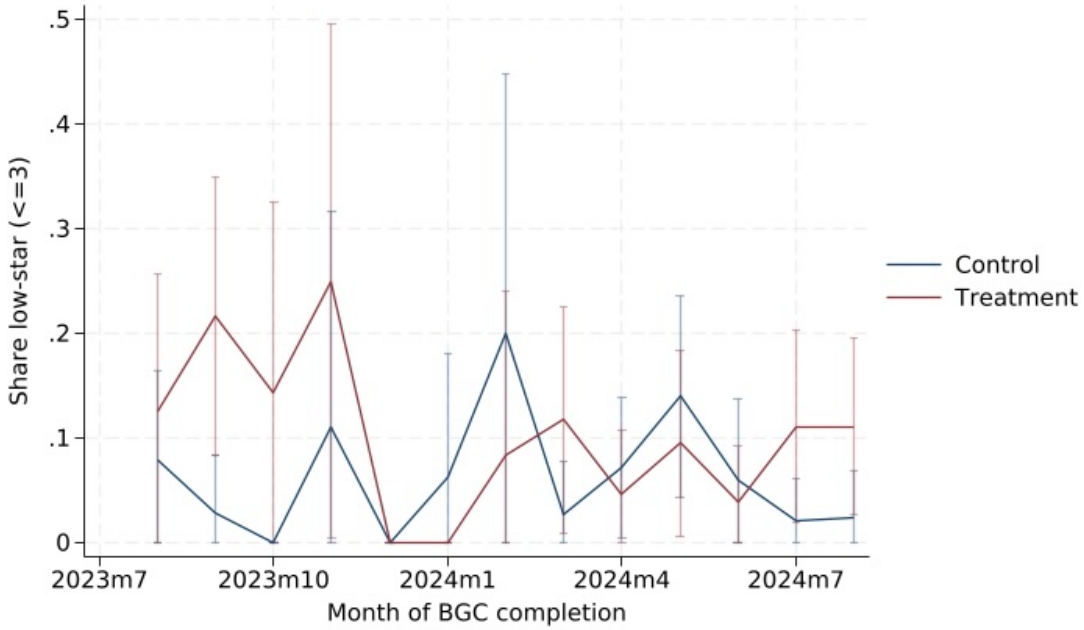
Note: Primary/secondary thresholds differ by outcome: for average rating, primary is the bottom-p5 cutoff and secondary is the low-star (≤3) convention used elsewhere in the paper; for verified reviews, primary is the 95th-percentile cutoff and secondary is the 99th percentile. Permutation tests reshuffle treatment assignment 1,000 times, following Young (2019). Sample restricted to always_clear==1 & new_user_exp==1 & any_before_created==0, excluding the December 2023 problem period.

Table E.9: Leave-One-Month-Out Permutation Results (Always Clear, New-User–No-Prior-Contact Sample)

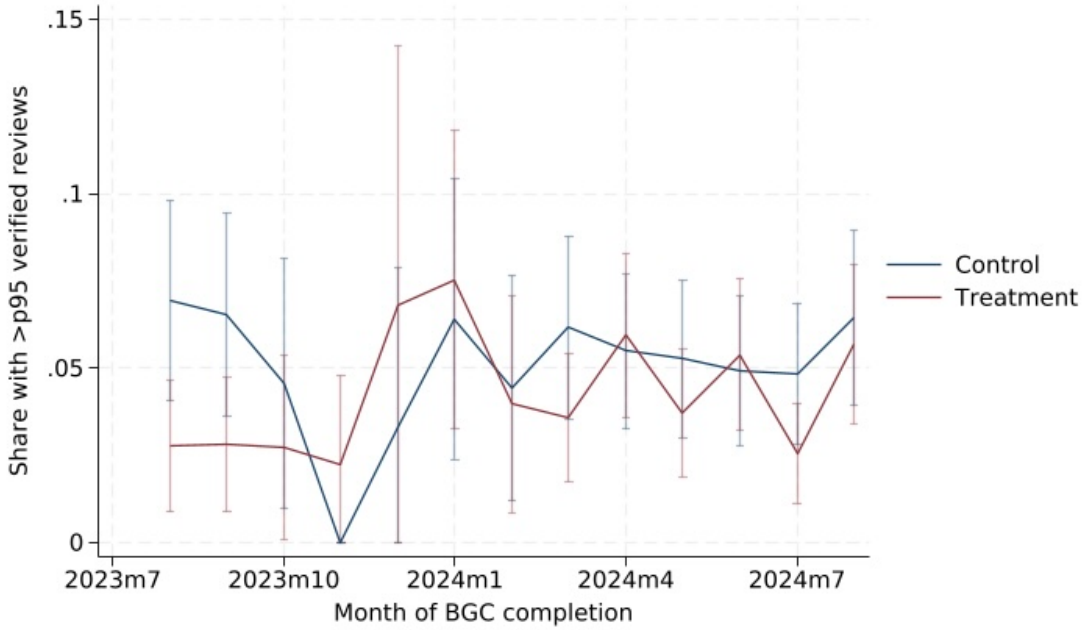
	Average Rating		Verified Reviews	
Month excluded	Low-star stat.	p-value	Tail-count stat.	p-value
Jul 2023	16	0.056	−41	0.008
Aug 2023	17	0.040	−30	0.008
Sep 2023	10	0.264	−33	0.048
Oct 2023	15	0.032	−41	0.016
Nov 2023	15	0.056	−46	0.004
Dec 2023	17	0.036	−44	0.004
Jan 2024	18	0.024	−45	0.012
Feb 2024	18	0.012	−42	0.004
Mar 2024	14	0.068	−37	0.024
Apr 2024	19	0.032	−44	0.000
May 2024	20	0.012	−38	0.012
Jun 2024	17	0.040	−46	0.008
Jul 2024	13	0.104	−34	0.044
Aug 2024	12	0.100	−41	0.008
Sep 2024	17	0.060	−40	0.016

Note: Low-star stat. is the tail-count permutation statistic for average rating (≤3 stars); tail-count stat. is the analogous statistic for verified reviews (>p95). Sample restricted to always_clear==1 & new_user_exp==1 & any_before_created==0, excluding the December 2023 problem period.

Figure E.5: Tail Rate by Month and Treatment (Always Clear, New-User–No-Prior-Contact Sample)



(a) Low-Star Rate (≤3 stars), Average Rating



(b) Extreme Rate (>p95), Verified Reviews

Note: Sample restricted to always_clear==1 & new_user_exp==1 & any_before_created==0, excluding the December 2023 problem period, July 2023 (a partial month, since the experiment began July 24, 2023), and September 2024 (too few always-clear observations in this subsample to be informative). Lines show 95 percent Wald confidence intervals by treatment arm.

Table E.10: Category/Service Balance by Treatment (Always Clear, New-User–No-Prior-Contact Sample)

	Average Rating		Verified Reviews	
Grouping	Chi2 p	Joint F p	Chi2 p	Joint F p
Occupation (top 15)	0.391	0.211	0.317	0.286
Most-quoted category (top 15)	0.292	0.217	0.026	0.022

Note: The chi-squared test compares the category distribution across treatment and control. The joint F-test is from a regression of treatment assignment on category-group indicators, controlling for number of categories and number of active services. Sample restricted to always_clear==1 & new_user_exp==1 & any_before_created==0, excluding the December 2023 problem period.

Table E.11: Category Deep Dive: Per-Category Balance and Confound Checks, Verified Reviews (Always Clear, New-User–No-Prior-Contact Sample)

Panel A: Per-category balance, sorted by Fisher's exact p-value

Category	N	Control	Treatment	Fisher's p	Mean reviews	Extreme rate
General Contracting	107	39	68	0.011	0.4	0.000
Pressure Washing	112	67	45	0.022	1.8	0.062
Air Conditioning Repair	68	40	28	0.114	1.7	0.088
Home Remodeling	68	40	28	0.114	0.4	0.015
Electrical Repair	150	83	67	0.137	3.8	0.113
Handyman	878	412	466	0.195	2.6	0.109
House Cleaning	771	395	376	0.195	0.8	0.023
Furniture Assembly	74	31	43	0.243	1.2	0.054
Furniture Moving	95	41	54	0.258	3.5	0.084
Lawn Mowing and Trimming	187	86	101	0.416	3.2	0.075
Junk Removal	234	120	114	0.506	3.2	0.103
Local Moving	201	103	98	0.521	3.1	0.104
Missing	1832	891	941	0.725	0.0	0.000
Interior Painting	260	125	135	0.801	1.2	0.042
Other (rank >15)	2088	1019	1069	0.835	1.6	0.055
Appliance Repair	70	34	36	1.000	7.8	0.114

Panel B: Confound checks

Treatment coef. on verified reviews, without category FE	−0.393 (SE 0.172)
Treatment coef. on verified reviews, with category FE	−0.390 (SE 0.172)
Treatment coef. on extreme-tail indicator, without category FE	−0.0139 (SE 0.0051)
Treatment coef. on extreme-tail indicator, with category FE	−0.0137 (SE 0.0050)
Full-sample category balance	χ^2 $p = 0.502$; joint F $p = 0.502$

Note: Panel A reports Fisher's exact tests comparing each category against all others combined, sorted by p-value; mean verified reviews and the extreme-tail rate (share above the 95th percentile) are reported for context and are not part of the balance test itself. Panel B reports the treatment coefficient on verified reviews and on the extreme-tail indicator, with and without category fixed effects, to test whether category composition confounds the verified-reviews result; and the chi-squared and joint F-test statistics from applying the same category-balance test (Table E.10) to the full experimental sample rather than the always-clear, new-user–no-prior-contact subsample alone.

Table E.12: Adjudication Timing and Assessment Pathway (Always Clear, New-User–No-Prior-Contact Sample)

	Control	Treatment	
Adjudication lag (days, adjudicated_at minus report_completed_at)			
Mean	4.847	5.004	
N	3,525	3,669	
Difference: -0.156 (SE 0.224), $p = 0.485$			
Rare non-standard assessment pathway			
	Control	Treatment	Total
N routed non-standard	4	13	17
N total	3,526	3,669	7,195
Fisher's exact $p = 0.050$ (two-sided), $p = 0.029$ (one-sided)			

Note: Adjudication lag is tested via a two-sample t-test and an OLS regression of lag on treatment, robust standard errors. The non-standard-pathway flag is `assessment_value==REVIEW` or `adjudication==POST_ADVERSE_ACTION`, capturing accounts routed through the platform's rare manual review/adverse-action process rather than the default automated (ELIGIBLE/ENGAGED) pathway that covers 99.8% of always-clear cases. At only 17 total cases, this cannot plausibly explain the tail imbalance documented in Tables E.8 and E.9, which involves hundreds of accounts; it is reported here for completeness rather than as an explanation. Sample restricted to always_clear==1 & new_user_exp==1 & any_before_created==0, excluding the December 2023 problem period.