Information Technologies and the Search for Top Talent in Competitive Job Markets

Xi Jing
Boston University
xjing99@bu.edu

Bin Gu
Boston University
bgu@bu.edu

Abraham Seidmann
Boston University
avis@bu.edu

Vera Tilson
University of Rochester
vera.tilson@simon.rochester.edu

Abstract

The search for highly talented professional employees has been dramatically reshaped by new information technologies. On the one hand, the rise of both dedicated and general-purpose job platforms has significantly reduced the cost and effort for applicants to apply to many more positions. In addition, the introduction of powerful LLM-based tools enables applicants to submit more polished and better-targeted materials, further lowering the signal-to-noise ratio in the application pool. On the other hand, recruiters now face an overwhelming number of applications per opening and have turned to more sophisticated selection rules and AI tools to identify candidates worth interviewing. Because application materials are not perfectly correlated with an applicant's true inherent quality or fit, and because applicants often conceal their job preferences, recruiters continue to rely on personal interviews—both to better gauge candidate quality and to market their positions. Overall, while applicants can now submit large batches of applications easily and at low cost, recruiters, flooded with submissions, find that the search process has become increasingly expensive and time-consuming.

This study situates the U.S. National Resident Matching Program (NRMP) within the broader problem of hiring in markets where specialized platforms dramatically increase application volume. As in other professional labor markets, residency programs compete for applicants in an environment where individuals can cheaply submit large numbers of applications, generating excess competition and noise. We focus on program-level strategies for managing the transition from applications to interviews—through adjustments in interview volume, screening accuracy, or interview cutoffs. Our analysis considers both symmetric settings, where programs are identical in applicant perceptions, and asymmetric settings, where one program is uniformly more desirable. A key and somewhat counterintuitive result is that additional effort by one program does not necessarily disadvantage its competitors and can, in some cases, improve overall match outcomes.

Keywords: Recruiting, Medical match, HR, Information technology, Competitive hiring

1. Introduction

Hiring has become cheap to enter and costly to finish. Digital platforms and LLM tools make it easy for applicants to send polished materials to many openings, but they also flood recruiters with noisy signals. Since applications only partially reveal quality and fit, interviews, which are expensive and scarce, have become the binding resource. Whether in campus recruiting, medical residency, or professional service contracting, employers face the same challenge: *how to identify and secure high-quality candidates when signals are noisy, interviews are expensive, and competitors are simultaneously searching in the same pool.* The core managerial problem is to spend those interviews where they most improve outcomes. This process is characterized by fundamental tradeoffs. On the one hand, an aggressive search strategy risks expending scarce interview budget on candidates who never reciprocate; on the other, a conservative strategy risks missing top candidates, leaving vacancies that damage reputation and performance. At its core, centralized hiring is an information problem: scarce and costly interviews must be allocated to reveal fit under uncertainty.

We focus on the medical residency match in the

URI: https://hdl.handle.net/10125/112108
978-0-9981331-9-5

United States as a high-stakes instantiation of these broader frictions. Every spring, tens of thousands of medical school graduates compete for residency positions through the NRMP Main Match. The theoretical foundation of the NRMP lies in the Gale–Shapley deferred acceptance algorithm, developed in 1962. The global significance of this work was recognized in 2012 when the Nobel Prize in Economics was awarded to Lloyd Shapley and Alvin Roth for their contributions to stable allocation theory and market design practice. Residency matching is one of the most notable real-world applications of these concepts, demonstrating how market design principles can shape the paths to critical career. At the same time, NRMP's monopoly position has drawn significant scrutiny. The U.S. Congress has repeatedly questioned whether the Match system functions as a monopoly, potentially restricting medical graduates' access to the labor market and raising concerns about fairness (U.S. Congress, 2025).

Medical school graduates first submit their applications via the Electronic Residency Application Service (ERAS). Programs then review the applications of candidates who have expressed interest in interviews, select a subset to interview, and rank only those they have interviewed. Applicants, in turn, rank only the programs where they have interviewed. Finally, the NRMP platform applies the Gale–Shapley deferred acceptance (DA) algorithm to produce stable matches. This multi-layered process underscores the platform's dual role: it not only conducts the centralized match but also structures the information environment within which applicants and programs perform their costly search.

Despite its success in filling most positions, the residency match has come under strain. The 2025 NRMP offered over 43,000 PGY-1 (Post-Graduate Year One, i.e., first-year residency) positions to more than 52,000 registered applicants, ultimately filling 94.1% of them (American Medical Association, 2025). Yet the build-up to match day revealed severe congestion. Application costs have fallen due to ERAS and virtual interviewing, producing *application inflation*: U.S. seniors submitted an average of 78 applications in 2023, international graduates 144, nearly double the figures of a decade prior (Carmody et al., 2021; Weissbart et al., 2015). Top-tier programs now receive thousands of applications, whereas less prestigious ones struggle for attention. Although a new preference signaling scheme introduced in the 2024–25 cycle drove a modest 7.3% decline in applications per candidate, with 30–40% drops in some highly competitive specialties, over-application remains the norm, exacerbating interview congestion and stakeholder stress (Association of American Medical Colleges, 2025).

At the root of this congestion are three layers of information search. First, observable signals such as test scores and formal records, eg. USMLE Step II scores and disciplinary records, are public and easy to assess. Second, inherent quality attributes like clinical reasoning and interpersonal skills can be uncovered only through direct interaction, where interviews remain the primary and labor-intensive channel for revealing these subtler signals. Third, and most elusive, are applicants' private preferences: a program may expend an interview on a candidate who never seriously considered them, while another program loses a potential fit due to misperceived interest. The interview season thus functions as a costly search game, where misaligned signals propagate to downstream mismatches and vacancies.

Beyond the applicant–program game, competition across programs introduces another layer of strategic complexity. Since applicants appear in the candidate pools of multiple programs, the programs are effectively competing for the overlapped talent pool. If one program decides to expand its interview scale or adopt more precise screening tools, it will not only change its own hiring opportunities but also reshape the distribution of candidates available to similar programs. The traditional view is that such measures will poach talent and harm competitors. This interdependence suggests that resident matching is not merely a human resource search problem but also a platform design issue, where information architecture and budget constraints jointly determine the overall market outcome.

For the research question, our study asks: (i) In a competitive, centralized hiring market, what strategies should a program adopt to improve its outcome? (ii) What are the externalities of such strategies for competing programs? We study these questions in a stylized two-program DA market with interview budget constraints, focusing on three operational levers: (1) enlarging the interview pool, (2) improving screening precision, (3) avoiding interviewing applicants with overqualified application quality signal.

Our study delivers several contribution to matching in HR market and platform design. First, we develop a stylized two-program game model that isolates how key program levers, interview capacity and screening precision, interact with the centralized DA algorithm. We provide formal results showing how these levers affect both program-level outcomes and cross-program externalities, expanding current IS theories of platform-mediated talent allocation. In addition, we prove a novel insight: under certain conditions, enhancements by one program (more capacity or better precision) can actually benefit competitors by expanding their attainable proposer set, leading to weakly improved match outcomes. This finding runs counter to the common belief that dominant participants necessarily harm weaker programs. We validate our findings using Monte Carlo simulations. This integration demonstrates the practical impact of theoretical levers and supports evidence-based guidance for residency program managers and platform designers. Our method also offers a template for analyzing strategic dynamics in other vertical, platform-mediated labor markets.

2. Related Literature

The residency match operates as a two-sided market in which applicants and programs must locate mutually agreeable counterparts while navigating severe congestion and information frictions.

2.1. Application inflation and interview congestion

Empirical studies show that U.S. medical school seniors submitted an average of 78 applications in 2023, while international medical graduates averaged 144, almost double the volumes observed a decade earlier (Carmody et al., 2021; Weissbart et al., 2015). Perceived competitiveness drives this escalation. In Emergency Medicine, for example, half of applicants acknowledge applying well beyond what they believe is necessary (R. D. Huang et al., 2020). Pandemic-era virtualization, accompanied by remote interviews, waived travel costs, and relaxed documentation, further magnified volumes (Meyer et al., 2022).

Low application costs do not come free. Pinker and Tilson (2013) find that digitized hiring slashes search costs but floods recruiters with lower-fit applicants, making screening the new bottleneck. Arnosti et al. (2021) show that when applications are unconstrained, programs squander screening resources on candidates destined to match elsewhere. The downstream consequence of application inflation is severe interview congestion. Programs hold finite interview slots, yet top-tier institutions attract thousands of applications, whereas lower-prestige programs struggle to secure sufficient interest. The resulting crowding effect allows highly competitive candidates to monopolize interview offers across multiple programs, leaving mid-tier candidates under-interviewed and some positions unfilled (Morgan et al., 2021; Whipple et al., 2019).

2.2. Program responses and proposed remedies

Programs have responded with various rules, notably minimum USMLE cut-offs, though these risk excluding holistically strong applicants (Sweet et al., 2019). Holistic review frameworks exist, yet their labor intensity limits scalability (Love et al., 2023).

Researchers propose several market-level levers. Preference signaling and application caps align applicant intentions with program capacity (Ashlagi et al., 2020; Love et al., 2021). Interview caps distribute opportunities more equitably (Morgan et al., 2021). Sequential invitation protocols and asynchronous offer mechanisms have been modeled to flatten temporal spikes in demand (Vohra & Yoder, 2024). Another approach is the imposition of costs to apply. N. Huang et al. (2025) investigate platform-imposed application costs as a lever for reducing congestion in online matching markets. Similarly, Fradkin et al. (2025) demonstrate through field experiments that providing competition information (number of prior applicants, vacancy age) nudges job searchers toward less crowded vacancies.

Complementary theoretical work indicates that strategically restricting certain actions (for example, limiting who can initiate contact or how many simultaneous interviews can be held), can steer the search toward higher-yield interactions and further alleviate congestion (Kanoria & Saban, 2021). Platforms can also dampen overcrowding by disclosing demand information (e.g. showing how many applications a program has already received), which nudges applicants away from oversubscribed programs and balances attention across the market (N. Huang et al., 2022). Finally, optional verification mechanisms have been shown to function as credible signals of applicant seriousness, increasing match rates and overall welfare (L. Shi & Viswanathan, 2023).

2.3. Matching under search and screening frictions

Classic two-sided matching theory, anchored in the Gale–Shapley deferred-acceptance algorithm (Gale & Shapley, 1962), historically assumed that players know their full preference rankings *ex-ante*. In practice, however, preferences are revealed only through costly interviews, which constrain the information sets available to both sides. Building on this observation,Lee and Schwarz (2017) formalize an interview game in which firms and workers jointly choose interview pools, illustrating how decentralized search can yield coordination failures and vacancies.

Field evidence indicates that the size of the choice set offered to users can meaningfully affect engagement and final match quality and calibrating these sets improves overall efficiency (Jung et al., 2022). More broadly, P. Shi (2023) shows that the relative performance of decentralized versus centralized mechanisms hinges on the structure of participant preferences, underscoring the need for context-specific platform design. On the screening side, an ethnographic study of a large-scale AI hiring deployment reveals that hybrid human-AI workflows, rather than fully automated filtering, produce the highest precision and stakeholder acceptance, highlighting the importance of human oversight in algorithmic classification (Van den Broek et al., 2020). Recent evidence underscores how scalable, AI-generated signals can mitigate search frictions in congested matching markets (Cheng et al., 2025). Complementing the focus on accuracy, Gehlot et al. (2022) model a multi-stage virtual-triage workflow with Colored Petri Nets and find that early-stage filtering smooths downstream capacity and improves overall system efficiency (Gehlot et al., 2022). Dynamic models reveal that greedy assignment rules systematically shorten the waiting time (Ashlagi et al., 2023).

Despite rich descriptive and theoretical work, the field lacks an integrated treatment that (i) jointly incorporates interview capacity, algorithmic screening accuracy, and limit resources; and (ii) considers both program-level competition and program-applicant game. By combining a game-theoretic model with Monte Carlo simulations, our study bridges this gap and contributes actionable guidance for both residency directors and platform designers.

3. Model

3.1. Setting up

We analyze a stylized residency market with two programs, A and B, of comparable reputation but located in different regions. Before turning to the model, it is useful to recall the roles of the two platforms that structure this market. ERAS is the application distribution platform: it collects applicants' materials and transmits them to programs designated by the applicants. NRMP is the allocation platform: at the end of the season, it takes rank-order lists from both applicants (restricted to programs that interviewed them) and programs (restricted to applicants they interviewed and whose quality is above the bar), and applies the Gale–Shapley deferred acceptance algorithm to produce a stable matching. Thus, ERAS governs the information layer, while NRMP governs the allocation layer. Figure 1 summarizes the process. Our analysis focuses on the intermediate stage: how programs use limited interview capacity and imperfect screening tools to decide whom to interview, which in turn shapes the feasible rank-order lists submitted to the NRMP.

Each program $s \in \{A, B\}$ offers k residency positions in a given cycle and may conduct $n_s \geq k$ interviews. We refer to n_s as the program's *number of interview*, and $g_s(n_s)$ as the cost of interviews. To evaluate applicants, each program employs an imperfect screening tool with *precision* $\alpha_s \in [0, 1)$. The market consists of N applicants indexed by i. Each applicant i possesses an inherent quality Q_i, assumed to follow a standard normal distribution, $Q_i \sim \mathcal{N}(0, 1)$. If applicant i is screened by program s, the program observes a noisy signal $S_{s,i} = \alpha_s Q_i + \sqrt{1 - \alpha_s^2}\, \varepsilon_i$, where $\varepsilon_i \sim \mathcal{N}(0, 1)$ is an idiosyncratic noise term. In addition, applicant i has a hidden first-choice preference $P_i \in \{A, B\}$, with $\Pr(P_i = A) = \pi$ and $\Pr(P_i = B) = 1 - \pi$. Programs know the distribution π but not individual realizations of P_i. Each program invites to interview its top n_s applicants ranked by $S_{s,i}$. After interviews, the allocation platform implements the applicant-proposing DA algorithm: applicants may rank only programs that interviewed them, and programs rank interviewed applicants according to their true qualities Q_i. This stylized two-program environment allows us to isolate *competitive spillovers*: when A expands interview slots n_A or improves its screening precision α_A, does program B's final matched cohort quality suffer or improve? We study two preference regimes: (i) a *symmetric demand benchmark*, $\pi = 1/2$, where applicants are equally likely to prefer A or B; and (ii) an *asymmetric demand case*, $\pi = 1$, where all applicants prefer A over B.

Table 1 is a summary of parameters and notations.

We adopt a compact set of maintained assumptions to keep the focus on managerial levers. First, both programs engage in full hiring: $n_A \geq k$ and $n_B \geq k$, and the combined interview pool suffices to fill all seats. Equivalently, the union of interview sets contains at least $2k$ distinct applicants. Second, all the applicants apply to both programs. Third, we define a symmetric demand benchmark (BD), where $n_A = n_B =: n$, $\alpha_A = \alpha_B =: \alpha$, and applicants are ex ante indifferent between programs, i.e., $\Pr(P_i = A) = \Pr(P_i = B) = 1/2$. All comparative statics are taken relative to this benchmark, with departures introduced separately either as an increase in A's number of interviews ($n_A \uparrow$ with α_A fixed) or as an improvement in its screening precision ($\alpha_A \uparrow$ with n_A fixed). Fourth, we retain key features

Figure 1. Timeline of Matching

Table 1. Summary of Notations.

Symbol	Description
i	Applicant index, $i \in \{1, 2, \ldots, N\}$
s	Program index, $s \in \{A, B\}$
n_A, n_B	The number of interviews of programs A and B
k	Open positions each program must fill ($k < n_s < N$ for $s \in \{A, B\}$)
$Q_i \sim \mathcal{N}(0, 1)$	Inherent applicant quality
α_s	Screening precision of program s (correlation between Q_i and $S_{s,i}$)
$\varepsilon_i \sim \mathcal{N}(0, 1)$	Independent noise terms
$S_{s,i} = \alpha_s Q_i + \sqrt{1 - \alpha_s^2}\,\varepsilon_i$	Signal observed by program s for applicant i
$P_i \in \{A, B\}$	Applicant i's first-choice program; e.g., $\Pr(P_i = A) = \pi$ (thus $\Pr(P_i = B) = 1 - \pi$)
$g_s(n_s)$	Weakly increasing interview cost function for program s (shape may be convex/linear/concave)

of the NRMP setting: a *two-program focus* (A vs. B) to isolate head-to-head spillovers; *interview gating* (an applicant may rank only programs that interviewed them), which makes n_s and α_s the operative levers; *program ranking by true quality* Q_i among interviewees, reflecting that interviews reveal enough information to induce an ordering; and *applicant-proposing DA* as the matching mechanism. Finally, to study how the distribution of applicants' preferences shapes outcomes, we also analyze an asymmetric demand case ($\pi = 1$), where all applicants prefer program A. This extreme benchmark highlights how program B's own investments in capacity and precision affect its outcomes when demand is highly concentrated, and serves as a starting point for understanding more general preference asymmetries.

Based on the model, we raise a few propositions (the proof is omitted due to the length limit).

Proposition 1 *Final-match non-deterioration of B when A increases interview capacity in the symmetric demand market starting from the symmetric benchmark.*

Proposition 2 *Final-match non-deterioration of B when A increases the screening precision in the symmetric demand market starting from the symmetric benchmark.*

Proposition 3 *Invariance of A's outcome to B's improvement on capacity or precision in the asymmetric demand market.*

Proposition 4 *Final-match non-deterioration of B when B increases interview capacity in asymmetric demand market.*

Proposition 5 *Final-match non-deterioration of B when B increases screening precision in asymmetric demand market.*

4. Simulation results

4.1. Monte-Carlo Simulation Under Normal Distribution

This section summarizes the Monte-Carlo experiments that operationalize the model in Section 3. We simulate two residency programs $S \in A, B$, each offering 10 positions, and draw an applicant pool that is five times larger than total capacity, yielding a candidate-to-position ratio of κ= 5. Thus, the pool is large enough to fill every slot.

Each applicant possesses an inherent quality $Q_i \sim \mathcal{N}(0, 1)$, which programs ascertain through interviews. A candidate is evaluated identically by both programs. Programs rank candidates for match solely on this inherent quality after interviews. However, the employers can only observe an application-based signal $S_{s,i}$ that correlates with Q_i at α_s during screening, which we vary to capture different screening technologies. We assume that all applicants submit applications to both programs, although each still holds an idiosyncratic preference ordering that the programs do not observe; only its population distribution is known.

After interviewing up to n_s candidates in total (i.e. n_s/k slots per position), the matching algorithm gives priority to applicants' preferences before programs' rankings to ensure stability. Average matched quality in primary round becomes the key performance metric.

We examine three competitive scenarios. The first one is symmetric programs with identical strategies. Programs are equally preferred by applicants and use the same screening technology and number of interviews. The second one keeps equal desirability, but one program tweaks either interview volume or screening accuracy. In third scenario, one program is more appealing and preferred by all applicants. The less appealing one experiments with extra interviews, stronger screening ability, or a cut-off that skips the very top applicants.

The baseline parameters are listed in Table 2.

4.2. Symmetric programs with different strategies

Figure 2 illustrates the impact of asymmetric interview capacity between identical programs as a function of one program's interview volume while the other remains fixed at two interviews per position. Interestingly, as the expanding program increases its interviews from two to ten applicants per position, both of the programs benefit in average quality of incoming class. This result aligns with Proposition 1.

Figure 3 shows the impact of asymmetric screening precision between identical programs as a function of one program's precision while the other remains fixed at 0.4, which illustrates that match outcomes hinge on how much the two programs' interview lists overlap. When both hospitals screen with equal precision ($\alpha = 0.4$), they invite almost exactly the same set of candidates. These applicants cannot accept both offers, so many invitations are squandered in head-to-head competition.

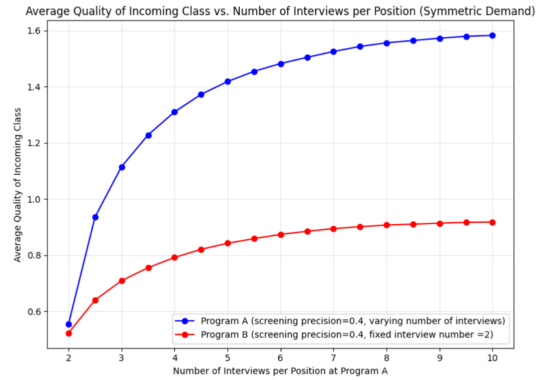

Figure 2. Average Quality of Incoming Class vs. Number of Interviews per Position (Symmetric Demand). Notes: The figure reports the average quality of the matched incoming class when one program maintains a fixed interview volume of 2.0 per position (blue), while the other increases its interview count from 2 to 10 (red). The x-axis shows the number of interviews allocated per residency position at program A.

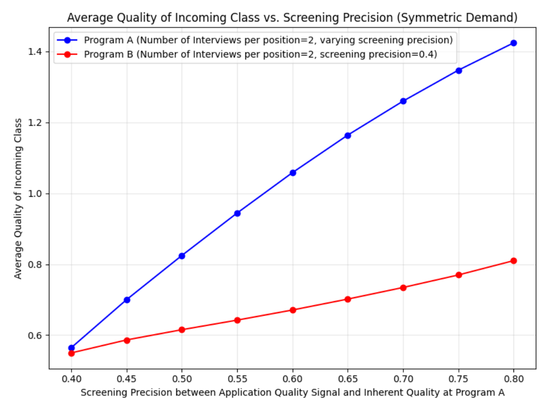

Figure 3. Average Quality of Incoming Class vs. Screening Precision (Symmetric Demand). Notes: The figure reports the average quality of the matched incoming class when one program maintains a fixed screening precision of 0.4 (blue), while the other increases its screening precision from 0.4 to 0.8 (red). The x-axis shows screening precision between application quality signal and inherent quality at program A.

If one program screens more accurately than the other, their interview pools diverge, concentrating on the absolute top tier while the fixed one shifts to the next tier. As the direct competition eases, both of the programs benefit in average quality of incoming class. This result aligns with Proposition 2.

Figures 4 and 5 show the respective effect of

Table 2. Key inputs for all simulation experiments.

Component	Value(s)
Residency programs	Two identical programs; $k = 10$ positions each
Applicant pool	$\kappa = 5$ per position, 100 applicants total
Application quality	$S_{s,i} \sim \mathcal{N}(0,1)$
Interview quality	$Q_{\text{i}} \sim \mathcal{N}(0,1)$
Interview volume n_s	$n_s/k \in \{2, \ldots, 10\}$ invitations
Screening accuracy α_s	$\alpha_s \in \{0.4, 0.5, 0.6, 0.7, 0.8\}$
Candidate preferences	Symmetric: no bias; Asymmetric: all weakly prefer Program A
Cut-off strategy x	$x \in \{0, 10, 20, 30, 40\}\%$ top scores skipped
Trials	1000 Monte-Carlo replications
Outcome metrics $\bar{Q}_{\text{s}}$	Average quality of incoming class

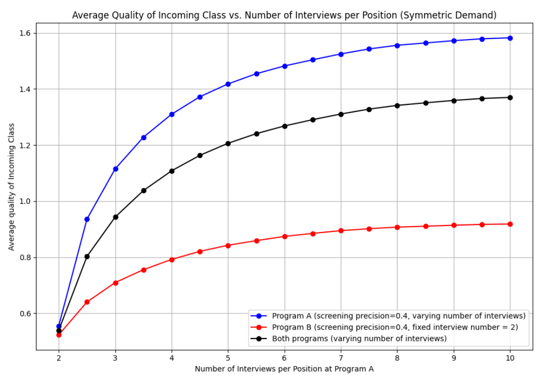

Figure 4. Comparison of Average Quality of Incoming Class vs. Number of Interviews per Position (Symmetric Demand). Notes: The figure compares the average quality of the matched incoming class across programs under symmetric demand when both programs take the same number of interview per position (black) with only one program increases interview numbers (blue) while the other fixed (red). The x-axis shows the number of interviews allocated per residency position.

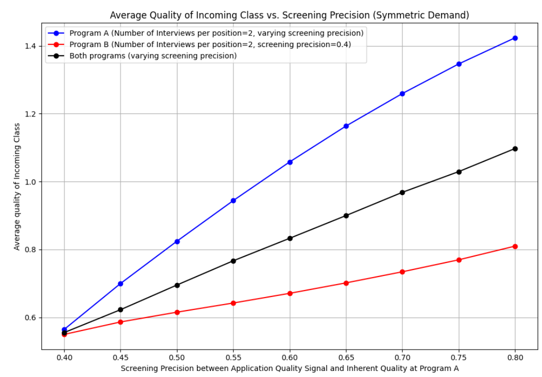

Figure 5. Comparison of Average Quality of Incoming Class vs. Screening Precision (Symmetric Demand). Notes: The figure compares the average quality of the matched incoming class across programs under symmetric demand when both programs take the same screening precision (black) with only one program increases screening precision (blue) while the other fixed (red). The x-axis shows the screening precision between application quality signal and inherent quality.

interview volume and screening precision on average quality of incoming class when both programs enjoy identical appeal and adopt the same strategies. When both programs increase the number of interviews simultaneously, they enjoy a moderate concave gain. As interview volume increases from two to ten interviews per position, we observe a concave relationship with rapidly diminishing returns. Figure 4 reveals that average match quality exhibits sharp initial gains but diminishes beyond six interviews per position. If both programs improve their screening precision at the same time, both benefit — but the gain is, again, more moderate. In addition, as screening precision increases from 0.40 to 0.80, there is a steady improvement in the average quality of the incoming class as illustrated in Figure 5.

4.3. Asymmetric environment

Now we assume that one program is more appealing, i.e. preferred by all the applicants. Figure 6 depicts how interview volume affects outcomes for a less appealing hospital as a function of interviews per position with the preferred hospital interviewing two applicants per position. The average quality of incoming class at the less-preferred program increases as it interviews more, suggesting that disadvantaged programs could utilize interview volume as a powerful strategy to improve the situation, align with Proposition 4.

Figure 7 shows outcomes for a less-appealing hospital as its screening correlation rises from 0.4 to 0.8 while the rival remains fixed at 0.4 and both programs interview exactly two applicants per position.

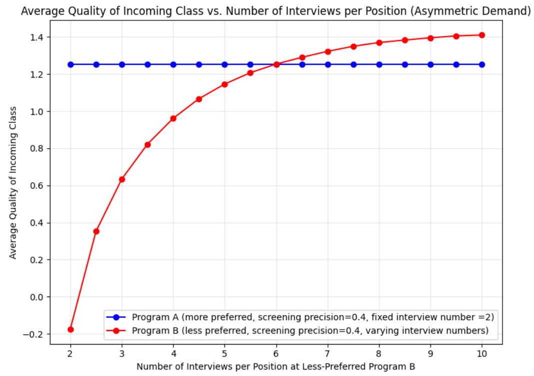

Figure 6. Average Quality of Incoming Class vs. Number of Interviews per Position (Asymmetric Demand). Notes: The figure reports the average quality of the matched incoming class when the more-preferred program maintains a fixed interview volume of 2.0 per position (blue), while the less-preferred one increases its interview count from 2 to 10 (red). The x-axis shows the number of interviews allocated per residency position at program B.

It reveals a steady but modest quality climb. Average match quality climbs up as less-preferred improves its screening precision. The figure suggests that accuracy improvement could mitigate the disadvantages of lower desirability, align with Proposition 5.

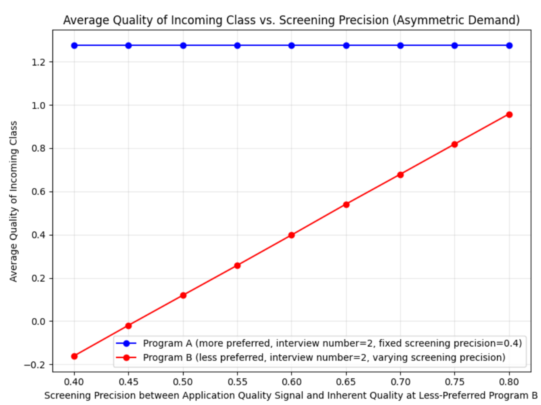

Figure 7. Average Quality of Incoming Class vs. Screening Precision (Asymmetric Demand). Notes: The figure reports the average quality of the matched incoming class when the more-preferred program maintains a fixed screening precision of 0.4 (blue), while the less-preferred one increases its screening precision from 0.4 to 0.8 (red). The x-axis shows screening precision between application quality signal and inherent quality at program B.

Both levers: increasing interviews per slots and the use of a more sophisticated screening tool are costly interventions especially when the application pool is very large. One budget-conscious option for the less preferred program is to avoid competing with the more appealing program for the very top candidates. Figure 8 reveals the effect of strategic upper-bound cutoffs as a function of the top percentage of applicants excluded by the less preferred program. As the cutoff percentage increases from 0 to 50%, we observe a distinctive inverted-U shape in performance metrics. The average match quality peaking at exactly 20% exclusion before declining, demonstrating that the optimal strategy involves avoiding precisely the top quintile of applicants ($2k/N = 20/100 = 20\%$). This non-monotonic relationship reflects the trade-off between candidate quality and acceptance probability. The optimum illustrates how deviations in either direction, whether competing too directly or excluding too many quality candidates, quickly degrade performance.

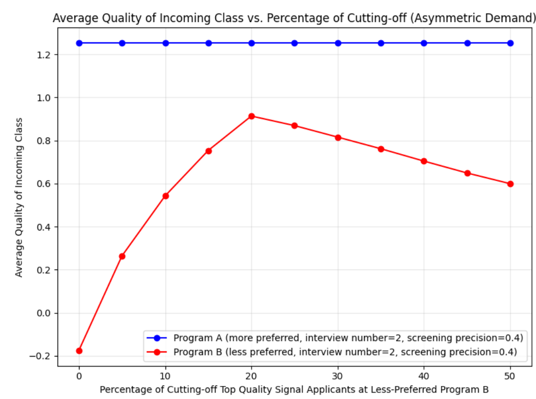

Figure 8. Average Quality of Incoming Class vs. Percentage of Cutting-off (Asymmetric Demand).
Notes: The more appealing program (blue) does not apply any filtering and consistently invites applicants across the full quality spectrum. The less appealing program (red) varies an upper cutoff threshold, excluding the top x% of applicants based on application scores to avoid competing for candidates unlikely to matriculate. The x-axis denotes the upper bound percentage of applicants skipped by the less appealing program.

5. Managerial Implications

Taken together, Sections 1–4 recast hiring as a problem of allocating scarce interviews under noisy signals and competition, with ERAS structuring the information layer and NRMP's DA algorithm governing allocation. Congestion arises because application materials provide only a partial signal of candidate fit, while interviews are costly and limited; consequently, programs must strategically select whom to screen and invite in a market where competitors are vying for the same applicants. The model and simulations clarify how two levers, number of interviews and screening precision, transform the set of applicants a program can feasibly land and, through DA, the quality of the incoming class. In this environment, rivalry is not inevitably zero-sum: at the symmetric baseline with shared noise, one program's operational improvements can leave the other no worse off and often indirectly better by reducing wasteful overlap.

For managers, the most dependable improvement comes from expanding interviews from very low levels up to the point of diminishing returns. In the symmetric environment where both programs use the same screening technology, the gains are steep as interviews per position move from two toward roughly five or six, then flatten markedly thereafter. The Monte-Carlo experiments show this concave pattern clearly (Figure 2), providing a practical ceiling beyond which additional invitations yield little marginal quality.

A second implication is counterintuitive but robust in the updated theory: do not fear a rival's upgrade. At the symmetric baseline with shared idiosyncratic noise and identical programs, a unilateral increase by Program A in either interview volume or screening precision cannot worsen Program B's final top-k quality and, by enlarging the set of applicants who reach B under DA, often improves it (Propositions 1 and 2). The mechanism is monotonicity of DA's top-k with respect to proposals: when A's policy change alters who flows downstream and when, B's attainable set expands or is upgraded without losing access to its previous best options. The simulation figures further corroborate this positive externality (see Figure 2 and Figure 3).

In asymmetric demand, when all applicants weakly prefer the same program, policy levers work differently but remain useful. If everyone prefers A, A's outcome is invariant to B's policy; B cannot harm A by changing its interview volume or precision (Proposition 3). But B can still help itself: increasing B's interview pool and improving B's screening precision both moves weakly improve B's final top-k (Propositions 4 and 5; see Figure 6 and 7).

A budget-friendly policy emerges from the asymmetric scenario: selective avoidance of the very top of the application-score distribution. Rather than fighting unwinnable head-to-head contests over elite candidates who are likely to accept offers elsewhere, a program can impose an upper cut-off and conduct interviews with high-quality but attainable candidates (Figure 8).

Our results show that the distribution of information, rather than increased competition, improves placement quality. These findings also inform the design of HR platforms. Expanding the number of interviews for one program or improving screening accuracy does not harm other programs at a baseline level; in fact, it may even create positive spillover effects by reducing shared inefficiencies. Therefore, policies at the platform level should focus on minimizing overlap rather than solely increasing volume, as this can enhance outcomes for all participants. These insights should be applicable to many competitive job markets.

References

American Medical Association. (2025). Biggest match day ever: Here's what 2025 numbers reveal [Accessed: 2025-09-18]. https://www.ama-assn.org/medical-students/preparing-residency/biggest-match-day-ever-here-s-what-2025-numbers-reveal

Arnosti, N., Johari, R., & Kanoria, Y. (2021). Managing congestion in matching markets. *Manufacturing and Service Operations Management*, *23*(4), 817–834.

Ashlagi, I., Braverman, M., Kanoria, Y., & Shi, P. (2020). Clearing matching markets efficiently: Informative signals and match recommendations. *Management Science*, *66*(5), 2163–2181.

Ashlagi, I., Nikzad, A., & Strack, P. (2023). Matching in dynamic imbalanced markets. *Review of Economic Studies*, *90*(1), 464–492.

Association of American Medical Colleges. (2025). Aamc leading dramatic improvements in transition to residency [Accessed: 2025-09-18]. https://www.aamc.org/about-us/aamc-leads/aamc-leading-dramatic-improvements-transition-residency-just-start

Carmody, J. B., Rosman, I. S., & Carlson, J. C. (2021). Application fever: Reviewing the causes, costs, and cures for residency application inflation. *Cureus*, *13*(3), e13804.

Cheng, Y., Gao, B., Zhang, R., & Li, X. (2025). "Word-of-AI" and Matching Quality: Evidence from a Natural Experiment on Online Review Platforms [SSRN Working Paper No. 5052570]. https://doi.org/10.2139/ssrn.5052570

Fradkin, A., Bhole, M., & Horton, J. J. (2025). Competition avoidance vs. herding in job search: Evidence from large-scale field experiments on an online job board. *Management Science*.

Gale, D., & Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, *69*(1), 9–15.

Gehlot, V., Sloane, E. B., Kirk, M., Miller, E., King, D., Schaffer, J., & Wickramasinghe, N. (2022). COVID–19 and beyond: Taming the virtual healthcare tiger with colored petri nets based modeling and simulation. *Proceedings of the 55th Hawaii International Conference on System Sciences (HICSS 55)*, 3773–3782.

Huang, N., Burtch, G., Huang, A., & Chen, P. (2025). Evaluating the efficacy of application costs for managing congestion in online matching markets. *Management Science (forthcoming)*.

Huang, N., Burtch, G., He, Y., & Hong, Y. (2022). Managing congestion in a matching market via demand information disclosure. *Information Systems Research*, *33*(4), 1196–1220.

Huang, R. D., Lutfy-Clayton, L., Franzen, D., Pelletier-Bui, A., Gordon, D. C., Jarou, Z., Cranford, J., & Hopson, L. R. (2020). More is more: Drivers of the increase in emergency medicine residency applications. *Western Journal of Emergency Medicine*, *22*(1), 77–85.

Jung, J., Lim, H., Lee, D., & Kim, C. (2022). The secret to finding a match: A field experiment on choice capacity design in an online dating platform. *Information Systems Research*, *33*(4), 1248–1263.

Kanoria, Y., & Saban, D. (2021). Facilitating the search for partners on matching platforms: Restricting agents' actions. *Management Science*, *67*(10), 5990–6029.

Lee, R. S., & Schwarz, M. (2017). Interviewing in two-sided matching markets. *RAND Journal of Economics*, *48*(3), 835–855.

Love, E. R., Dexter, F., & Karan, S. B. (2021). Reducing over-interviewing in the anesthesiology residency match. *Cureus*, *13*(8), e17538.

Love, E. R., Reminick, J. I., & Dexter, F. (2023). Over-application and interviewing in the 2021 united states primary care virtual recruitment season. *Cureus*, *15*(5), e39084.

Meyer, A. M., Hart, A. A., & Keith, J. N. (2022). Covid-19 increased residency applications and how virtual interviews impacted applicants. *Cureus*, *14*(6), e26096.

Morgan, H. K., Winkel, A. F., Standiford, T., Muñoz, R., Strand, E. A., Marzano, D. A., Ogburn, T., Major, C. A., Cox, S., & Hammoud, M. M. (2021). The case for capping residency interviews. *Journal of Surgical Education*, *78*(3), 755–762.

Pinker, E., & Tilson, V. (2013). The impact of technology on the labor procurement process. *2013 46th Hawaii International Conference on System Sciences*, 4176–4185.

Shi, L., & Viswanathan, S. (2023). Optional verification and signaling in online matching markets: Evidence from a randomized field experiment. *Information Systems Research*, *34*(4), 1603–1621.

Shi, P. (2023). Optimal matchmaking strategy in two-sided marketplaces. *Management Science*, *69*(3), 1323–1340.

Sweet, M. L., Williams, C. M., Stewart, E., Chudgar, S. M., Angus, S. V., Kisielewski, M., & Willett, L. L. (2019). Internal medicine residency program responses to the increase of residency applications: Differences by program type and characteristics. *Journal of Graduate Medical Education*, *11*(6), 698–703.

U.S. Congress. (2025). House event: Discussion on NRMP and Medical Residency Match [Accessed: Sep. 18, 2025]. https://www.congress.gov/event/119th-congress/house-event/118236/text

Van den Broek, E., Sergeeva, A., & Huysman, M. (2020). When the machine meets the expert: An ethnography of developing ai for hiring. *MIS quarterly*, *45*(3), 1557–1580.

Vohra, A., & Yoder, N. (2024). Matching with costly interviews: The benefits of asynchronous offers [Working Paper, University of Georgia].

Weissbart, S. J., Kim, S. J., Feinn, R. S., & Stock, J. A. (2015). Relationship between the number of residency applications and the yearly match rate: Time to start thinking about an application limit? *Journal of Graduate Medical Education*, *7*(1), 81–85.

Whipple, M. E., Law, A. B., & Bly, R. A. (2019). A computer simulation model to analyze the application process for competitive residency programs. *Journal of Graduate Medical Education*, *11*(1), 30–35.