An Experimental Evaluation of Deferred Acceptance: Evidence from Over 100 Army Officer Labor Markets*

Jonathan M.V. Davis Kyle Greenberg Damon Jones

University of Oregon *West Point* *University of Chicago*

August 20, 2023

Abstract

We present evidence from a randomized trial of the impact of matching workers to jobs using the deferred acceptance (DA) algorithm. Our setting is the U.S. Army's annual many-to-one marketplace that matches 10,000 officers to units. Officers and jobs are partitioned into over 100 distinct markets, our unit of randomization. Matching with DA reduced officers' attrition in their first year in their new match by 16.7 percent, but we can rule out more than a 10 percent reduction in attrition by the end of their second year. Matching with DA had precise zero effects on performance evaluations and promotions. Although matching with DA increased truthful preference reporting by a statistically significant 10 percent, many officers matched by DA misreport their true preferences. We present new evidence suggesting that communication and coordination of preferences may limit the benefits of DA in matching markets where each side actively ranks the other.

*All views expressed in this manuscript are those of the authors and do not represent the views or official positions of the United States Military Academy, the United States Army, or the Department of Defense. Andrew Hoover and Nina Kerkebane provided exceptional research assistance. We thank Alex Rees-Jones and participants at LERA at ASSA, SOLE, NBER SI Labor Studies and Personnel Economics programs, and Stanford SITE for helpful comments. Although Kyle Greenberg is a lieutenant colonel in the U.S. Army, none of the authors received financial compensation from the Department of Defense for the purpose of completing this project. This research is supported by NSF Grant SES-2018246 and a J-PAL Workforce of the Future Grant. This RCT was registered in the American Economic Association Registry under trial number AEARCTR-0004718.

1 Introduction

Over the last three decades, insights from market design have had a substantial influence on matching markets without prices. Results from this field have been used to help redesign mechanisms to match new doctors to hospitals (Roth and Peranson, 1999), assign students to schools in several large school districts (Abdulkadiroğlu and Sönmez, 2003; Abdulkadiroğlu et al., 2009; Pathak and Sönmez, 2008), match newly commissioned Army officers to military occupations (Sönmez, 2013; Sönmez and Switzer, 2013; Greenberg et al., 2023), and match kidney donors to recipients (Roth et al., 2003). In addition, many organizations have adopted tools from market design to facilitate internal worker-to-division matching markets (Barron and Vardy, 2005; Cowgill and Koning, 2018; Cowgill et al., 2022; Davis, 2022).

Stability, the match property in which no unmatched pair of agents prefer being matched together over their assigned match, is considered a crucial characteristic of successful market design interventions (Roth, 1984, 1990, 1991; Roth and Xing, 1994; Roth and Peranson, 1999; Kagel and Roth, 2000). When a match is unstable, some agents have "justified envy" which creates opportunities to deviate from the assigned match. The deferred acceptance (DA) algorithm is the simplest and most widely used example of a stable matching algorithm (Gale and Shapley, 1962; Roth, 2008). Market designers have also advocated for DA because it is both strategyproof—that is, truthful reporting of rankings is a dominant strategy—and it yields the optimal stable match for the proposing side of the market (Abdulkadiroğlu and Sönmez, 2003; Abdulkadiroğlu et al., 2006; Chen and Sönmez, 2006; Pathak and Sönmez, 2008). However, these theoretical benefits might not be realized in practice. For example, if participants do not understand or trust that the algorithm is strategyproof, they may still misreport their preferences (Rees-Jones, 2017). Or they may try to reduce their uncertainty about their match by entering into informal agreements with agents on the other side of the market to "rank each other first" (Roth and Xing, 1994).

This paper provides evidence from a randomized controlled trial evaluating the impacts

of matching workers to jobs using the DA algorithm. Our setting is the internal market within the United States Army where officers are matched to units. Officers generally rotate units within the Army every three years. The Army's Human Resources Command (HRC) coordinates this reorganization in an annual matching market that includes over 14,000 officers to be matched across roughly 500 units. Officers and potential positions at units are segmented into disjoint markets based on their rank and military occupation. We randomly assigned a subset of markets to a treatment group where officers and units were matched using an officer-proposing DA algorithm. The remainder are assigned to a control group where officers and units were matched according to the Army's traditional process, which involves career managers manually making match decisions using officer and unit preferences without the aid of any specific algorithm.

Our setting is ideal for two reasons. First, our experimental sample includes 115 disjoint markets. This provides a large enough sample to have comparable treatment and control groups and the statistical power to precisely estimate treatment effects of interest at the market level. Second, the Army collects high-quality data on participants' preferences and relevant market outcomes. We are able to link these data to surveys that ask participants about strategic preference manipulations and their satisfaction with their assigned matches. We also link to administrative performance and retention data. Together, these features provide us with the unique opportunity to credibly measure the impact of DA on immediate and longer-run market outcomes.

For our main outcomes, we selected and pre-registered proxies for well-known goals of market design mechanisms: retention within the Army, which proxies for officer satisfaction; officers' evaluation reports and promotions, which measure officer performance; and survey evidence on truthful preference reporting, which reflects whether DA is strategyproof for officers. Overall, we find little evidence that DA leads to meaningful improvements in retention, performance evaluations, or promotions two years into their new position. And while we find evidence that DA increases truthful preference reporting, effect sizes are modest and most officers in DA markets admit to misreporting preferences.

The officer-proposing DA yields the officer optimal stable match, which means that all officers prefer the match selected to all other stable matches. Consistent with this, matching with DA in our setting reduces attrition in the first post-match year by a statistically significant 1.1 percentage points ($p = 0.03$), a 16.7 percent reduction relative to the control group's attrition rate. However, this is sensitive to how we adjust for multiple hypothesis testing and by two years out matching with DA only reduces attrition by a statistically insignificant 0.3 percentage point (pp). Our results at two years are precise enough to rule out a 1.5 pp reduction in attrition, less than one-tenth of two-year attrition in the control group (15.1 percent).

To the extent that match quality can be improved by DA, we might expect improvements in productivity and promotion. However, matching with DA has precise zero effects on the probability that an officer receives the highest possible performance evaluation.[1] Roughly half of officers in the control group receive the highest possible performance evaluation in the first and second year after starting their new positions. Our estimates suggest that matching with DA has little impact on this performance outcome, with confidence intervals ranging from a 2.2 pp reduction up to a 1.0 pp increase in the first year and from a 3.2 pp reduction up to a 1.2 pp increase in the second year. We find similarly precise zero effects on the likelihood that an officer is promoted to the next rank and on officers' promotion board percentile ranking relative to other officers considered for promotion.

Another advantage of DA is that it is strategyproof for the proposing side of the market, meaning that truthful preference reporting is a dominant strategy for officers. We find moderate evidence in support of this prediction. In a survey administered three weeks before the marketplace closed, officers in DA markets are 2.4 pp more likely to report always submitting their true preferences ($p = 0.001$), a 10 percent increase on the 24 percent of officers in control. Officers in DA markets are also 1.2 pp more likely to state that they

[1] Officers' evaluation reports have a significant impact on whether an officer is promoted to the next rank. We only observe promotion outcomes for the subset of officers in our sample who were considered for promotion within two years post-match.

have accurately reported their top choice ($p = 0.061$), relative to the 84 percent of officers who indicated truthful reporting of their top choice in control group markets. However, in a post-market survey administered when officers learned of their matches, only 69 percent of officers stated that they truthfully reported their top choice, with no statistically significant difference between the treatment and control markets.

So what explains the relatively limited effects of DA on retention, performance, and truthful preference reporting? One possibility is that strategic cross-market communication between officers and units may limit the benefits of having a strategyproof mechanism. In both DA and control markets, roughly 45% of matches are "first-to-first" pairings—that is, a match where an officer has ranked a job listing as their first choice and where the unit has ranked the same officer as their first choice for that particular listing. This is true, even though officers' and units' exact rank-ordered preferences are hidden. We derive a test for the presence of strategic communication by leveraging cases where units have multiple, nearly identical job postings. In these cases, we can simulate the rate of first-to-first matches under the null hypothesis of no coordination. The high observed rate of such pairings far exceeds what we would expect under the null, and because we focus on sets of identical jobs at the same unit, the patterns are not simply the result of highly correlated preferences between officers and units. We further show that this pattern is not simply an artifact of the initial rollout of DA: the same patterns prevail two years after our study, by which time DA had been used in *all* markets by the Army and messaging of DA's properties had improved.[2]

If career managers typically honor first-to-first pairings because of the ease with which they can be identified without the help of a computer algorithm, then they will achieve many of the same matches as a DA algorithm. However, if some first-to-first matches in DA markets are strategic and differ from the true preferences of officers and units, the benefits of switching to DA may be attenuated, akin to the process described by Roth and

[2]In Appendix A, we show, using a simple theoretical example, that officers may benefit from strategic communication and misreporting their preferences if units' preferences are responsive to this communication (Antler, 2015).

Xing (1994) as "stage 4 unraveling." Survey evidence corroborates that officers frequently deviate from their true preferences to achieve a first-to-first pairing.

These limitations have implications for other settings where proximity and repeated interaction between those on either side of the market are non-trivial, such as in other internal labor markets or even the National Resident Matching Program (NRMP) for medical doctors. In fact, studies of physicians and residency program directors report frequent communication about rankings after interviews, even though this type of communication is prohibited by the NRMP code of conduct (Anderson et al., 1999; Carek et al., 2000; Teichman et al., 2000; Sbicca et al., 2010; Berriochoa et al., 2018).

An alternative explanation for the limited effects of DA found in our setting is that the officers may not have understood that truthful preference reporting is a weakly dominant strategy (Chen and Sönmez, 2002; Hassidim et al., 2017; Rees-Jones, 2018; Rees-Jones and Skowronek, 2018). Although career managers knew whether the officers they managed were part of treatment or control, we were not able to observe their exact messaging to officers in DA markets, which could result in considerable heterogeneity in awareness of DA's properties. However, we do not believe that this completely explains our results. In subsequent years, as DA was rolled out to the entire marketplace and when we have better information how DA's properties were communicated to officers, we observe similar patterns in 1-to-1 matching and self-reported truthfulness.

This paper contributes to several literatures. First, we contribute to the literature on market design by providing the first evidence from a randomized trial on the impact of matching workers to jobs using DA. Despite the rapidly growing adoption of market design tools over the past three decades, there is little empirical evidence of the impact of labor market design on market and match outcomes. The leading studies to date use time series comparisons within a single market (Niederle and Roth, 2003b), cross-sectional comparisons between markets (Niederle and Roth, 2003a), or difference-in-difference methods (Davis, 2022). One challenge is that experimental or quasi-experimental variation in mechanisms across multiple, comparable markets is uncommon. What is much more common

is a one-time conversion to a new mechanism within a single market (Roth and Peranson, 1999; Abdulkadiroğlu et al., 2005a,b). Counterfactual outcomes at the market level are therefore difficult to estimate when there is only one market under observation.

Our preference and survey data allow us to analyze how matching mechanisms deviate from theoretical ideals when implemented in practice. Recent research suggests that doctors misreport their true preferences in the incentive-compatible NRMP (Hassidim et al., 2017; Rees-Jones, 2018; Rees-Jones and Skowronek, 2018). In our setting, we can take advantage of the presence of identical postings within a unit to test for strategic behavior based on submitted rankings, which complements evidence from self-reported surveys. Relatedly, Echenique et al. (2022) posit that doctors' preferences for hospitals deviate from the truth in part because doctors can only preference hospitals they interview with.[3] Our results suggest that participants in two-sided markets may strategically communicate (often untruthful) preferences in order to obtain higher rankings from participants on the other side of the market, a possibility that is consistent with evidence from several surveys that typically consist of a few hundred doctors and residency program directors (Anderson et al., 1999; Carek et al., 2000; Teichman et al., 2000; Sbicca et al., 2010, 2012; Berriochoa et al., 2018). We contribute to prior evidence using a survey that is administered to all officers participating in the marketplace and that has a higher response rate than is typical in surveys of doctors in the NRMP.

Second, our study contributes to organizational and personnel economics by providing new evidence on the trade-offs of different assignment mechanisms within organizations. Market design tools are designed to produce matches with certain properties, such as incentive-compatibility, stability, transparency, and strategic simplicity, but organizations that organize these markets and participants in the market may have other competing objectives (Cowgill et al., 2022). For example, Haegele (2021) shows mangers' incentives to hoard talent on their teams creates a misallocation of workers to jobs within firms. Our

[3] In the context of centralized school choice markets with strategy-proof assignment mechanisms, Arteaga et al. (2022) find that beliefs about admissions probabilities influence choices by shaping how applicants search for schools.

study offers a unique opportunity to estimate improvements in officer satisfaction due to DA, as measured by retention in the Army, against any changes in officer performance evaluations, which we broadly interpret as a proxy for more general organizational objectives. We focus primarily on horizontal job changes within an organization. Huitfeldt et al. (2023) study the vertical structure of labor markets within a firm. Relatedly, Benson et al. (2019) study how firms make promotion decisions and show that they may prioritize current performance at the expense of other characteristics that are more predictive of managerial talent.

Third, this research adds to the literature on personnel considerations within military organizations. Several papers have studied the assignment of cadets to branches of the military (Sönmez, 2013; Sönmez and Switzer, 2013; Schlegel, 2015; Jagadeesan, 2019; Greenberg et al., 2023). Lewis et al. (2022) study the assignment of Coast Guard servicemen to ships and recommend focusing on the assignments of officers. Greenberg et al. (2022) study the long-term impacts of voluntary enlistment in the U.S. Army, Bruhn et al. (2023) study the effects of combat deployments, and several others study the long-term effects of compulsory military service in the U.S. and elsewhere (e.g., Angrist, 1990; Bound and Turner, 2002; Bedard and Deschênes, 2006; Angrist et al., 2010; Angrist and Chen, 2011; Angrist et al., 2011; Card and Cardoso, 2012; Bingley et al., 2020). More generally, our paper contributes to the broad literature on experiments in labor economics (Horton et al., 2011; Charness and Kuhn, 2011; List and Rasul, 2011). Our study is unique in that we randomize entire markets to understand how different market clearing rules affect outcomes and we focus on the impact of different matching mechanisms using experimental variation.

2 Institutional Details of the Internal Matching Market

Since 2017, the U.S. Army has used an online, interactive module to match most officers to positions at Army units within an internal labor market.[4] The online marketplace, known as the Army Talent Alignment Process (ATAP),[5] allows officers to build profiles that units can see (and vice-versa),[6] permits officers who are scheduled to change assignments within 6 to 9 months to submit preferences over available jobs, and allows units to submit preferences over officers expected to move. The Army's human resources division, known as the Human Resources Command (HRC), manages the marketplace and partitions all officers and jobs within the marketplace into distinct markets. Each market is defined by a combination of officer rank and officer occupation (e.g. a market for "infantry captains" and a separate market for "military intelligence majors").

In addition to being associated with a specific rank and a specific occupation, each job also belongs to one of roughly 500 different Army units. Although the distinct markets within the marketplace are many-to-one in the sense that multiple officers can match to a single unit, in practice officers submit preferences over specific jobs at a unit as part of a one-to-one market. For example, if a market has 10 units that each have 5 distinct jobs within the market, then each officer in that market can rank up to 50 jobs. Units provide descriptions for each job listed in the online marketplace, and job descriptions may vary when a market contains multiple job listings that belong to the same unit. Units with multiple job listings in the same market must submit separate rank-order lists of officers

[4]This internal labor market does not include new Second Lieutenants, who receive their initial assignment through their respective sources of commission (e.g. Officer Candidates School, the Reserve Officer Training Corps, or the United States Military Academy).

[5]The marketplace is also frequently referred to by the name of the online platform, the Assignment Interactive Module—version 2.0 (AIM2).

[6]Officer profiles include all information on the standard Officer Record Brief (including assignment history, civilian education degree information, military education, and military awards) and additional self-reported details, including previous civilian and military employment and education, professional skills and certifications, cultural experiences, and travel. Units can provide specific descriptions for each job in the marketplace, contact information for the job's current incumbent or point of contact, and general information about their unit.

for each listing, and these preferences need not be identical across listings, even when such job listings have identical descriptions.[7]

Each cycle of the online marketplace is open for 6 to 8 weeks, during which time officers may submit preferences for all jobs within their (rank-by-occupation) market.[8] Officers can adjust their preferences for jobs at any time while the marketplace is open. Likewise, units may submit preferences over officers and can change their preferences at any time. Officers are not required to rank-order all jobs in their market and jobs are not required to rank-order all officers. The final version of their preference lists at the market's scheduled closing date are used to determine (for DA) or inform (for control markets) matches.

Officers' exact preferences over jobs and units' exact preferences over officers are hidden from each other. However, units can observe a signal if an officer ranks one of the unit's jobs among the officer's top 10 percent of all possible choices. For example, if a market has 200 jobs, then the officer interest signal will appear next to 20 jobs regardless of how many jobs the officer leaves unranked. This signal is essentially costless because officers can change which jobs they list in their top 10 percent of choices at any time, and only preferences submitted at the time the marketplace closes are relevant to eventual matches. On the other side of the market, officers observe a signal if a unit ranks them anywhere on their rank-ordered list for a job in the officer's market. Officers and units are permitted to conduct informal interviews and to communicate outside of the online marketplace, but there is no strict requirement to do so.[9] Relatedly, officers can submit preferences over all jobs in their market regardless of whether they have interviewed with any of the units (and vice-versa).

Career managers at HRC are responsible for clearing markets by matching officers to

[7] In what follows, we often use the terms "a job's preferences over officers" and "a unit's preferences over officers" interchangeably.

[8] As a practical matter, there is little variation in wages for jobs within the same market as military base pay is a function of an officer's military rank and years of service. Officers receive a housing allowance that varies according to local housing prices near the base an officer is assigned. Officers assigned to bases with high costs of living may also receive an additional cost of living allowance.

[9] The Army's online platform does not have a functionality that allows officers and units to request and schedule interviews. As such, we are unable to observe which officers interviewed with which units.

jobs within distinct markets. These managers are officers who serve two to three years at HRC, before typically returning to a non-HRC position within their normal military occupation. After career managers clear a distinct market, they place officers on orders to move to their assigned units in the coming months. Between 6 and 9 months after the marketplace closes, officers report to a new unit. Depending on the timing of their contracts, officers can decide not to renew and exit the Army if they are unhappy with the match. In particular, some may exit before they begin their next assignment.[10]

3 Experimental Design

Drawing on prior research on the impacts of the deferred acceptance algorithm on match outcomes (Davis, 2022) and personnel economics within the context of the military (Greenberg et al., 2022; Bruhn et al., 2023), the research team suggested that the Army test the impact of matching using DA with a randomized controlled trial. The Army, however, implemented the matching mechanisms and maintained final decision-making authority over all aspects of the matching.

Our experiment took place during the officer marketplace open from October 11th through December 6th, 2019. This marketplace included more than 14,000 officers scheduled to move in the summer of 2020. The Army's practice of segmenting officers and units into disjoint markets defined by rank and specialty offers an ideal setting for randomizing at the market level. Furthermore, the matching mechanisms we describe below are implemented at the market level, making a disjoint market of officers and positions the appropriate unit of analysis for this study.

Our experimental sample includes 9,577 officers assigned to 115 distinct markets.[11] Be-

[10]Officers who move to a different base incur a one-year service obligation (United States Army, 2019). This obligation rarely binds for our primary retention results that follow, which are measured at 15 and 27 months after officers are scheduled to move.

[11]Our pre-analysis plan originally indicated 118 distinct markets were part of the experiment. However, prior to the listing of the marketplace, HRC made the decision not to execute two markets during the assignment cycle. HRC originally intended for 5 officers to be in one of these markets and for 1 officer to be

fore the marketplace opened, we worked with the Army to randomly assign these disjoint markets to either a treatment or control condition. Randomization was stratified by the rank of officers in the market and "skill clusters." The Army decided on skill clusters so as to group markets with similar skill requirements. For example, infantry and armor officers comprise one skill cluster, and officers with occupations related to logistics, finance, and acquisitions comprise another skill cluster. Skill clusters included anywhere from 2 to 25 markets. Strata defined by rank and skill cluster included between 2 and 10 markets.

3.1 Control Markets

Career managers matched officers in control group markets to jobs according to the Army's traditional matching process, which was neither automated nor reliant on a specific algorithm. Under this process, managers were responsible for pairing officers to jobs with an emphasis on officer and unit preferences, but also ensuring officers with unique assignment considerations—such as those with exceptional family considerations or spouses also in the Army—paired with jobs at locations that accommodated their needs. For the marketplace that took place during our experiment, the Army further instructed assignment officers to attempt to honor first-to-first pairings in control group markets—i.e. jobs where the officer ranked the job number 1 and where the unit ranked the officer number 1—consistent with how career managers traditionally matched officers to jobs since these are relatively easy to observe.[12]

in the other market. A third market included zero officers. Our sample does not include all 14,000 officers because, prior to the randomization, the Army decided to exclude roughly 4,000 officers in specialty occupations (medical service professionals, lawyers, chaplains, and some cyber and aviation officers with specific qualifications) from the experiment. The Army further excluded roughly 400 officers in the rank of first lieutenant who were part of special markets for officers scheduled to move outside of the Army's normal cycle. Because these exclusions were based on occupation and rank, these officers were not competing in markets with officers included in the experiment.

[12]We often use the term "first-to-first pairing" and "first-to-first match" synonymously. However, career managers were not obligated to honor all first-to-first pairings that existed at the close of the marketplace. As such, some job listings where an officer ranked the listing as their most preferred choice, and where a unit likewise ranked the same officer as their most preferred choice, did not in fact result in a first-to-first match.

Although each manager had leeway to pair officers to jobs according to their own process within these guidelines, our conversations with managers of control group markets suggest that most typically focused first on pairing officers with unique assignment considerations to jobs. For example, managers try to coordinate the placement of officers married to servicemembers in other marketplaces. Typically around 10 percent of officers in a market have a unique assignment consideration. Then career managers moved on to implementing first-to-first pairings or other scenarios where officers and units had mutually high ratings for each other. Finally, career managers matched remaining officers to jobs according to a process of the manager's choosing. While this process is somewhat of a black box, it is similar to human resources-driven assignment at other large organizations.[13]

3.2 Deferred Acceptance Markets

For officers in treatment markets, career managers first matched officers with unique assignment considerations to jobs, then career managers executed an officer-proposing deferred acceptance (DA) algorithm to match all remaining officers to remaining jobs in the market. In an officer-proposing algorithm, officers first "apply" to their top job choice.[14] All officers who are the highest-ranked, from the perspective of the job/unit, are placed on hold in their first choice. The other officers are "rejected." In the next round, "rejected" officers apply to their next most preferred job. Each job then "holds" the highest-ranked current applicant, either on hold from the previous round or newly applying to the job/unit in the current round. Officers not put on hold or who are removed from being on hold are rejected. The process continues until all officers are either on hold at a job or are rejected by all jobs at which point all "held" matches are finalized. Officer-proposing

[13]For example, the World Bank's staff policy manual states "Staff Members in positions at grades GF-GH whose professional disciplines are utilized in more than one department may be subject to planned periodic reassignment" but does not explain the assignment procedure (World Bank, 2019).

[14]The algorithm is run on a computer using submitted preferences. We describe officers "proposing" and being accepted or rejected at different steps in the algorithm for ease of exposition.

DA yields the officer-optimal stable match (Gale and Shapley, 1962), which means that all members of the proposing side prefer the DA stable match to all other possible stable matches. Importantly, managers of treatment markets reviewed all potential matches to ensure no officers matched to jobs for which they were not qualified. This review resulted in the adjustment of roughly 5 to 10 percent of officers in treatment markets.

By design, every market in both the treatment and control groups had at least as many job listings as officers. Before matching officers to jobs in both treatment and control group markets, HRC reduced the number of job listings in the marketplace to match the number of officers. Moreover, for markets in the treatment group, HRC imputed missing preferences for any jobs that officers left unranked and for any officers that units left unranked.[15] As a result, it was not possible for an officer to go "unmatched" due to a surplus of applicants or a lack of demand for officers. HRC's choice of which jobs to fill and which jobs to leave vacant were functions of the baseline vacancy rates across units, which were not directly influenced by officer and unit preferences.

4 Empirical Methods and Data

4.1 Estimation and Inference

For outcomes measured at the officer level, we estimate treatment effects using the following officer-level regression:

$$Y_i = \alpha + \beta DA_{m(i)} + X_i'\gamma + \delta_{b(m(i))} + \varepsilon_i, \tag{1}$$

where i indexes individual officers and $m(i)$ indicates officer i's market (determined by the

[15] In both cases, missing preferences were either randomly imputed or determined by the Army's preferences over jobs (or a combination of randomization and Army preferences), with the requirement that initially unranked positions or officers be ranked as less preferred than ranked positions or officers. We do not observe the imputed preferences.

officer's rank and occupation). Y_i is the outcome of interest and $DA_{m(i)}$ is an indicator for whether the market was randomly assigned to use deferred acceptance or the status quo matching mechanism. The coefficient β is the causal impact of being in a market randomly assigned to match officers to jobs using DA instead of the status quo manager-driven approach, or the intent-to-treat effect of the experiment. X_i is a set of pre-randomization officer characteristics including indicators for gender, race and ethnicity, family structure, birth year, source of commission, baseline performance, and years in rank. These controls are not necessary for identification because treatment was randomly assigned, but are included to help improve our statistical power. $\delta_{b(m(i))}$ represents a set of strata or block fixed effects, which account for any incidental differences in treatment probabilities across strata. When an outcome is measured at the unit level, we estimate an analogous job-level regression but without officer-specific controls.

Our inference is based on standard errors clustered by market because treatment status is randomly assigned at the market level (Abadie et al., 2023). Table B.1 shows versions of our inference using the wild bootstrap (Cameron et al., 2008) or adjustments for multiple hypothesis testing (Anderson, 2008).

4.1.1 Administrative Marketplace Data

Our data include the rank and occupation of each officer and job in the marketplace, allowing us to reconstruct each officer's full choice set of jobs and each job's full choice set of officers. We observe each officer's preferences over all jobs in their respective market and each job's preferences over all officers in the job's market. The ability to reconstruct choice sets allows us to observe when an officer leaves a job unranked and when a job leaves an officer unranked. We also observe the specific mapping of jobs to units.

4.1.2 Officer Surveys

We link administrative officer data with responses to two surveys. The first was a mid-market survey that HRC administered before the marketplace closed. During the final three weeks of the marketplace, while officers and units were still eligible to update their preferences, HRC required all officers who logged into the marketplace to complete this survey. Officers were not permitted to view or change their preferences for jobs until they completed the survey, resulting in a high response rate (88 percent). The survey inquired about the truthfulness of officers' preferences over jobs and about officers' perceptions in the marketplace. The second survey was a shorter post-market survey administered by HRC when officers learned of the job with which they had matched, typically 2 or 3 months after the conclusion of the marketplace.[16] Importantly, one question in this post-market survey inquired about the truthfulness of officers' preferences. Table B.2 contains the precise wording of each survey question reported on in our analysis.

4.1.3 Army Service, Evaluation, and Promotion Data

Our data links all officers in treatment and control markets to administrative service records that include the age, race, sex, marital status, source of commission (e.g. the Reserve Officer Training Corps [ROTC], West Point, etc.), rank, and Army occupation. We also have data on the performance ranking of each officer, determined in the month prior to the start of the marketplace. Army service records indicate the specific location and unit an officer is assigned to in a specific month and also allow us to observe if an officer is still on active duty, which is critical for constructing the retention outcomes that follow.

We also link officers in our sample to their evaluation reports through September 2022. Evaluation reports are the most important factor in determining whether an officer is pro-

[16]Officers learned the job with which they were matched just before completing the post-market survey, but could not obtain their orders until they completed the survey. However, only 54 percent of officers completed the post-market survey due to a technical delay in the survey prompt. Responses to both officer surveys are balanced across DA and control group markets.

moted to higher ranks. We define strong performance as receiving an evaluation report with a rating of "Most Qualified." A senior rater cannot give a rating of "Most Qualified" to more than 49 percent of the officers they evaluate. Officers are required to receive at least one evaluation report every 12 months, but can be evaluated sooner than 12 months if their rater (i.e., direct supervisor) or senior rater (i.e., the supervisor of their direct supervisor) changes, which occurs whenever an officer changes jobs.

Additional measures of officer performance include an indicator for whether an officer was promoted post-randomization and, for officers who were considered for promotion between October 2020 and September 2022, their promotion board percentile rankings relative to other officers considered for promotion at the same board. Officers are considered for promotion to the next higher rank once every five or six years, which explains why percentile rankings are missing for two-thirds of our sample.[17]

4.2 Baseline Summary Statistics and Balance Tests

Table 1 presents summary statistics of baseline officer and market characteristics separately for the treatment group (column 1) and the control group (column 2). Consistent with the demographic composition of all Army officers, officers in both treatment and control groups are predominately male, married, and born in 1982 or later (with an average age of 36 at the start of the market). Roughly 15 percent of officers are Black, 10 percent are Hispanic/Latino, and 66 percent are White. The average market had around 80 to 85 officers and between 110 to 120 jobs.

To formally test for baseline differences in covariate means between treatment and control groups, column (3) of Table 1 reports estimates from a regression of the baseline co-

[17]Technically, each cohort of officers has three opportunities to be promoted to the next rank: a below the zone opportunity, a primary zone opportunity one year later, and an above the zone opportunity one year later for anyone not selected in the primary zone. Since most officers are promoted in the primary zone, our percentile ranking reflects an officer's ranking from their primary zone promotion board. For the handful of officers who are promoted to the next rank below the zone, we adjust their percentile ranking so that they rank higher than all other officers in their original cohort who were considered in the primary zone.

variate in the left column on an indicator for whether the officer (or market) was part of the treatment group and fixed effects for market strata (defined by rank and skill cluster). Columns (4) and (5) report the standard error and p-value from the same regressions, respectively. Among the 23 comparisons reported in column (3), two are statistically significant at the 10 percent level and one of these is significant at the 5 percent level, consistent with what we would expect from random chance. Joint tests of significance among the officer characteristics and among the market characteristics fail to reject the null hypothesis that the treatment and control groups are balanced.[18] We additionally control for these baseline characteristics in equation 1, as was specified in our pre-analysis plan.

5 Immediate Impacts on Matching Methods and Matches

5.1 Impacts on Matching Mechanisms

Before presenting the effects of DA assignment, we first document compliance with random assignment in treatment and control markets. Table 2, Panel A shows the impact of a market being randomly assigned to match with DA on the probability that any non-first-to-first matches within a market were determined by DA.[19] We exclude first-to-first matches even though DA will always match a pair that ranks each other first because they were also prioritized by career managers in the control group. Markets randomly assigned to match with DA are 76.4 pp more likely to match with DA than control group markets. The corresponding F-statistic is 123.33, and, thus, random assignment is a highly relevant instrument for actually matching with DA at the market level.

The fact that treatment markets are not 100 pp more likely to match with DA implies there was some non-compliance with random assignment. In the treatment group, 53 of 59

[18]The joint test uses versions of covariates with missing values imputed as zero, along with indicators for missingness included as additional controls. This version of controls is also included in our main regressions.

[19]No markets in our sample had exclusively first-to-first matches.

markets had any non-first-to-first DA matches. In the control group, 8 of 56 markets had at least one non first-to-first DA match.[20] If we regress an indicator for whether each of the 115 markets had any non first-to-first DA matches on treatment status and strata fixed effects, the coefficient on treatment is 0.76 with a standard error of 0.07. This corresponds to a first-stage F-statistic of 123.33. Therefore, random assignment is a highly relevant instrument for whether DA was actually used to determine matches.

Table 2, Panel B shows the impact of being in a market randomly assigned to match with DA on officers' match types. Nearly 46 percent of matches in both treatment and control markets are first-to-first matches. Although DA will always match an officer with a job if they both rank each other first, the Army separately tracked first-to-first matches because it asked career managers in control markets to attempt to honor such matches (see Section 3.1). For non-first-to-first matches, we observe a flag for whether matches were directly determined by DA. As expected, matches in DA markets are 27.1 pp more likely to have a non-first-to-first match determined by DA than officers in control markets, which only have a DA match about 1 percent of the time. This difference is significantly different from zero ($p < 0.001$). Because the process in control markets might still result in the same match as would have occurred under DA, our measure is best interpreted as a measure of manager compliance, as opposed to deviation from a DA counterfactual.

5.2 Impacts on Justified Envy

A theoretical benefit of DA is that it produces a stable match. A match is stable if no officer and job prefer to be matched together over their assigned match. When this is not the case, we say that the officer has "justified envy." The first row of Table 2, Panel B shows 9.6 percent of officers in control markets have justified envy for at least one position. Matching with DA reduces the prevalence of justified envy by 3.2 pp. This proportionally

[20]The 8 control group markets with at least one non first-to-first DA match all had overall DA match rates, including first-to-first matches, of over 50 percent. A total of 60 officers in these markets (out of 4,776 in the control group) had a non first-to-first DA match.

large (one-third) and statistically significant ($p < 0.001$) reduction confirms that random assignment to matching with DA caused a material change in matches. Relatedly, DA reduced the average number of jobs for which an officer has justified envy by 0.053, from a baseline average of 0.139 in the control group. This 38 percent reduction is statistically significant at conventional levels ($p < 0.001$).

Of course, if all matches were determined by DA, we would expect all justified envy to be eliminated. Non-compliance with random assignment at the market level (as described in Section 5.1) in part explains why justified envy was not entirely eliminated. Scaling by the "first stage" of 0.76 (i.e. the treatment's effect on the likelihood that a market has at least one non first-to-first DA match), suggests that DA reduced justified envy by approximately 50 percent.[21] Non-compliance with DA-recommended assignments in treated markets explains the remaining instances of justified envy.

The Army, and HRC in particular, had the power to overrule the DA assignment and place an officer in a different match if deemed necessary. These changes could not have been driven by officers lobbying for different placements, because officers do not see their assignments until the finalized list is released. Instead, deviations from DA are driven by officers being moved to meet certain constraints—such as ensuring only qualified officers are assigned to positions that involve training other officers—or to satisfy other Army objectives. Although the remaining 8 to 9 percent of jobs in the treatment group where officers have justified envy may be interpreted as a significant deviation from the algorithm, it is important to note that even a single changed match can create justified envy for many officers if the position was desirable and the moved officer was not ranked highly by the new assignment. Such deviations from official DA assignments are not unique to our setting. For example, there are waivers in the NRMP[22] and some school choice markets reserve slots for principals.[23]

[21] Control complier markets have similar rates of justified envy, on average, as control markets.
[22] https://www.nrmp.org/policy/requesting-a-waiver/
[23] https://chicagoschooloptions.com/forums/topic/spring-2023-sehs-principals-discretion/

5.3 Impacts on Match Rank and Match Satisfaction

The officer-proposing version of DA used in our study yields the officer-optimal stable match. Therefore, we might expect officers and units in DA markets to be happier with their matches. As a first test of this hypothesis, we measure the impact of DA matching on the average ranking of matches. Table 3 shows the impact of DA on officers' and units' satisfaction with their assigned match based on their stated preferences and officers' responses to the post-market survey.[24]

Panel A shows that about 85 percent of officers in both DA and control group markets were matched with a unit that they ranked. Among this subset of officers, officers in control markets ranked their match 7.9, on average, and officers in DA markets ranked their match 6.6, on average, for a difference of 1.3 ranks. This difference is significant at conventional levels ($p = 0.039$). Table B.1 shows that this significance is sensitive to using the wild bootstrap or making adjustments for multiple hypothesis testing. To the extent that rankings reflect true preferences, officers in DA markets matched to slightly more preferred jobs. However, truthful reporting is only a dominant strategy within DA markets. Nonetheless, we can interpret a ranked match that is more preferred as an outcome closer to an officer or unit's strategic goal. We return to the question of separating strategic behavior from truthful reporting in Section 7.

Panel B reports the impact on officers' answers to three questions from the post-market survey inquiring about their satisfaction. Responses to these questions were reported on a 5-point Likert scale. We standardize responses using the control group mean and standard deviation. We find positive, but statistically insignificant and economically modest treatment effects on officers' responses to questions about how satisfied they are with the match they received, how satisfied they are with the marketplace overall, and how likely they are to stay in the Army.

Panel C shows the impact of DA on *units'* stated preferences over matches. In general,

[24]Three officers did not submit any preferences and are excluded from analyses using preference reports.

units may benefit from misreporting their preferences, even when matches are determined by DA, in an officer-proposing design. These estimates will therefore be less likely to reflect the impact of DA on units' satisfaction with the outcomes. In both DA and control group markets, about 70 percent of units match to an officer they ranked. Units in DA markets prefer their matches by 0.4 ranks. This is statistically significant ($p = 0.014$) and proportionally large improvement given that the average unit in the control group ranked their match 2.7. The significance of this result is sensitive to some adjustments for multiple hypothesis testing (see Table B.1).

6 Longer-Run Impacts

In this section, we move beyond the initial features of the matches and document the impact of matching with DA on officers' retention, performance, and promotions through their first two years in their new position.

6.1 Officer Retention

Table 4, Panel A reports the effect of being assigned to match with DA on retention. The experimental marketplace closed in December 2019 and officers began receiving orders for their next assignment starting in February 2020, with instructions to report to follow-on assignments in the summer of 2020. The onset of the COVID-19 pandemic in the United States in Spring 2020 delayed some moves, but nearly all officers moved by the end of September 2020. Our three primary retention outcomes are (1) an indicator for still being in the active duty Army as of September 30th, 2020, which measures any attrition that might occur after officers learn the results of the marketplace; (2) an indicator for still being in the Army as of September 30th, 2021, which we broadly interpret as the primary "first-year" retention outcome; and (3) an indicator for still being in the Army as of September

30th 2022, which we interpret as "second-year" retention.[25]

The first row of Panel A indicates that relative to officers in control markets, officers in treatment markets that matched with the DA algorithm were a statistically insignificant 0.3 pp more likely to be in the Army through September 2020. The second row reveals that DA increases retention through one year (September 2021) by a statistically significant 1.1 pp (p = 0.025). This significance is sensitive to adjusting for the familywise error rate (p=0.15), but is marginally significant after adjusting for the false discovery rate (q=0.08). One-year retention in the control group is high (93.4 percent), and our treatment effect therefore implies DA reduces attrition by 16.7 percent $\left(\frac{1.1}{100-93.4} = 0.167\right)$.

However, these effects fade-out by the second year. Matching with DA increases retention through the second year (September 2022) by only 0.3 pp. The 95 percent confidence interval around this estimate rules out increases in retention larger than 1.5 pp or reductions in retention of more than 0.9 pp.

6.2 Performance Outcomes

Panel B of Table 4 reports impacts on performance outcomes in officers' first year in their new match (October 2020 through September 2021). As described in Section 4.1.3, officers receive evaluation reports whenever their supervisor changes or after serving under the same supervisor for a total of 12 months, whichever comes first. Most officers moved in the summer of 2020 and should therefore have received at least one evaluation report from their new position by September 2021. To ensure that evaluations are predominantly from officers' new assignments, we only construct performance outcomes from evaluations with rating periods that ended in October 2020 or later.

In the first row of Table 4, Panel B, we find that officers in DA markets were slightly more likely to have received a performance evaluation with rating periods ending be-

[25]Our pre-analysis plan said we would measure retention through July of each year. We extended the window to September because of the COVID-19 delays.

tween October 2020 and September 2021, consistent with the positive effects we observed on one-year retention in Panel A. The second row of Panel B suggests that, according to observed performance evaluations, matches resulting from DA did not lead to improved performance in the subsequent assignment relative to matches in control markets. Officers in treated markets were 0.6 pp less likely to receive at least one good, or "Most Qualified," evaluation than officers in control markets (control mean of 48.3 percent). This estimate is indistinguishable from 0, with the 95 percent confidence interval ranging from a 2.2 pp reduction to a 1.0 pp increase. Of course, this result conditions on having at least one evaluation with the rating period ending between October 2020 and September 2021, which is imbalanced between treatment and control. In Table B.3, we show that this finding is not sensitive to how we treat missing performance evaluations. We find no effect on performance whether we impute missing performance evaluations with the market average or all negative or all positive outcomes.

Panel C shows the impact on evaluations in officers' second year in the position (October 2021 through September 2022). As of April 2023, 80.6 percent of officers in the control group had an evaluation during this time period.[26] We again find precise zero effects on the impact of matching with DA on officers' performance evaluations in the second year after an officer moves. Panel D reports the impact of matching with DA on officers' promotion outcomes. About 30 percent of officers in control group markets had been promoted in their first two years in the match. Mirroring the performance evaluation results, our estimated impact of DA on promotions is a precise zero.

We dig deeper into the promotion outcome by looking at officers' percentile ranks by their official promotion boards. Promotion boards consist of a panel of senior officers who review and rank-order the personnel files of all officers within a particular cohort to determine which officers from the cohort will be promoted to the next rank. We only observe this outcome for 32 percent of our sample, primarily because officers are typically only considered for promotion once every five or six years. However, there is no difference

[26] Evaluations appear in our administrative data with a lag.

in the rate at which we observe promotion board outcomes between treatment and control markets. Among officers where we can observe promotion board outcomes, matching with DA increased an officer's promotion board ranking by less than 1 percentile. This point estimate is statistically insignificant, and we can rule out an increase of 2 percentiles and a decrease of 1 percentile. This result is unchanged if we include a fixed effect for the specific promotion board that evaluated the officer.

Taken together, these results suggest that matching with DA had at most a relatively small impact on match quality. Even so, matching with DA might still be optimal from an organization's perspective because it is straightforward to implement and easy for workers and managers to navigate.

7 Strategic Behavior in Matching Markets

One explanation for the modest impact of DA on these outcomes is that officers may have deviated from truthful preference reporting for strategic reasons. Officers should not be able to benefit from strategically misreporting their preferences because DA is strategyproof (Pathak and Sönmez, 2008). However, the theoretical properties of DA rely on assumptions that may not hold in practice. Participants in DA markets have been shown to misreport their preferences in a variety of settings. For example, using experimental evidence collected outside of the NRMP from a small share of the market, Rees-Jones (2018) finds that many doctors who recently participated in the NRMP try to strategically manipulate their preferences even though the matching mechanism is strategyproof.

A benefit of our setting is that we can match officers to their self-reported survey responses, to their preferences for units, and to units' preferences for them. Moreover, the response rates to the surveys are also quite high (88 percent for the mid-market survey and 54 percent in the post-market survey). We explore the possibility of strategic behavior by first documenting the impact of DA on truthful preference reporting based on officer sur-

veys. Next, we document patterns in preference reports that are consistent with strategic preference coordination between officers and units. Finally, we confirm that the patterns persist in subsequent years, despite improved messaging of DA's properties.[27]

7.1 Evidence from Officer Surveys

Table 5 reports results from survey questions administered to officers as part of the mid-market survey and the post-market survey (both described in Section 4.1.2) that pertain to strategic behavior. The mid-market survey was administered to officers when they were actually participating in the matching market. A high share of officers (87 percent in control markets, 88 percent in treatment markets) responded to this survey. The mid-market survey asked about strategic behavior in three ways. The impact of DA on responses to these questions is shown in Table 5, Panel A.

First, officers were asked if they were guaranteed to match with their top choice, would their stated top choice remain their top choice? Matching with DA increased the rate at which officers indicate truthful reporting of their first choice by 1.2 pp. This effect is marginally statistically significant ($p = 0.061$) but is small relative to the control group's 84 percent rate of truthfully reporting the most preferred choice and is not robust to adjustment for multiple hypothesis testing. Moreover, officers were asked this question again in the post-market survey, which was administered when officers were notified about their matches. As seen in Panel B of Table 5, officers in DA markets were no more likely to indicate that they truthfully reported their first choice than officers in control markets.[28]

The mid-market survey also asked officers about the extent to which their reported preferences reflected their true preferences beyond their top choice. Responses were on a Likert scale, which we standardized using the control group mean and standard deviation

[27] In Appendix A, we show, using a simple theoretical example, that officers may benefit from communicating their preferences and coordinating first-to-first matches even when matches are determined with DA.

[28] The mid-market and end-of-market estimates of DA's effect on truthful first choices are not statistically distinguishable, so some or all of this difference could simply reflect noise.

(SD). As a result, the control group mean is zero by construction. We find a statistically significant ($p < 0.01$) 0.05 SD increase in the extent of truthful reporting in DA markets. This is driven entirely by a statistically significant 2.4 pp increase in the share of officers stating that their reported preferences always reflect their true preferences in the treatment group. This effect constitutes a 10 percent increase in completely honest reporting, relative to the control group's 24 percent rate. While, in our view, this is a proportionally large increase, nearly 3 in 4 officers in DA markets still did not report always honestly reporting.

Finally, the mid-market survey asked officers whether they ranked any positions higher because they were aware that ranking a job among their top 10 percent of possible choices sent a signal to units. Matching with DA reduced this signaling by 2.6 pp. While this effect is statistically significant ($p = 0.01$), more than half of officers in DA markets still responded affirmatively that they had strategically misreported their preferences to send this signal. In the next two subsections, we further explore the evidence of and consequences of strategic communication.

7.2 Evidence from Officer and Unit Preferences

Some communication across market sides is usually necessary in two-sided markets. Workers and jobs need to gather information about how much they may or may not like potential matches. Recent research has shown that idealized properties of the NRMP may not be realized because hospitals and doctors generally have to interview each other before learning their preferences (Echenique et al., 2022). While jobs are often prohibited from asking workers directly about their preferences, as was true in our setting and in the NRMP, workers may still find it beneficial to signal their interest in certain jobs.[29] And both doctors and residency program directors report frequent communication about rankings after interviews even though this type of communication is prohibited by the

[29]See for example: https://blog.matcharesident.com/residency-programs-number-one-choice/.

NRMP's code of conduct (Anderson et al., 1999; Carek et al., 2000; Teichman et al., 2000; Carek, 2012). As we have mentioned, the marketplace platform informs units if an officer ranks one of their jobs among their top 10 percent of choices, which the officer can change at any time. But officers and units may try to gain an advantage by coordinating outside of the official mechanism. Roth and Xing (1994) refer to this type of coordination within a centralized marketplace as "stage 4 unraveling."

A straightforward way for officers and units to coordinate is to agree to rank each other first. This guarantees a match when DA is used and likely guarantees a match even when DA is not used, given the Army's commitment to honoring first-to-first matches. A high rate of first-to-first matches, however, is not necessarily evidence of this coordination. It could reflect a high degree of correlation in officers' and units' true preferences for each other. In order to overcome this potential confound, we focus our attention on units with multiple jobs. The matching market is many-to-one, and thus, many units are looking to hire multiple officers. Often, units are looking to hire multiple officers for nearly identical jobs.[30] Within a set of nearly identical jobs at the same unit, correlation in officers' and units' true preferences is held constant.

We further restrict attention to officers and groups of similar jobs with potential first-to-first matches—that is, at least one of the identical jobs at a unit ranked the officer first and the officer ranked at least one of the identical jobs first. Because identical jobs are perfect substitutes from the officer's perspective, we would expect the officer's top choice to be uniformly distributed over these identical positions. An abnormally high incidence of the officer choosing as a top choice the particular job listing that ranked them first is evidence of coordination.

We test the null hypothesis that officers' top choices are uniformly distributed over positions using randomization tests separately in treatment and control markets. Holding

[30]We say a pair of jobs is nearly identical if the Jaro-Winkler string similarity score between their job descriptions is greater than 0.9 (with 1.0 indicating the descriptions are identical). We require job similarity to be transitive so if job A and B are nearly identical and jobs B and C are nearly identical, we say A and C are nearly identical even if the similarity score is lower than 0.9.

unit preferences fixed, we randomize which of the identical jobs the officer ranks first 10,000 times. We calculate the share of first-to-first pairings in each randomization and then calculate a p-value using the share of randomizations with a first-to-first matching rate at least as large as the observed rate. Figure 1 shows the results. In both DA (right panel) and control markets (left panel), the observed share of first-to-first matches is more than double what we would expect if officers were randomizing over identical jobs. The observed rates are 48 pp larger than even the largest first-to-first rate in the simulations.[31] This provides strong evidence against the null hypothesis that officers randomize over identical jobs and suggests officers and units coordinated which jobs to rank first.

7.3 Evidence from Post-Experiment DA Markets

The high first-to-first match rate and evidence of coordination during our experimental market could potentially be due to officers in DA markets not fully understanding that truth-telling is an optimal strategy, a common problem in other settings that use DA (Chen and Sönmez, 2002; Hassidim et al., 2017; Rees-Jones, 2018; Rees-Jones and Skowronek, 2018). We are not able to observe exactly how the DA mechanism was explained to officers during our experiment and were not permitted to add questions measuring understanding to officer surveys. However, we believe officers and units would have strategically coordinated rankings during our experiment even if most officers understood that truth-telling is an optimal strategy in DA markets. Even though the Army did not adopt standardized messaging that differed for officers in treatment and control markets, career managers knew if their market was part of the treatment or control group and were permitted to communicate their market's matching process to officers and units. As discussed earlier in this section, the statistically significant, albeit economically small, increase in truthful preference reporting as indicated through the mid-market officer survey

[31] The actual share of first-to-first matches in Figure 1, just over 90 percent, exceeds the overall rate of first-to-first matches in our sample (roughly 45 percent, see Table 2), because the randomization tests restrict to officer-by-job pairs among identical jobs where the unit ranks the officer as their top choice for at least one of the identical jobs and where the officer ranks one of the identical jobs as their top choice.

is consistent with officers in treatment markets having some understanding that DA is strategyproof.

For additional evidence, we turn to markets that occurred one year after (October - December 2020) and two years after (October - December 2021) the randomized trial. The Army adopted DA for all markets at the conclusion of the randomized trial. Officers in more recent markets have had more time to learn the implications of DA from publicly available sources (e.g., Greenberg et al. (2020)) and from career managers.[32] Additionally, prior to the 2021 marketplace, the Army adopted a user agreement modeled after the NRMP's Match Code of Conduct in part to address concerns that units and officers were not respecting the confidentiality of preferences. All officers and units were required to acknowledge the user agreement, shown in Figure B.1, the first time they entered the marketplace. Similar to how the NRMP Match Code of Conduct asks program directors to not request an applicant disclose ranking preferences or intentions, the Army's user agreement asks officers and units to not ask the other side to disclose their preferences in any way.[33]

Despite these efforts, the incidence of first-to-first matching actually increased in more recent markets, as seen in Figure B.2. Furthermore, relative to officers in the experimental marketplace, officers in the marketplaces that occurred one and two years later were more likely to acknowledge that their highest-ranked job was not their true first preference.[34]

Survey responses from officers participating in marketplaces one and two years after the randomized trial also reveal that a majority of officers acknowledged altering their

[32]For example, before the October - December 2021 marketplace, one career manager sent an e-mail to moving officers with the following message: *"Finally, when you make your final adjustments to your preferences when the market closes, be sure to put down your TRUE preferences…don't be afraid to put a job #1 even if you're not sure you will get it."*

[33]The NRMP Match Code of Conduct for Programs is available at `https://www.nrmp.org/wp-content/uploads/2022/08/NRMP-Match-Code-of-Conduct_Programs_Final.pdf` (accessed 28 July 2023).

[34]Among officers in treated markets of the randomized trial who responded to the post-market survey, 30 percent acknowledged their first choice job was not their true first preference (see Panel B, Table 5.). Among officers in markets corresponding to treatment group markets (according to officers' rank and occupation) that took place one year later and two years later, 46 percent and 38 percent, respectively, acknowledged that their first choice job was not their true top preference (See Tables B.4 and B.5).

preferences in order to achieve a first-to-first match (see Tables B.4 and B.5).[35] Officers who altered their preferences to achieve a first-to-first match were also more likely to acknowledge that their stated first choice assignment was not their true first choice (Tables B.6 and B.7), suggesting that strategic preference coordination is not only the result of officers naturally preferring to be assigned to units that likewise rank them highly. Overall, strategic preference reporting appears to have increased in the years following the randomized trial, and results from officer surveys suggest that much of this was driven by agents' desire to strategically coordinate first-to-first matches.

8 Conclusion

This paper presents results from a randomized controlled trial of the impact of matching workers to jobs using the deferred acceptance algorithm. Our setting is matching US Army officers to units. We randomized 115 disjoint markets with nearly 10,000 officers and hundreds of units to treatment and control conditions. This is a large enough sample to yield precisely estimated treatment effects. We analyze the impacts using high-quality administrative data on both participants' preferences and relevant market outcomes. We are able to link these data to surveys that ask participants about strategic preference manipulations and their satisfaction with their assigned matches. These features provide us with the unique opportunity to credibly measure the impact of DA on immediate and longer-run market outcomes. This is the first RCT to measure the impact of matching workers to jobs using DA that we are aware of.

Matching with DA reduced officers' attrition in their first year in their new match by 16.7 percent, but these retention benefits fade out by the second year. We find no evidence that matching with DA affected officers' performance in their new positions based on performance evaluations or promotions. We show that matching with DA significantly

[35] Unfortunately, the survey administered to officers in markets that were part of the randomized trial did not ask officers if they altered preferences for jobs for the purpose of achieving a first-to-first match.

reduces strategic preference manipulation but the reduction is relatively small and many officers misreport their preferences.

One explanation for the modest impact of DA on these outcomes is that officers may have deviated from truthful preference reporting for strategic reasons. We present new evidence of communication and coordination of preference reports across the two sides of the market. Workers may benefit from strategic coordination of first-to-first matches with a potential job if this coordination improves their ranking with the job. This type of endogenous preference formation may explain the low rates of truthful preference reporting and high rates of coordination we observe. Extensive preference communication has also been documented by surveys of a few hundred participants in the NRMP (Anderson et al., 1999; Carek et al., 2000; Teichman et al., 2000; Sbicca et al., 2010; Carek, 2012; Berriochoa et al., 2018). We expect this type of communication is common in most two-sided labor markets where some communication across sides is necessary. However, such communication is less likely in matching applications where communication across sides is either uncommon or unnecessary, such as school choice markets where schools' rankings of students are based on lotteries, test scores, or distance rules. Future research on the effects of DA in matching markets with limited communication across sides is clearly warranted.

Even with these modest impacts on retention and performance, DA matching might still be optimal from an organization's perspective because it is straightforward to implement and easy for workers and managers to navigate. Revealed preference would suggest this is the case: the Army chose to adopt DA for all markets in the years after this RCT.

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. (2023). When should you adjust standard errors for clustering? *Quarterly Journal of Economics*, 138(1):1–35.

Abdulkadiroğlu, A., Pathak, P. A., and Roth, A. E. (2009). Strategy-proofness versus effi-

ciency in matching with indifferences: Redesigning the nyc high school match. *American Economic Review*, 99(5):1954–78.

Abdulkadiroğlu, A., Pathak, P. A., Roth, A. E., and Sönmez, T. (2006). Changing the boston school choice mechanism. *NBER Working Paper*, 11965.

Abdulkadiroğlu, A. and Sönmez, T. (2003). School choice: A mechanism design approach. *American Economic Review*, 93(3):729–747.

Abdulkadiroğlu, A., Pathak, P. A., and Roth, A. E. (2005a). The New York City High School Match. *American Economic Review*, 95(2):364–367.

Abdulkadiroğlu, A., Pathak, P. A., Roth, A. E., and Sönmez, T. (2005b). The Boston Public School Match. *American Economic Review*, 95(2):368–371.

Anderson, K. D., Jacobs, D. M., and Blue, A. V. (1999). Is match ethics an oxymoron? *American Journal of Surgery*, 177(3):237–239.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American Statistical Association*, 103(484):1481–1495.

Angrist, J. D. (1990). Lifetime earnings and the vietnam era draft lottery: evidence from social security administrative records. *American Economic Review*, pages 313–336.

Angrist, J. D. and Chen, S. H. (2011). Schooling and the vietnam-era gi bill: Evidence from the draft lottery. *American Economic Journal: Applied Economics*, 3(2):96–118.

Angrist, J. D., Chen, S. H., and Frandsen, B. R. (2010). Did vietnam veterans get sicker in the 1990s? the complicated effects of military service on self-reported health. *Journal of Public Economics*, 94(11-12):824–837.

Angrist, J. D., Chen, S. H., and Song, J. (2011). Long-term consequences of vietnam-era conscription: New estimates using social security data. *American Economic Review*, 101(3):334–38.

Antler, Y. (2015). Two-sided matching with endogenous preferences. *American Economic Journal: Microeconomics*, 7(3):241–258.

Arteaga, F., Kapor, A. J., Neilson, C. A., and Zimmerman, S. D. (2022). Smart matching platforms and heterogeneous beliefs in centralized school choice. *Quarterly Journal of Economics*, 137(3):1791–1848.

Barron, G. and Vardy, F. (2005). The Internal Job Market of the IMF's Economist Program. *IMF Staff Papers*, 52(3):410–429.

Bedard, K. and Deschênes, O. (2006). The long-term impact of military service on health: Evidence from world war ii and korean war veterans. *American Economic Review*, 96(1):176–194.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300.

Benson, A., Li, D., and Shue, K. (2019). Promotions and the peter principle. *Quarterly Journal of Economics*, 134(4):2085–2134.

Berriochoa, C., Reddy, C. A., Dorsey, S., Campbell, S., Poblete-Lopez, C., Schlenk, R., Spencer, A., Lee, J., Eagleton, M., and Tendulkar, R. D. (2018). The residency match: Interview experiences, postinterview communication, and associated distress. *Journal of Graduate Medical Education*, 10(4):403–408.

Bingley, P., Lundborg, P., and Lyk-Jensen, S. V. (2020). The opportunity costs of mandatory military service: Evidence from a draft lottery. *Journal of Labor Economics*, 38(1):39–66.

Bound, J. and Turner, S. (2002). Going to war and going to college: Did world war ii and the gi bill increase educational attainment for returning veterans? *Journal of Labor Economics*, 20(4):784–815.

Bruhn, J. M., Greenberg, K., Gudgeon, M., Rose, E. K., and Shem-Tov, Y. (2023). The effects of combat deployments on veterans' outcomes. *NBER Working Paper*, 30622.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics*, 90(3):414–427.

Card, D. and Cardoso, A. R. (2012). Can compulsory military service raise civilian wages? evidence from the peacetime draft in portugal. *American Economic Journal: Applied Economics*, 4(4):57–93.

Carek, P. J. (2012). Postinterview communications between residency programs and candidates: what are the rules? *Journal of Graduate Medical Education*, 4(2):263–264.

Carek, P. J., Anderson, K. D., Blue, A. V., and Mavis, B. E. (2000). Recruitment behavior and program directors: How ethical are their perspective about the match process? *Family Medicine*, 32(4):258–260.

Charness, G. and Kuhn, P. (2011). Lab labor: What can labor economists learn from the lab? *Handbook of Labor Economics*, 4:229–330.

Chen, Y. and Sönmez, T. (2002). Improving efficiency of on-campus housing: An experimental study. *American Economic Review*, 92(5):1669–1686.

Chen, Y. and Sönmez, T. (2006). School choice: an experimental study. *Journal of Economic theory*, 127(1):202–231.

Cowgill, B., Davis, J., Montagnes, B. P., and Perkowski, P. (2022). Matchmaking principals: Theory and evidence from internal labor markets. *Available at SSRN*.

Cowgill, B. and Koning, R. (2018). Matching markets for googlers. Harvard Business School Case 718-487.

Davis, J. M. (2022). Labor market design can improve match outcomes: Evidence from matching teach for america teachers to schools. Working Paper.

Echenique, F., González, R., Wilson, A. J., and Yariv, L. (2022). Top of the batch: Interviews and the match. *American Economic Review: Insights*, 4(2):223–38.

Gale, D. and Shapley, L. S. . (1962). College Admissions and the Stability of Marriage. *American Mathematical Monthly*, 69(1):9–15.

Greenberg, K., Crow, M., and Wojtaszek, C. (2020). Winning in the marketplace: How officers and units can get the most out of the army talent alignment process. *Modern War Institute*.

Greenberg, K., Gudgeon, M., Isen, A., Miller, C., and Patterson, R. (2022). Army service in the all-volunteer era. *Quarterly Journal of Economics*, 137(4):2363–2418.

Greenberg, K., Pathak, P. A., and Sönmez, T. (2023). Redesigning the us army's branching progress: A case study in minimalist market design. Technical report.

Haegele, I. (2021). Talent hoarding in organizations.

Hassidim, A., Marciano, D., Romm, A., and Shorrer, R. I. (2017). The mechanism is truthful, why aren't you? *American Economic Review*, 107(5):220–24.

Horton, J. J., Rand, D. G., and Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 14:399–425.

Huitfeldt, I., Kostøl, A. R., Nimczik, J., and Weber, A. (2023). Internal labor markets: A worker flow approach. *Journal of Econometrics*, 233(2):661–688.

Jagadeesan, R. (2019). Cadet-branch matching in a kelso-crawford economy. *American Economic Journal: Microeconomics*, 11(3):191–224.

Jones, D., Molitor, D., and Reif, J. (2019). What do workplace wellness programs do? evidence from the illinois workplace wellness study. *Quarterly Journal of Economics*, 134(4):1747–1791.

Kagel, J. H. and Roth, A. E. (2000). The dynamics of reorganization in matching markets: A laboratory experiment motivated by a natural experiment. *Quarterly Journal of Economics*, 115(1):201–235.

Lewis, J. L., Davenport, A. C., Tannehill, B., Lewis, A., Marrone, J. V., Smith, V. M., and Bicksler, B. (2022). *Improving Precommissioning Assignments and Readiness on the U.S. Coast Guard Offshore Patrol Cutter*. RAND Corporation.

List, J. A. and Rasul, I. (2011). Field experiments in labor economics. In *Handbook of Labor Economics*, volume 4, pages 103–228. Elsevier.

Niederle, M. and Roth, A. E. (2003a). Relationship Between Wages and Presence of a Match in Medical Fellowships. *Journal of the American Medical Association*, 290(9):1153–1154.

Niederle, M. and Roth, A. E. (2003b). Unraveling Reduces Mobility in a Labor Market: Gastroenterology with and without a Centralized Match. *Journal of Political Economy*, 111(6):1342–1352.

Pathak, P. A. and Sönmez, T. (2008). Leveling the playing field: Sincere and sophisticated players in the Boston mechanism. *American Economic Review*, 98(4):1636–1652.

Rees-Jones, A. (2017). Mistaken play in the deferred acceptance algorithm: Implications for positive assortative matching. *American Economic Review*, 107(5):225–29.

Rees-Jones, A. (2018). Suboptimal behavior in strategy-proof mechanisms: Evidence from the residency match. *Games and Economic Behavior*, 108:317–330.

Rees-Jones, A. and Skowronek, S. (2018). An experimental investigation of preference misrepresentation in the residency match. *Proceedings of the National Academy of Sciences*, 115(45):11471–11476.

Roth, A. E. (1984). The evolution of the labor market for medical interns and residents: a case study in game theory. *Journal of Political Economy*, 92(6):991–1016.

Roth, A. E. (1990). New physicians: a natural experiment in market organization. *Science*, 250(4987):1524–1528.

Roth, A. E. (1991). A Natural Experiment in the Organization of Entry-Level Labor Markets : Regional Markets for New Physicians and Surgeons in the United Kingdom. *American Economic Review*, 81(3):415–440.

Roth, A. E. (2008). Deferred acceptance algorithms: history, theory, practice, and open questions. *International Journal of Game Theory*, 36(3-4):537–569.

Roth, A. E. and Peranson, E. (1999). The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design. *American Economic Review*, 89(4):748–780.

Roth, A. E., Sönmez, T., and Unver, M. U. (2003). Kidney exchange. *Quarterly Journal of Economics*, 119(2):457—-488.

Roth, A. E. and Xing, X. (1994). Jumping the gun: Imperfections and institutions related to the timing of market transactions. *American Economic Review*, 84(4):992–1044.

Sbicca, J. A., Gorell, E. S., Kanzler, M. H., and Lane, A. T. (2010). The integrity of the dermatology national resident matching program: results of a national study. *Journal of the American Academy of Dermatology*, 63(4):594–601.

Sbicca, J. A., Gorell, E. S., Peng, D. H., and Lane, A. T. (2012). A follow-up survey of the integrity of the dermatology national resident matching program. *Journal of the American Academy of Dermatology*, 67(3):429–435.

Schlegel, J. C. (2015). Contracts versus salaries in matching: A general result. *Journal of Economic Theory*, 159(A):552–573.

Sönmez, T. (2013). Bidding for Army Career Specialties: Improving the ROTC Branching Mechanism. *Journal of Political Economy*, 121(1):186–219.

Sönmez, T. and Switzer, T. B. (2013). Matching With (Branch-of-Choice) Contracts at the United States Military Academy. *Econometrica*, 81(2):451–488.

Teichman, Anderson, K. D., Dorough, M. M., Stein, C. R., Optenberg, S. A., and Thompson, I. M. (2000). The urology residency matching program in practice. *Journal of Urology*, 163(6):1878–1887.

United States Army (2019). Army regulation 614-100: Officer assignment policies, details, and transfers. Available at: https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/ARN16964_R614_100_FINAL.pdf, Last accessed: July 31, 2022.

Westfall, P. H., Young, S. S., and Wright, S. P. (1993). On adjusting p-values for multiplicity. *Biometrics*, 49(3):941–945.

World Bank (2019). Staff rule 5.01 - reassignment. Available at: https://ppfdocuments.azureedge.net/0b288045-d3f0-44d6-b4d8-7521a85f96de.pdf, Last accessed: August 18, 2023.

Table 1: Descriptive Statistics and Balance Tests

	Treatment Mean	Control Mean	Regression Adj. Difference	SE	*P*-value	*N*
Panel A: Officer Characteristics						
Female	0.149	0.133	0.001	0.017	0.944	9,577
White	0.650	0.675	-0.024	0.017	0.151	9,577
Black	0.160	0.139	0.026**	0.013	0.043	9,577
Hispanic	0.096	0.097	-0.007	0.006	0.292	9,577
Married	0.740	0.762	-0.002	0.010	0.828	9,577
Children	0.609	0.639	-0.004	0.014	0.776	9,577
Married with Children	0.561	0.593	-0.007	0.014	0.640	9,577
Birth year 1962-1966	0.008	0.00	0.004*	0.002	0.056	9,577
Birth year 1967-1971	0.038	0.040	0.006	0.005	0.233	9,577
Birth year 1972-1976	0.106	0.118	0.011	0.008	0.178	9,577
Birth year 1977-1981	0.219	0.254	-0.012	0.011	0.257	9,577
Birth year 1982-1986	0.287	0.294	-0.003	0.015	0.842	9,577
Birth year 1987-1991	0.213	0.182	-0.009	0.009	0.325	9,577
Birth year 1992-1996	0.127	0.105	0.003	0.008	0.664	9,577
ROTC	0.537	0.522	-0.013	0.011	0.246	9,577
USMA	0.138	0.142	0.001	0.014	0.931	9,577
Performance Quartile 1	0.235	0.199	0.008	0.012	0.505	8,759
Performance Quartile 2	0.243	0.249	-0.002	0.009	0.794	8,759
Performance Quartile 3	0.260	0.262	0.004	0.008	0.623	8,759
Performance Quartile 4	0.262	0.290	-0.009	0.014	0.485	8,759
More than 3 years in the current rank	0.498	0.503	0.002	0.014	0.900	9,565
Joint Test, Officer Characteristics					0.107	9,577
Panel B: Market Characteristics						
Number of Officers	81.373	85.286	-1.629	15.749	0.918	115
Number of Jobs	109.585	123.750	-9.150	21.319	0.669	115
Joint Test, Market Size					0.877	115

Notes: The sample includes 9,577 officers. Treatment coefficients and standard errors are estimated from a regression of each covariate on a treatment indicator and strata fixed effects for every combination of rank and skill group (described in Section 3). Standard errors are clustered by market. The joint test is from an F-test on the null hypothesis that all of the baseline covariates are zero in a regression of treatment on the baseline covariates and strata fixed effects with inference clustered by market. For the joint test, missing values of each covariate are imputed with the mean of the covariate within the market and a missing indicator is added to the covariates included in the joint test. *** is significant at the 1% level; ** is significant at the 5% level; * is significant at the 10% level.

Table 2: Compliance with Random Assignment

Variable	N	Control Mean	Coefficient	SE	*P*-value
Panel A: Market Mechanism					
Market Used DA	115	0.143	0.764***	0.069	0.000
Panel B: Type of Match					
First-to-First Match	9,577	0.455	-0.008	0.011	0.497
DA Match (Not First-to-First)	9,577	0.013	0.271***	0.020	0.000
Panel C: Justified Envy					
Justified Envy For Any Job	9,574	0.096	-0.032***	0.008	0.000
Average Number of Jobs Justifiably Envied Per Officer	9,574	0.139	-0.053***	0.012	0.000

Notes: This table summarizes the impact of treatment on market matching mechanisms, match type and justified envy. Panel A shows the impact of being randomly assigned to the treatment group on the market matching mechanism. Panel B shows the impact on officers' match types. Panel C shows the impact on the extent to which officers' have justified envy. An officer has justified envy for a job if the officer prefers the job over her assigned match and the job similarly prefers the officer over its match. Three officers did not submit any preferences and are excluded from all analyses using preference reports. All regressions control for baseline covariates described in section 4 and block fixed effects. Standard errors clustered by market. *** is significant at the 1% level; ** is significant at the 5% level; * is significant at the 10% level.

Table 3: Officers' and Units' Satisfaction with Match

Variable	N	Control Mean	Coefficient	SE	*P*-value
Panel A: Officers' Preferences Over Match					
Ranked Match	9,574	0.848	0.006	0.011	0.596
Rank of Match	8,131	7.851	-1.320**	0.633	0.039
Panel B: Officers' Reported Satisfaction					
Is the officer in the survey data at all?	9,577	0.529	0.013	0.040	0.741
Rate your overall satisfaction with the assignment you received (Standardized)	5,224	0.000	0.048	0.051	0.343
Rate your overall satisfaction with the AIM2 marketplace (Standardized)	5,223	0.000	0.045	0.032	0.168
How likely are you to stay active in the US Army (Standardized)	5,135	-0.000	0.004	0.040	0.930
Panel C: Units' Preferences Over Match					
Ranked Match	9,967	0.702	0.003	0.014	0.822
Rank of Match	7,034	2.740	-0.366**	0.147	0.014

Notes: This table summarizes the impact of matching with DA on officers' and units' preferences over matches. Panel A shows the impact on officers' preferences over matches. Panel B shows the impact on officers' self-reported satisfaction with the match and the marketplace from the post-market survey. Panel C shows the impact on units' preferences over matches (based on their reported preferences). Three officers did not submit any preferences and are excluded from all analyses using preference reports. All regressions control for baseline covariates described in section 4 and strata fixed effects (Equation (1)). Standard errors clustered by market. *** is significant at the 1% level; ** is significant at the 5% level; * is significant at the 10% level.

Table 4: Impact of DA on Longer-Run Outcomes

Variable	N	Control Mean	Coefficient	SE	*P*-value
Panel A: Retention					
Still in Army as of 30 September 2020	9,577	0.985	0.003	0.002	0.147
Still in Army as of 30 September 2021	9,577	0.934	0.011**	0.005	0.025
Still in Army as of 30 September 2022	9,577	0.849	0.003	0.006	0.582
Panel B: Performance in First Year					
Officer Received an Evaluation (Sept. 2021)	9,577	0.890	0.014*	0.008	0.081
Share of Evaluations that are 'Most Qualified' (Sept. 2021)	8,646	0.483	-0.006	0.008	0.462
Panel C: Performance in Second Year					
Officer Received an Evaluation (Sept. 2022)	9,577	0.806	0.014	0.009	0.116
Share of Evaluations that are 'Most Qualified' (Sept. 2022)	7,868	0.498	-0.010	0.011	0.364
Panel D: Promotion Outcomes					
Promoted (Sept. 2022)	9,577	0.296	0.005	0.010	0.642
Has Promotion Percentile (Sept 2022)	9,577	0.321	0.002	0.009	0.844
Promotion Board Percentile (Sept 2022)	3,083	0.497	0.008	0.007	0.254
Promotion Board Percentile w/ Board FE (Sept 2022)	3,083	0.497	0.007	0.006	0.224

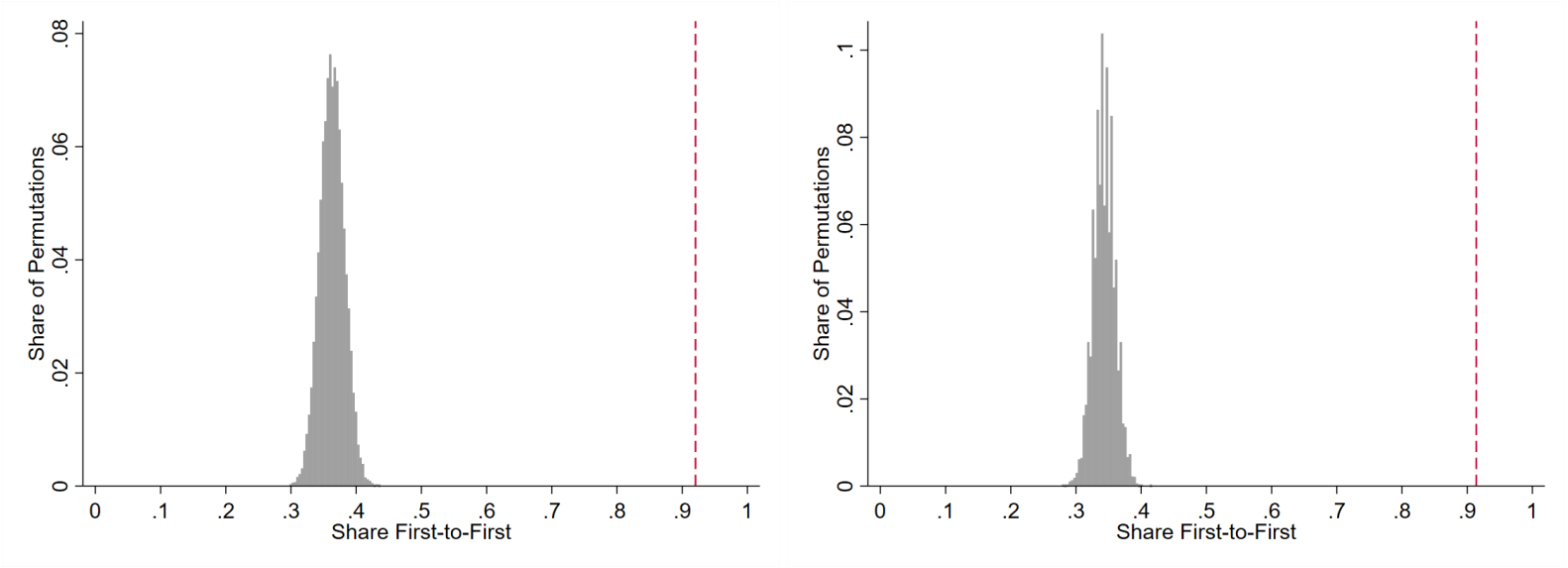
Notes: This table shows the impact of matching with DA on retention (Panel A), performance (Panels B and C), and Promotions (Panel D). Retention is measured as an indicator variable for being in the army on September 30th of 2020 (the year the new match started) and 2021 (one year after the new match started). Performance is measured using officers' evaluation reports. Evaluation reports are the most important factor in determining whether an officer is promoted to the next highest rank. Officers with strong evaluation reports have better performance rankings, and ultimately better chances of being promoted, than officers with weak evaluation reports. We define strong performance as receiving an evaluation report with a rating of "Most Qualified." A rating of "Most Qualified" is the best, and senior raters cannot give this rating to more than 49% of the officers they evaluate. The Army's Evaluation Entry System does not permit senior raters to break this cap of 49%. All regressions control for baseline covariates described in section 4 and strata fixed effects (Equation (1)). Standard errors clustered by market. *** is significant at the 1% level; ** is significant at the 5% level; * is significant at the 10% level.

Table 5: Survey Evidence on the Impact of DA on Strategic Behavior of Officers

Variable	N	Control Mean	Coefficient	SE	*P*-value
Panel A: Mid-Market Survey					
Is the officer in the survey data at all?	9,577	0.874	0.004	0.010	0.653
If position guaranteed, would this position be ranked #1?	8,354	0.844	0.012*	0.006	0.061
Extent submitted preferences reflect your true preferences? (Standardized)	8,350	0.000	0.050***	0.017	0.005
Submitted preferences always reflect true preferences?	8,427	0.238	0.024***	0.007	0.001
Did you rank any position higher because units see if they're in your top 10%?	8,356	0.545	-0.026***	0.010	0.010
Panel B: Post-Market Survey					
Is the officer in the survey data at all?	9,577	0.529	0.013	0.040	0.741
If position guaranteed, would this position be ranked #1?	5,223	0.694	-0.004	0.015	0.768

Notes: This table reports the impact of matching with DA on officers' self-reported strategic behavior. Results in Panel A are based on responses to a survey administered during the final three weeks of the marketplace, when HRC required all officers who logged into the marketplace to complete the survey. Results in Panel B are based on responses to a survey administered when officers learned of their match outcomes, typically 2-3 months after the marketplace closed. All outcomes are indicator variables except the extent submitted preferences reflect true preferences (responses to this question were on a 6 point Likert scale–we standardize this outcome using the control group mean and standard deviation). All regressions control for baseline covariates described in section 4 and strata fixed effects (Equation (1)). Standard errors clustered by market. *** is significant at the 1% level; ** is significant at the 5% level; * is significant at the 10% level.

Figure 1: Preference Coordination Test: Simulated First-to-First (solid histogram) vs. Actual First-to-First (dashed line)



(a) Control Markets

(b) DA Markets

Notes: These figures test the null hypothesis that officers' top choices are uniformly distributed over nearly identical positions using separate randomization tests in treatment and control markets. Holding unit preferences fixed, we randomize which of the identical jobs the officer ranks first 10,000 times. The solid bars show the distribution of the share of first-to-first matches (where the officer ranks the job number 1 and the unit ranks the officer number 1) across iterations. The vertical dashed lines show the actual share of first-to-first matches in this sample. The sample is restricted to officer-job group combinations with potential first-to-first matches because the officer ranked one of the jobs first and at least one of the nearly identical jobs ranked the officer first. We say a pair of jobs is nearly identical if the Jaro-Winkler string similarity score between their job descriptions is greater than 0.9 (with 1.0 indicating the descriptions are identical). We require job similarity be transitive so if job A and B are nearly identical and jobs B and C are nearly identical, we say A and C are nearly identical even if the similarity score is lower than 0.9.

Online Appendix

A A Model of DA with preference signaling

Standard models of DA implicitly assume that preferences are independent of the actions of the other agents in the market. Antler (2015) extends the standard model by assuming that each agent's preference reports are common knowledge and may have an endogenous effect on other agents' preferences. This generalization allows for an officer to potentially benefit from sharing that he or she is ranking a unit as its top choice. Antler (2015) shows that DA is not guaranteed to yield a stable match if an agent's preferences can endogenously depend on other agents' preference reports.

Using a simple counterexample, we show that preference communication can also undermine strategyproofness. In the example, officers can win favor with particular units by ranking that unit first.

Suppose that there are three workers (1,2,3) and three jobs (A,B,C). All workers' true preferences are $u_j(A) > u_j(B) > u_j(C)$ for $j = 1, 2, 3$. Let R^* denote the true ranking of A, B, and C in that order of preference. Let $R^{A \leftrightarrow B}$ be the rank order list that instead moves B to first, A to second, while leaving C in third place.

We assume that jobs A and C are indifferent between all workers and randomize their rank-ordered lists. Job B, however, prefers workers who rank it first to other workers, but randomizes within these two groups. This feature of job B's preferences is a deviation from standard DA models.

Assume workers 2 and 3 truthfully report their preferences in their rankings. Worker 1 must decide how to rank the jobs. The only potentially beneficial manipulation is misreporting B as their first choice instead of A.

Denote worker 1's expected utility from submitting preference list R by $EU_1(R)$. If worker 1 truthfully reports their preferences, R^*, their expected utility is:

$$EU_1(R^*) = \frac{1}{3}u_1(A) + \frac{1}{3}u_1(B) + \frac{1}{3}u_1(C),$$

because all workers have the same preferences and all jobs randomize their rank-ordered lists.

If instead, worker 1 submits $R^{A \leftrightarrow B}$, their expected utility is $EU_1(R^{A \leftrightarrow B}, 1) = u_1(B)$. That is, job B will rank worker 1 first because worker 1 is the only worker who ranked job B first. Worker 1 and job B will therefore be a first-to-first match. If $2u_1(B) > u_1(A) + u_1(C)$, worker 1 is better off misreporting their preferences and ranking job B first than they would be if they had truthfully reported their preferences. This demonstrates that worker-proposing DA is not necessarily strategyproof for workers if even a single job views workers more favorably who rank it highly.[36]

This example shows that, in theory, workers may have incentives to strategically coordinate with units that are not their top choice in order to arrange a first-to-first match. Here, we have assumed that preference reports are common knowledge to simplify the analysis. In practice, workers may need to tell a job that they are ranking it first and jobs must assess the credibility of this report. This type of communication is likely common in most labor market matching problems. For example, both doctors and residency program directors report frequent communication about rankings after interviews even though this type of communication is prohibited by the NRMP's code of conduct (Anderson et al., 1999; Carek et al., 2000; Teichman et al., 2000; Sbicca et al., 2010; Carek, 2012; Berriochoa et al., 2018). Whether this communication is credible or just cheap talk may depend on the potential social consequences of being caught lying. These consequences are likely higher in our setting than in other labor markets. Nevertheless, doctors mention exactly these types of concerns: "It's a small world, especially if you're applying to a competitive specialty or applying heavily to a certain geographic area. If you tell Program A you're ranking them #1, and they rank you back #1, but you actually rank Program B #1 and end up there, Program A's likely going to notice you lied when you don't end up on their list... maybe three years from now you'll apply for a job or fellowship affiliated with Program A and they might remember 'the kid we really liked a few years ago until he lied to us.'"[37]

B Additional Results

[36] The particular Nash Equilibria of this game are of less interest than this result. If workers are identical, however, two workers truthfully reporting and a single worker misreporting is a Nash Equilibrium if $2u_1(B) > u_1(A) + u_1(C)$ and $\frac{2}{3}u_1(A) + \frac{1}{6}u_1(C) > u_1(B)$.

[37] Post on a message board on Jul 23, 2013 https://forums.studentdoctor.net/threads/letters-of-intent-whats-the-deal.1361671/

Table B.1: Multiple Hypothesis Testing Adjusted Inference

Variable	Asymptotic *P*-value	Wild Bootstrap *P*-value	FWER *P*-value	FDR *Q*-value
Panel A: Match Characteristics Outcome Family				
First-to-First Match	0.497	0.531	0.553	0.498
DA Match (Not First-to-First)	0.000	0.000	0.000	0.001
Justified Envy For Any Job	0.000	0.000	0.010	0.001
Count of Jobs Where Officers Have Justified Envy	0.000	0.000	0.010	0.001
Panel B: Preferences Outcome Family				
Ranked Match	0.596	0.684	0.671	0.596
Rank of Match	0.039	0.117	0.319	0.079
Panel C: Survey of Officers' Preferences Outcome Family				
Is the officer in the survey data at all?	0.741	0.815	0.958	0.930
Rate your overall satisfaction with the assignment you received (Standardized)	0.343	0.435	0.843	0.686
Rate your overall satisfaction with the AIM2 marketplace (Standardized)	0.168	0.258	0.729	0.674
How likely are you to stay active in the US Army (Standardized)	0.930	0.936	0.958	0.930
Panel D: Units' Submitted Preferences Outcome Family				
Ranked Match	0.822	0.854	0.854	0.822
Rank of Match	0.014	0.075	0.234	0.029
Panel E: Retention Outcome Family				
Still in Army as of 30 September 2020	0.147	0.214	0.382	0.221
Still in Army as of 30 September 2021	0.025	0.045	0.150	0.075
Still in Army as of 30 September 2022	0.582	0.642	0.637	0.583
Panel F: Performance Outcome Family				
Officer Received an Evaluation (Sept. 2021)	0.081	0.127	0.417	0.233
Share of Evaluations that are 'Most Qualified' (Sept. 2021)	0.462	0.501	0.648	0.462
Officer Received an Evaluation (Sept. 2022)	0.116	0.199	0.426	0.233
Share of Evaluations that are 'Most Qualified' (Sept. 2022)	0.364	0.427	0.648	0.462
Panel G: Promotion Outcome Family				
Promoted (Sept. 2022)	0.642	0.692	0.887	0.845
Has Promotion Percentile (Sept 2022)	0.844	0.868	0.887	0.845
Promotion Board Percentile (Sept 2022)	0.254	0.349	0.667	0.763
Panel G: Strategic Preference Reporting Outcome Family				
If position guaranteed, would this position be ranked #1?	0.061	0.097	0.463	0.106
Extent submitted preferences reflect your true preferences? (Standardized)	0.005	0.019	0.160	0.018
Submitted preferences always reflect true preferences?	0.001	0.004	0.090	0.010
Did you rank any position higher because units see if they're in your top 10%?	0.010	0.013	0.203	0.023
Is the officer in the survey data at all? (Post-Market)	0.741	0.803	0.976	0.769
If position guaranteed, would this position be ranked #1? (Post-Market)	0.768	0.778	0.976	0.769

Notes: The asymptotic p-value is the conventional p-value based on our standard errors clustered by market. The The family-wise error rate (FWER) is the probability of rejecting any true null hypothesis belonging to a "family" of hypotheses. Families are defined by the group of outcomes in each panel. We calculate FWER adjusted p-values using the free step-down resampling methodology of Westfall et al. (1993) using the implementation of Jones et al. (2019). The false discovery rate (FDR) is the expected proportion of false rejections within a family of outcomes (Benjamini and Hochberg, 1995). We calculate FDR adjusted q-values using the implementation of Anderson (2008).

Table B.2: Officer Survey Questions

October-December 2019 Mid-Market Survey (see Panel A, Table 5)

Q13: "If the Army could guarantee you orders to any assignment in your AIM2 marketplace, would this assignment be the position that you ranked number one in your marketplace (your responses to this survey will be kept anonymous; units will not know how you answered this or any other question)?"

Q14: "Did you ever rank any positions in your marketplace higher than normal because units can see if you ranked one of their positions among your top 10 percent of possible choices?"

Q15:"To what extent did the preferences you submitted via AIM2 reflect your true preferences for positions?" *(Possible Responses: "Never", "Rarely", "Some of the Time", "Most of the Time", "Almost Always", "Always")*

October-December 2019 Post-Market Survey (See Panel B, Table 3 and Panel B, Table 5)

Q1: "Rate your overall satisfaction with the assignment you received." *(Possible Responses: "Extremely Positive", "Positive", "Neutral", "Negative", Extremely Negative.")*

Q2: "Rate your overall satisfaction with the AIM2 marketplace." *(Possible Responses: "Extremely Positive", "Positive", "Neutral", "Negative", Extremely Negative.")*

Q3: "If the Army could guarantee you orders to any assignment in your AIM2 marketplace, would this assignment be the position that you ranked number one in your marketplace (your responses to this survey will be kept anonymous; units will not know how you answered this or any other question)?"

Q4: "Did you ever rank any positions in your marketplace higher than normal because units could see if you ranked one of their positions among your top 10 percent of possible choices?"

October-December 2020 Post-Market Survey (See Table B.4, Table B.6)

Q3: "If the Army could guarantee you orders to ANY assignment in your AIM2 marketplace as long as you ranked it number one, would you change the job that you ranked number one in your marketplace?"

Q4: "During the market, did you ever alter your assignment preferences in an attempt to secure a "one to one" match?"

Q5: Did you ever rank any positions in your marketplace higher (or lower) than you otherwise would have because units could see if you ranked one of their positions among your top 10 percent of possible choices?"

October-December 2021 Post-Market Survey (See Table B.5, Table B.7)

Q1: During the market did you ever alter your assignment preferences in an attempt to secure a "one to one" match?"

Q2: "Where does the position you listed a #1 in the AIM Marketplace fall on your true preference list? Your responses to this survey will be kept anonymous; units will not know how you answered this question" *(Possible Responses: "My #1 Preference was truly my #1 Preference", "One of my top three preferences", "One of my top five preferences", "One of my top 10 preferences", "Outside of my top 10 preferences")*

Notes: This table lists the exact working of all officer survey questions and possible responses for questions reported in this paper. Questions without responses listed were Yes/No questions.

Table B.3: Sensitivity of Performance Results to Imputation Techniques

Variable	N	Control Mean	Coefficient	SE	*P*-value
Panel A: Performance in First Year					
Share of Evaluations that are 'Most Qualified', Imp Mean (Sept. 2021)	9,576	0.481	-0.005	0.008	0.553
Share of Evaluations that are 'Most Qualified', Imp 0 (Sept. 2021)	9,577	0.432	0.002	0.008	0.823
Share of Evaluations that are 'Most Qualified', Imp 1 (Sept. 2021)	9,577	0.538	-0.012	0.009	0.149
Panel B: Performance in Second Year					
Share of Evaluations that are 'Most Qualified', Imp Mean (Sept. 2022)	9,577	0.493	-0.016	0.011	0.145
Share of Evaluations that are 'Most Qualified', Imp 0 (Sept. 2022)	9,577	0.405	-0.003	0.009	0.735
Share of Evaluations that are 'Most Qualified', Imp 1 (Sept. 2022)	9,577	0.593	-0.018	0.011	0.124

Notes: This table shows the sensitivity of our performance results to various techniques for imputing missing data. Performance is measured using officers' evaluation reports. Evaluation reports are the most important factor in determining whether an officer is promoted to the next highest rank. Officers with strong evaluation reports have better performance rankings, and ultimately better chances of being promoted, than officers with weak evaluation reports. We define strong performance as receiving an evaluation report with a rating of "Most Qualified." A rating of "Most Qualified" is the best, and senior raters cannot give this rating to more than 49% of the officers they evaluate. The Army's Evaluation Entry System does not permit senior raters to break this cap of 49%. In this table, we show the results if we assume officers missing reports would have received an average evaluation, the maximum possible evaluation, or the worst possible evaluation. All regressions control for baseline covariates described in section 4 and strata fixed effects (Equation (1)). Standard errors clustered by market. *** is significant at the 1% level; ** is significant at the 5% level; * is significant at the 10% level.

Table B.4: October-December 2020 Marketplace Officer Survey Response

Survey Question	Yes	No	N
Q4: During the market did you ever alter your assignment preferences in an attempt to secure a ”one to one” match?	66%	34%	3,905
Q3: If the Army could guarantee you orders to ANY assignment in your AIM2 marketplace as long as you ranked it number one, would you change the job you ranked #1?	46%	54%	3,907

Notes: This table reports results from a survey administered to officers in markets that took place from October through December 2020, one year after the marketplace in the randomized trial. The results are from officers who were in the same rank and occupation as officers in the treated (DA) markets of the randomized trial. There were 5,138 officers in such markets from October-December 2020, of which 3,905 (76%) responded to the survey. All officers in October-December 2020 markets were matched to jobs according to DA.

Table B.5: October-December 2021 Marketplace Officer Survey Response

Survey Question	Percent Yes	Percent No	N
Q1: During the market did you ever alter your assignment preferences in an attempt to secure a one to one match?	58%	42%	2,750
Q2: Where does the position you listed as #1 in the AIM Marketplace fall on your true preference list?			2750
My #1 Preference was truly my #1 preference	62%		1714
One of my top three preferences	21%		581
One of my top five preferences	7%		185
One of my top ten preferences	5%		128
Outside my top 10 preferences	5%		142

Notes: This table reports results from a survey administered to officers in markets that took place from October through December 2021, two years after the marketplace in our experiment. The results are from officers who were in the same rank and occupation as officers in the treated (DA) markets of the randomized trial. There were 3,932 officers in such markets from October-December 2021, of which 2,750 (70%) responded to the survey. All officers in October-December 2021 markets were matched to jobs according to DA.

Table B.6: Additional October-December 2020 Marketplace Officer Survey Responses

	Q4 Resp: "Yes" (N = 2,560)		Q4 Resp: "No" (N = 1,345)	
Survey Question	Yes	No	Yes	No
Q3: If the Army could guarantee you orders to ANY assignment in your AIM2 marketplace as long as you ranked it number one, would you change the job you ranked #1?	53%	47%	34%	66%
Q5: Did you rank any position higher because units see if you ranked one of their positions among your top 10% of possible choices?	67%	33%	32%	68%

Notes: This table reports results from a survey administered to officers in markets that took place from October through December 2020, one year after the marketplace in the randomized trial. The results are split by officers' responses to Question 4 of the same survey, which asks officers if they ever altered their assignment preferences in an attempt to secure a one to one match (see Table B.4).

Table B.7: Additional October-December 2021 Marketplace Officer Survey Responses

	Q1 Resp: "Yes" (N = 1,584)	Q1 Resp: "No" (N = 1,166)
Survey Question		
Q2: Where does the position you listed as #1 in the AIM Marketplace fall on your true preference list?		
My #1 Preference was truly my #1 preference	53%	75%
One of my top three preferences	28%	12%
One of my top five preferences	9%	4%
One of my top ten preferences	6%	3%
Outside my top 10 preferences	4%	6%

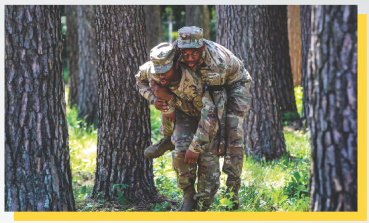
Notes: This table reports results from a survey administered to officers in markets that took place from October through December 2021, two years after the marketplace in the randomized trial. The results are split by officers' responses to Question 1 of the same survey, which asks officers if they ever altered their assignment preferences in an attempt to secure a one to one match (see Table B.5).

Figure B.1: Army Marketplace User Agreement (Implemented Prior to 2021 Marketplace)

ATAP User Agreement

Introduction – ATAP is About Trust.

Trust is the lifeblood of our profession. In fact, trust-building is a big part of why we've created the Army Talent Alignment Process (ATAP) - to increase talent alignment, transparency, and trust. These benefits make the Army more effective, but only if all ATAP users behave honestly and ethically.



"Honest and Ethical" Means Living the Army Values AND Following Marketplace Rules.

ATAP functions best when units and officers trust the assignment marketplace enough to share accurate and granular talent information, resulting in better talent alignment. To build and safeguard that trust, the Army has created a set of **marketplace rules** governing all ATAP interactions. Grounded in applicable federal EO/EEO guidelines and DoD/Army policy, these rules apply to **ALL parties**. Ignoring these rules not only reduces trust but **may have legal consequences** as well.

Marketplace Rules - The "DOs and DON'Ts"

Respect the confidentiality of preferences.

While units and officers may verbally express interest in each other, neither party will ask the other to disclose their preferences in any way. Officer and unit preferences are strictly **confidential**.

Protect officers from retaliation.

Officers expressing good faith concerns with the marketplace will **not** be penalized or retaliated against by units/commanders during the hiring process, to include intentional mischaracterizations of their abilities or motivations.

Don't ask coercive or illegal interview questions.

This includes questions regarding age, gender, religion, sexual orientation, and marital or family status. To prevent this from occurring, units should consider using standardized, prescreened interview questions for all job candidates.

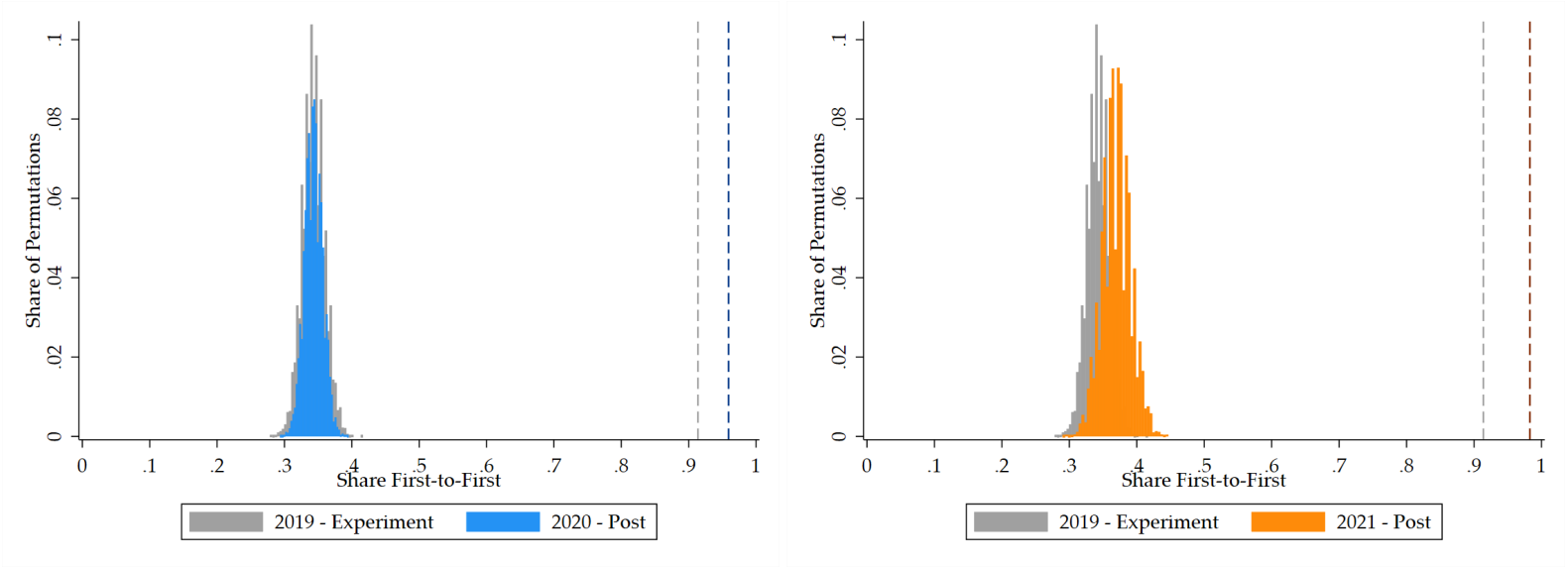
Don't ask for evaluation reports.

While officers may elect to share OERs voluntarily, units cannot compel them to do so, nor may units ask HRC for access to OERs.

Summing Up. The Army is committed to ensuring that ATAP is an inclusive, fair, and professionally rewarding experience for all parties. To help ensure this, please report conduct that is inconsistent with this User Agreement at the link below. Prior to filing a report, we encourage you to review the frequently asked questions.

Notes: All participants in the marketplace that opened in October 2021 were required to acknowledge the user agreement above the first time they logged into the online platform. The marketplace that opened in October 2021 was the second major marketplace after the randomized trial and used DA for all markets.

Figure B.2: Preference Coordination Tests: Post-Experiment Markets



(a) First Post-Experiment Market (Oct 2020)

(b) Second Post-Experiment Market (Oct 2021)

Notes: These figures test the null hypothesis that officers' top choices are uniformly distributed over nearly identical positions using a randomization test in an analogous set of markets in the first and second year after the RCT took place. All of these markets used DA to match officers to jobs, and corresponded to treated (DA) markets in the randomized trial (based on military rank and occupation). Holding unit preferences fixed, we randomize which of the identical jobs the officer ranks first 10,000 times. The solid blue (orange) bars show the distribution of the share of one-to-one matches across iterations in the markets that took place one (two) year after the randomized trial. These bars are overlayed on top of gray solid bars that show the distribution of the share of one-to-one matches from the same exercise executed on DA markets from the randomized trial. The vertical dashed lines show the actual share of one-to-one matches in the corresponding samples. Samples are restricted to officer-job group combinations with potential one-to-one matches because the officer ranked one of the jobs first and at least one of the nearly identical jobs ranked the officer first. We say a pair of jobs is nearly identical if the Jaro-Winkler string similarity score between their job descriptions is greater than 0.9 (with 1.0 indicating the descriptions are identical). We require job similarity be transitive so if job A and B are nearly identical and jobs B and C are nearly identical, we say A and C are nearly identical even if the similarity score is lower than 0.9. See the notes of Figure 1 for additional details.