

# Applicant Characteristics Associated with Program Rank Order Lists

*On-screen:* [NBME, The Match National Resident Matching Program.

Applicant Characteristics Associated with Program Rank Order Lists - a Collaboration.

Sara Balestrieri, PhD, ScM, Director of Research, NRMP.

Christopher Feddock, MD, MBA, MS, Vice President of Educational Strategy, NBME.

October 5, 2024.]

Sara: Welcome to our session entitled Applicant Characteristics Associated With Program Rank Order Lists A Collaboration. My name is Sara Balistreri and I am the Director of research at NRMP.

Chris: Hi Sara. I'm Christopher Feddock, Vice President of Educational Strategy at NBME.

## **Agenda**

Here's our agenda for the session. We're going to begin by reviewing the objectives followed by providing some background information about why this topic is important and how this project came to be.

Next, we'll review our research questions and recap the methods used to carry on our analysis. Then we'll dive into our results focused on results related to applicant type, sex, underrepresentation in medicine (URiM), and USMLE scores.

And finally. we'll summarize the findings we've shared with you and discuss some of our anticipated next steps, before opening the floor for questions from all of you.

## **Session Objectives**

The objectives for our session are to describe the applicant characteristics of residency program rank order lists (ROLs), across different specialties and to differentiate between applicant characteristics on the competitive and non-competitive segments of the rank order lists.

As we'll elaborate on momentarily, the goal of this work is to increase transparency about program ranking behaviors, particularly in light of both programs and specially

stated efforts to align the composition of the resident cohorts with that of the patient populations they serve.

We are going to focus our presentation on a few applicant characteristics, but this work is part of a larger collaboration looking at a wide range of applicant characteristics.

## **Background**

So, what do we hope to gain from looking at the characteristics of applicants included on program rank order lists? Well, we know that applicants submit many residency applications to try to match with a desirable program for applicants. This leads to a burdensome application process due to the inordinate amount of time spent identifying places to apply and perhaps to making decisions on where to preference signal. Meanwhile, program directors must review an increasing large number of applications.

We know that both program directors and applicants lack transparent high quality information that may inform decision making. In fact, many of the proposed solutions, for example, application caps, interview caps, and even preference signaling place additional strain on applicants because transparency of information is still lacking to really inform those application and signaling decisions.

Given the limited information about characteristics of applicants who programs are ranking, applicants are unable to determine whether their application has a reasonable chance of success with a specific program.

With so many applications, program directors have limited information from which to evaluate applicants. As a result, there is a strong agreement about the need for increased transparency in the UME to GME transition.

Further, many have professed goals for the system and for individual specialties such as goals regarding the composition of the physician workforce. However, there's little information about whether those professed goals or intentions have led to action as may be seen in ranking behaviors.

Therefore, this project aims to increase transparency by characterizing the patterns of desirable applicant attributes that reflect program decisions.

## **Research Questions**

Our first research question is by specialty, what program ranking patterns exist with respect to applicant characteristics? We mentioned earlier that our presentation today is

going to focus on just a piece of a bigger collaborative project. That larger project will include a wider array of applicant characteristics, but today we're going to focus on just a few.

We'll take a look at program ranking behaviors in relation to applicant type including US MD seniors or graduates US DO seniors or graduates, US international medical graduates or US IMGs and non-US IMGs.

Then we'll turn to looking at applicant's sex and whether applicants who are underrepresented in medicine. And finally we'll take a look at applicant's USMLE performance, specifically whether they ever failed Step 1 or Step 2 as well as their overall Step 2 CK score.

Next, we will look at what differences exist in applicant characteristics between those on the competitive segment of the program rank order list versus the remainder of the rank order list. We define competitive as the top 25% of each program's rank order list. While those on the remainder of the rank order list, that is the remaining 75% we consider to be non-competitive. This is subjective, but common categorization as anecdotal evidence suggests programs focus more on applicant's rank near the top of their list. So in the end we're comparing the characteristics of applicants in the top 25% to the characteristics of the remaining 75% of applicants.

## **Specialties Included in Our Presentation**

*On-screen:* [A flow chart with three boxes, one leading to second, leading to third.]

1. Unrestricted Program - Match Year Level File:
  - a. N=18,756 program - Match years
  - b. N=44 specialties.
2. Meeting Inclusion Criteria:
  - a. N=13,984 program - Match years
  - b. N=21 specialties.
3. Included in This Presentation:
  - a. N=4,130 program - Match years
  - b. N=5 specialties.]

So, we combine the rank level data from the 2022, 2023 and 2024 match cycles. This produced a program match year file containing almost 19,000 program match years across 44 different specialties.

Now you will hear us refer to the program match year throughout this presentation. And to be clear, we are referring to a single program's rank order list.

Because most programs continue to participate in the match year after year, many of the programs in our sample included three separate rank order lists, one for each of those years. Therefore, programs are counted three separate times as each rank order list is included in our analysis. This is why we'll often use program match year or you will hear us use the shorthand just program but know that we are always referring to program match years as we did not stratify our analyses by match year.

Now, because there are many specialties that have variable numbers of programs and some with very small program size that could impact the reliability of our estimates. We set some very specific specialty inclusion criteria.

So, in at least two out of the three match years, specialties had to have a minimum of 20 programs and at least 50% of the programs in a given match year needed to have a rank order list length of at least 30 applicants. Further, individual program match years needed to have a minimum rank order list length of 30 applicants in order to be included.

When we reduced the number of program match years and specialties using those inclusion criteria, we ended up with almost 14,000 program match years across 21 different specialties.

While data on all 21 of those specialties is forthcoming, given our limited time with you today, we narrowed our focus to five specialties and they were really based on the following criteria.

First was a high percentage of programs and data retained. Second was an inclusion of a variety of specialties including both primary care and surgical. And last was choosing specialties that really demonstrated the range of behaviors across the variables that we'll be talking about today.

So, the specialties we'll present on today are family medicine, obstetrics and gynecology, otolaryngology, pathology and diagnostic radiology. So the data will present represents approximately 4,130 total program match years across those five specialties.

## **Contextualizing the Data**

Before we move on to our results, we want to take a moment to contextualize the data we'll be presenting. First the data reflects the characteristics of applicants included on program rank order lists. Data will be presented by specialty. So, although data are based on individual programs, the data have been aggregated at the specialty level so that individual programs are not identifiable.

One limitation of our studies that we do not have applicant pool data. In other words, we do not have data to reflect the applications to a specialty or to specific programs. So programs can only rank applicants who apply. Therefore, the applicant pool in part shapes these ranking behaviors that we'll be discussing.

Last, multiple factors likely contribute to ranking behaviors. Some of these factors include decisions made by the applicant. Others reflect decisions made by the program.

*On-screen:* [Multiple factors may contribute to ranking behaviors:

- Advice to applicants on competitiveness for specialty and programs
- Applicant decisions to apply to a specialty and individual programs
- Program mission, values and leadership
- Program interview offers
- Applicant interview acceptances
- Program rank list decisions.]

So if we were to walk through the application process, advice that applicants receive from advisors, peers and even social media will inform them where they should target their applications, what programs they apply to this in turn shapes the applicant pool from which programs can even select applicants for interviews. Applicants who accept and attend those interviews are considered for rank order lists.

So the relationship between applicant's behaviors and programs ranking behaviors are intertwined. Therefore we want to reiterate that the data presented reflects the characteristics of applicants on rank order lists but how or why those applicants ended up on rank order lists is quite complex and difficult to unpack.

With that context in mind, let's turn to our results.

## **Results: Applicant Type**

Sara: We'll begin today by discussing program ranking patterns in relation to applicant type.

*On-screen:* [Graph of Applicant Type Composition of Ranked Applicants. Average Percent of Program ROL Comprised of Each Applicant Type Group.

Otolaryngology:

- U.S. MD 98.0%
- U.S. DO 1.0%
- U.S. IMG 0.4%
- Non-U.S. IMG 0.6%

Diagnostic Radiology:

- U.S. MD 77.3%
- U.S. DO 15.7%
- U.S. IMG 2.3%
- Non-U.S. IMG 4.6%

OB/GYN:

- U.S. MD 76.8%
- U.S. DO 19.2%
- U.S. IMG 2.3%
- Non-U.S. IMG 1.7%

Pathology:

- U.S. MD 43.5%
- U.S. DO 17.0%
- U.S. IMG 9.6%
- Non-U.S. IMG 29.9%

Family Medicine:

- U.S. MD 37.7%
- U.S. DO 41.3%
- U.S. IMG 14.0%
- Non-U.S. IMG 7.0%.]

First let's look at the average percentage of program rank order lists comprised by each applicant type with a focus on four different applicant types. First, we have US MDs, which combines US MD Seniors and graduates, US DOs, which combines US DO seniors and graduates, US IMGs and non-US IMGs.

Each bar on this slide totals a hundred percent with each color representing one of the four applicant groups. I also wanted to give you a reminder that these compositions are calculated at the program level than averaged across all programs within a specialty. As can be easily seen at the color breakdown applicant type composition varies widely across these five specialties.

On average, otolaryngology program ROLs consist of nearly only US MD applicants, which we can see in the 98% within the gray bar for this specialty.

US MDs also account for about three quarters of program ROLs in diagnostic radiology and OB/GYN.

The applicant type making up the largest percentage of family medicine ROLs is US DOs followed closely by US MDs.

US and non-US IMG Applicants make up very small percentages of Oto ROLs and small percentages of both diagnostic radiology and OB/GYN ROLs. Meanwhile, these two applicant groups account for significant percentages of the applicants on pathology and family medicine ROLs.

We do see differences between US IMG and non-US IMG representation. For example, there are twice as many US IMGs on family medicine ROLs compared to non-US IMGs.

In pathology, we actually see there are three times as many non-US IMGs compared to US IMGs. Given these vast composition differences, we'll want to keep some of these findings in mind as we take a deeper dive into each of these applicant types.

*On-screen:* [Graph of Percent of Programs Ranking Only U.S. MD Applications.

Otolaryngology:

- Only U.S. MDs on ROL 48.9%
- Only U.S. MDs, Competitive Only 36.5%

Note: 14.6% of programs competitively rank applicant types other than U.S. MDs.

Diagnostic Radiology:

- Only U.S. MDs on ROL 4.7%
- Only U.S. MDs, Competitive Only 22.3%

OB/GYN:

- Only U.S. MDs on ROL 10.2%

- Only U.S. MDs, Competitive Only 21.7%

#### Pathology:

- Only U.S. MDs on ROL 0.0%
- Only U.S. MDs, Competitive Only 7.2%

#### Family Medicine:

- Only U.S. MDs on ROL 0.2%
- Only U.S. MDs, Competitive Only 3.2%]

We'll start by looking at the percentage of programs in each specialty ranking only a single applicant type which commonly occurs for US MD applicants. As you can see on this slide, nearly 50% of Oto programs rank only US MDs, which given the previous slide is not all that surprising.

On the previous slide we also saw that both diagnostic radiology and OB/GYN ROLs were about three quarters US MDs. But here we see that about 5% of diagnostic radiology and 10% of OB/GYN programs only rank US MDs.

While ROLs in these specialties are predominantly US MDs, most programs rank applicant types other than US MDs.

Finally, as you can see on this slide next to no programs in pathology and family medicine rank only US MDs.

This graph again displays the percent of programs in each specialty ranking only US MDs but this time we're focused on programs ranking only US MDs competitively. The blue bar is a repeat from the previous slide, which represents programs that ranked only US MDs anywhere on their rank order list. Meaning these programs only ranked US MDs in both the competitive and non-competitive sections of the ROL.

To put it a different way, the blue bar represents programs who included only US MDs in the top 25% of their ROL and the bottom 75% of their ROL. The orange bar represents the percent of programs ranking only US MDs in the competitive section of the rank order list but did rank other applicant types on the remainder of the ROL.

If we add the blue and orange bars together, we get the percentage of programs who only rank US MDs competitively. For example, while 85.4% of Oto programs ranked only US MDs competitively, 14.6 of programs competitively ranked at least one applicant who was not a US MD.



In the remaining four specialties, it is common to see programs competitively ranking other applicant types.

*On-screen:* [Graph of Percent of Programs Ranking No U.S. DO Applicants.

Otolaryngology:

- No U.S. DOs on ROL 75.9%
- No U.S. DOs, Competitive Only 20.3%

Note: 3.8% of programs competitively rank U.S. DO applicants.

Diagnostic Radiology:

- No U.S. DOs on ROL 13.0%
- No U.S. DOs, Competitive Only 28.8%

OB/GYN:

- No U.S. DOs on ROL 15.9%
- No U.S. DOs, Competitive Only 23.4%

Pathology:

- No U.S. DOs on ROL 3.9%
- No U.S. DOs, Competitive Only 23.1%

Family Medicine:

- No U.S. DOs on ROL 1.0%
- No U.S. DOs, Competitive Only 5.1%]

We're now going to transition our focus to types of applicants who are absent from rank order lists compared to what we just covered, looking at what applicants were only present on rank order lists.

We're going to start with the percentage of programs that did not rank any US DOs. Patterns are very similar to the results looking at programs ranking only US MDs.

Over 75% of Oto programs did not rank any US DOs, whereas fewer than 4% and 1% of pathology and family medicine programs include no US DOs.

The blue and orange bars on this slide represent the same concepts as the blue and orange slide I previously presented. Taken together, the blue and orange bars represent the percent of programs that did not rank any USDO applicants competitively.

For example, 96.2% of Oto programs did not rank any US DOs competitively, meaning 3.8% of programs ranked at least one US DO applicant.

As we saw with the only US MD results, with the exception of Oto, most programs in the remaining specialties rank US DOs competitively.

In family medicine, 6.1% of programs did not rank any US DOs competitively, while 93.9% of the programs in family medicine did.

There's a slightly stronger tendency for programs to not rank US DOs competitively compared to program tendency to only rank US MDs competitively. For example, while 93% of pathology programs competitively rank applicants other than US MDs, 27% of pathology programs did not rank any US DO applicants competitively.

*On-screen:* [Graph of % of Programs Ranking No U.S. IMG Applicants.

Otolaryngology:

- No U.S. IMGs on ROL 82.2%
- No U.S. IMGs, Competitive Only 12.4%

Diagnostic Radiology:

- No U.S. IMGs on ROL 43.6%
- No U.S. IMGs, Competitive Only 36.0%

OB/GYN:

- No U.S. IMGs on ROL 63.5%
- No U.S. IMGs, Competitive Only 20.7%

Note: 15.8% of programs competitively rank U.S. IMG applicants.

Pathology:

- No U.S. IMGs on ROL 10.6%
- No U.S. IMGs, Competitive Only 38.9%

Family Medicine:

- No U.S. IMGs on ROL 24.0%
- No U.S. IMGs, Competitive Only 26.6%]

Let's now look at the percentage of programs ranking no US IMGs. This ranges from 10.6% of pathology programs to 82.2% of Oto programs.

Compared to ranking no DOs, we see higher rates of ranking no US IMGs across these specialties. For example, while 3.9% of pathology programs ranked no DOs, here we see that 10.6% of programs in pathology rank no US IMGs.

Here we see the percent of programs ranking no US IMGs competitively. Taking OB/GYN as an example, 63.5% of OB/GYN programs do not rank US IMGs at all, competitively or non-com competitively, and 20.7% do not rank them competitively, but do rank them non-competitively.

This leads to a total of 84.2 programs in OB/GYN not ranking US IMGs competitively. This also means that 15.8% of OB/GYN programs do rank US IMGs competitively.

Overall, we see a smaller percentage of programs that competitively rank US IMGs. Even in pathology and family medicine, only about half of the programs competitively rank US IMGs.

This contrasts with DOs where, with the exception of Oto, most programs in the other four specialties did competitively rank DOs.

*On-screen:* [Graph of Percent of Programs Ranking No Non-U.S. IMG Applicants.

Otolaryngology:

- No Non-U.S. IMGs on ROL 76.5%
- No Non-U.S. IMGs, Competitive Only 16.5%

Diagnostic Radiology:

- No Non-U.S. IMGs on ROL 35.5%
- No Non-U.S. IMGs, Competitive Only 32.5%

OB/GYN:

- No Non-U.S. IMGs on ROL 64.7%
- No Non-U.S. IMGs, Competitive Only 19.9%

Note: 15.4% of programs competitively rank non-U.S. IMG applicants.

Pathology:

- No Non-U.S. IMGs on ROL 2.5%
- No Non-U.S. IMGs, Competitive Only 21.4%

Family Medicine:

- No Non-U.S. IMGs on ROL 43.4%
- No Non-U.S. IMGs, Competitive Only 26.1%]

Our final pair of applicant type results are going to look at the percentage of programs ranking no non-US IMGs, with a focus on how these behaviors differ from ranking US IMGs.

We observe slightly lower percentages of Oto, diagnostic radiology, and pathology programs ranking no non-US IMGs compared to US IMGs.

We also see similar percentages of OB/GYN programs ranking no non-US and no US IMGs.

In family medicine, we see the percentage of programs ranking no non-US IMGs is nearly double those that did not rank US IMGs. This is a comparison of about 43% to 24%.

Taking our OB/GYN example, again, 64.7% of programs did not rank non-US IMGs competitively or non-competitively, and 19.9% did not rank them competitively, but did rank them non-competitively.

This results in 84.6% of OB/GYN programs that do not rank non-US IMGs competitively and 15.4% that do rank them competitively.

This is nearly identical to the US IMG results I presented previously. However, these trends are dependent on specialty. You may recall that about half the programs in both pathology and family medicine competitively rank US IMG applicants. Here we see about three quarters of pathology programs competitively rank non-US IMGs while only 30.5% of family medicine programs do.

## **Results: Sex**

Next, we'll discuss findings related to applicant sex, focusing specifically on the percentage of program ROLs comprised of female applicants.

Many of the results throughout the rest of this presentation will be presented as box and whisker plots. I am going to take a moment to orient you to what is on the screen. This set of box plots show by specialty the distribution of the percentage of applicants ranked by each program who identified as female.

The solid black bar in the center of each green box represents the median or 50th percentile for each specialty. While the outside ends of the green boxes represent the 25th and 75th percentile for each specialty.

For example, 50% of OB/GYN programs fell below and 50% fell above the specialty median of 90% of ranked applicants being female. 25% of OB/GYN ROLs were less than 86% female, which was the 25th percentile and 25% of ROLs were more than 93% female, which was the 75th percentile.

The whiskers or horizontal lines on either side of the box represent the minimum and maximum values. Looking at each plot from left to right represents the full range of data for each specialty. The individual dots represent outliers that fall outside the whisker plots.

In order to compare the specialties to all residency programs, even outside of the five specialties we're reporting today, we've also plotted the 25<sup>th</sup>, 50<sup>th</sup>, and 75th percentiles for the entire sample of nearly 14,000 program match years. These values are indicated by the three dotted vertical lines.

Looking at just these five specialties, we see the distribution of ROLs comprised of females varies widely. All OB/GYN ROLs are more than 50% female with most ranging from 75 to a hundred percent.

As indicated by the OB/GYN green box being narrower, most OB/GYN programs fell between a tighter range than other specialties, although it is worth noting the outliers to the left.

Diagnostic radiology ROLs range from 10 to 50% and the boxes to the left of the overall sample 25th percentile, meaning nearly 90% of programs have rank lists with less than 43.5% female applicants.

Family medicine has the widest range of ROLs with programs generally falling between 40 and 80% female.

*On-screen:* [Pathology ranges from approximately 37.5 to 80%. The box ranges from approximately 52 to 62%. The median is approximately 55%

Otolaryngology ranges from approximately 35 to 70%. The box ranges from approximately 45 to 55%.]

Now we're going to compare the competitive and non-competitive sections of the ROL. Here, the green boxes represent the percentage of the competitive ROL comprised of females, and the yellow represents the percentage of the non-competitive ROL comprised of females.

For OB/GYN, the median lines in the green and yellow boxes are very similar and the boxes entirely overlap. This suggests that in this specialty, which ranks predominantly females overall, these sections of the ROL are very similar, meaning there's very little evidence that programs differentially rank applicants competitively or non-competitive based on sex.

The other specialties appear to have greater female representation in the competitive section of the ROL as indicated by the median lines in the green boxes being farther to the right than the yellow boxes.

This difference is most notable in diagnostic radiology where we see a greater gap between the ROL section medians and very little overlap of the green and yellow boxes. While diagnostic radiology programs rank predominantly male applicants, this does suggest a trend that female applicants may be ranked more competitively.

*On-screen:* [For Family Medicine, the competitive' s median is about 65% and the non-competitive median is about 60%

For Pathology, competitive' s median is about 60% and the non-competitive median is about 55%.

For Otolaryngology, competitive' s median is about 57% and the non-competitive median is about 48%]

## **Results: Underrepresented in Medicine (URiM)**

Now let's turn our attention to findings related to applicants being underrepresented in medicine. A group which includes Black or African American, Hispanic, Latino, American Indian, Alaskan Native and Pacific Islander applicants.

A reminder that the vertical dotted lines represent the 25th 50th and 75th percentiles for programs across all specialties, not just the five on this slide.

We see that over 45% of OB/GYN programs had ROLs comprised of greater than 24% URiM applicants. This suggests OB/GYN programs tended to rank URiM applicants at a higher rate than other specialties.

Diagnostic radiology had lower representation of URiM on program ROLs. However, looking at the outlier dots, we see that there are several programs in this specialty with greater URiM representation.

Inclusive of outliers, family medicine has the widest range of URIM representation, with many outlier programs having greater representation.

*On-screen:* [Pathology ranges from approximately 5 to 37.5%. The box ranges from approximately 13 to 23%. The median is approximately 20%

Family Medicine ranges from approximately 1 to 50%. The box ranges from approximately 11 to 27%. The median is approximately 18%

Otolaryngology ranges from approximately 1 to 35%. The box ranges from approximately 11 to 20%. The median is approximately 15.5%

Diagnostic Radiology ranges from approximately 1 to 30%. The box ranges from approximately 11 to 18%. The median is approximately 14%]

When we compare the competitive and non-competitive sections of the rank order list on URiM representation, we do not see substantial differences. We can infer this based on the overlap of the green and yellow boxes.

The only somewhat noticeable difference is that the competitive boxes are comprised of slightly higher percentages of URiM compared to non-competitive in OB/GYN indicated by the median line of the green box being to the right of the median line of the yellow box.

We also observed numerous family medicine outliers for both the competitive and non-competitive plots.

*On-screen:* [Graph of Percent of Programs Ranking No URiM Applicants.

Otolaryngology:

- No URiM Applicants on ROL 2.4%
- No URiM Applicants Ranked, Competitive Only 17.4%

Note: 80.2% of programs competitively rank URiM applicants.

Diagnostic Radiology:

- No URiM Applicants on ROL 1.0%
- No URiM Applicants Ranked, Competitive Only 10.3%

OB/GYN:

- No URiM Applicants on ROL 0.1%
- No URiM Applicants Ranked, Competitive Only 6.5%

Pathology:

- No URiM Applicants on ROL 0.0%
- No URiM Applicants Ranked, Competitive Only 9.6%

Family Medicine:

- No URiM Applicants on ROL 0.4%
- No URiM Applicants Ranked, Competitive Only 10.7%]

Finally, we are bringing back the blue and orange graph used earlier in the applicant type results. Here we see the percentage of program years ranking no URiM applicants competitively. The total of the blue and orange bars indicate the total percentage of programs in each specialty that did not rank URiM applicants in the competitive section of the ROL.

For example, 19.8% of Oto programs did not rank any URiM applicants competitively, which also means that 80.2% of Oto programs did rank them competitively.

I have two caveats to give you. The first one in some cases the top quartile of ROLs in our sample was really small, so in some instances not ranking any URiM applicants competitively may be a random occurrence rather than indicative of a larger ranking pattern.

The second and final caveat is that given specialties vary in number of programs within that specialty. The observed percentages also translate to larger numbers of programs for larger specialties such as family medicine.

With that, I'm going to pass it back to Chris to step us through the USMLE results.

## **Results: USMLE**

Chris: The characteristic we'll focus on today is USMLE, for which we examine the percentage of programs that ranked any applicants who ever failed Step 1 or Step 2 CK, as well as an overall assessment of Step 2 CK scores.

*On-screen:* [Graph of USMLE: Ranked any applicants who ever failed.

Otolaryngology:



- Step 1 5.7%
- Step 2 2.2%

#### Diagnostic Radiology:

- Step 1 20.8%
- Step 2 12.9%

#### OB/GYN:

- Step 1 28.4%
- Step 2 17.0%

#### Pathology:

- Step 1 50.6%
- Step 2 52.4%

#### Family Medicine:

- Step 1 89.2%
- Step 2 71.1%]

This bar chart shows the percentage of programs in each specialty that ranked at least one applicant who ever failed USMLE Step 1 or Step 2 CK exams.

So, the percentage of programs ranking any applicant, who ever failed Step 1 is shown in the dark blue bars and the percentage of programs ranking any applicant, who ever failed Step 2 CK is shown in the lighter blue bars.

Overall, you can see that most programs are more likely to rank an applicant who has failed Step 1 rather than Step 2 CK, with pathology being the exception where the percentages of programs are relatively equal.

However, what probably stands out to you the most are the differences across specialties. So, for example, while a low percentage of Otolaryngology programs ranked any applicants who have ever failed Step 1 or Step 2 CK, you can see that over half of pathology programs ranked at least one applicant who ever failed Step 1 and similar for Step 2.

Meanwhile, almost 90% of family medicine programs ranked at least one applicant who failed Step 1, and a bit over 70% ranked at least one applicant, whoever failed Step 2 CK.

From considering the whole rank order list to focus on just the competitive rankings, the top 25% of applicants ranked on each program rank order list.

*On-screen:* [Graph of USMLE: Competitively ranked any applicants who ever failed Step 1.

- Otolaryngology: 1.6%
- Diagnostic Radiology: 3.0%
- OB/GYN: 10.0%
- Pathology: 12.6%
- Family Medicine: 53.1%]

This bar chart shows the percentage of programs in each specialty that competitively ranked any applicants who ever failed Step 1. As with the overall ranking behaviors, we observe that a very small percentage of otolaryngology programs competitively rank any applicants who ever failed Step 1.

For the other specialties we see a similar trend, that programs that rank applicants, who ever failed Step 1, tend to rank them non-competitively.

For example, while almost 21% of diagnostic radiology programs ranked a student who failed Step 1 anywhere on the program rank order list, we can see that only about 3% do so competitively.

This trend of programs more likely to non-competitive rank applicants who are failed Step 1 is also observed in OB/GYN and pathology.

Family medicine provides a bit of a contrast to this trend, recall that almost 90% of family medicine programs rank any applicants who ever failed Step 1. Here we see that 53% of those programs do so competitively.

*On-screen:* [Graph of USMLE: Competitively ranked any applicants who ever failed Step 2 CK.

- Otolaryngology: 0.0%
- Diagnostic Radiology: 1.4%
- OB/GYN: 3.8%
- Pathology: 9.0%
- Family Medicine: 28.9%]

So now let's look at the same bar chart, but this time focusing on the percentage of programs that competitively rank any applicants who have ever failed Step 2 CK. We see some very similar patterns to those that we observed on the prior slide. Coupled with the overall rank order list findings, these again demonstrate a trend of programs being more likely to non-competitively rank applicants who have ever failed.

However, we observed that across these five specialties and even lower percentage of programs competitively rank any applicants who have ever failed Step 2 CK, compared to the percentages observed for Step 1.

Even pathology, which had a slightly higher percentage of programs, ranking applicants who failed Step 2 CK Compared to Step 1, those applicants are predominantly ranked non competitively.

Let's transition to looking at the program level means Step 2 CK scores. Again, we have a figure similar to the box plots that you saw previously in which the middle line in each green box represents the median Step 2 CK score while the left and right edges of the box represent the 25th and 75th percentile scores respectively for each specialty.

We've also again plotted the overall 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> percentiles across all programs regardless of specialty using the vertical dotted and dashed lines. Based on this figure, we observe that Otolaryngology programs mainly rank applicants with high Step 2 CK scores.

Further, you can see that it is a very narrow green box which shows that it has a very concentrated distribution, with the mean for most programs varying between 250 and 265.

In contrast, family medicine programs rank applicants with a wide range of Step 2 CK scores, with program means tending to have lower mean scores compared to the other four specialties presented here. For example, about three fourths of family medicine programs have a rank list with an applicant means Step 2 CK score under 244.

*On-screen:* [Diagnostic Radiology ranges from approximately 243 to 263. The box ranges from approximately 251 to 257. The median is approximately 254.

OB/GYN ranges from approximately 237 to 260. The box ranges from approximately 247 to 253. The median is approximately 249.

Pathology ranges from approximately 229 to 254. The box ranges from approximately 238 to 248. The median is approximately 243.

Family medicine ranges from approximately 221 to 254. The box ranges from approximately 234 to 243. The median is approximately 238.]

Overall, we observe that although some specialties tend to have higher or lower program level means Step 2 CK scores., for many specialties, the program level means Step 2 CK scores vary widely as indicated by the wider green boxes, longer whiskers and outlier programs represented by the dots.

Now let's take a look at the program level mean Step 2 CK scores across the five specialties divided into the competitive versus non-competitive rank order list segments.

So, for each program match here we calculated the mean Step 2 CK score of the competitive rank order list segment as shown in the green box and separately calculated the mean Step 2 CK score. The non-competitive segment shown in the yellow box.

We observe that across specialties applicants ranked competitively tend to have higher Step 2 CK scores as indicated by the median lines in the green boxes being farther to the right than the median lines in the yellow boxes.

Whereas this is largely true across specialties, we see in OB/GYN really a minimal difference between the competitive and non-competitive segments as the two median lines overlap, demonstrating that a significant difference in Step 2 CK scores does not seem to exist between the competitive and non-competitive rank order lists across the OB/GYN specialty.

Here, we display the program level minimum score distributions by specialty, and this is really referring to the minimum Step 2 CK score ranked on each program rank order list.

The distributions for maximum Step 2 CK scores were very similar so we decided not to display them. The first vertical dotted line shows the 25th percentile for all program match years followed by the middle dash line which represents the median and then the far dotted line to the right, which represents the 75th percentile across all programs.

For minimum Step 2 CK scores, we observe that specialties have similar distributions as the mean scores, but there really are smaller differences between those specialties.

Getting a little bit more in depth, we observe that more than 75% of Otolaryngology programs do not rank applicants with Step 2 CK scores under 225, with very few outlier programs in that distribution.

Meanwhile, over 50% of family medicine programs rank applicants with Step 2 CK scores as low as 215 with several outlier programs that have particularly low Step 2 CK scores, as well as a few outlier programs with particularly high minimum Step 2 CK scores compared to other family medicine programs.

*On-screen:* [Diagnostic Radiology, OB/GYN, and Pathology have medians that fall between Otolaryngology and Family medicine for program-level minimum step 2 CK scores. OB/GYN and Pathology contain more outliers that are similar to Family Medicine.

Finally, let's look at the program level means Step 2 CK distributions again, this time comparing the competitive and non-competitive rank order list segments. Overall, we observed that the minimum USMLE Step 2 CK scores are much higher on the rank order list competitive segment compared to the non-competitive segment, as represented by the green boxes being much further to the right than the yellow boxes.

For example, over 75% of pathology programs rank applicants with USMLE Step 2 CK score less than 218 on the rank order list non-competitive segment, compared to 75% of the programs that rank applicants with USMLE Step 2 CK score greater than 218 on the competitive segment of the rank order list.

## **Summary and Next Steps**

So, we've covered a lot of ground in a short amount of time. In an effort to summarize some of our findings, let's talk about what we've observed and then about some of the next steps for this work.

*On-screen:* [Summary. Describe the applicant characteristics of residency program ROLs across specialties:

- Applicant characteristics differ across specialty ROLs.

Differentiate between applicant characteristics on the competitive and non-competitive segments of the ROL:

- Even when ranked, certain characteristics are more commonly found on the non-competitive segment.

Multiple factors may contribute to ranking behaviors:

- Advice to applicants on competitiveness for specialty and programs
- Applicant decisions to apply to a specialty and individual programs

- Program mission, values and leadership
- Program interview offers
- Applicant interview acceptances
- Program rank list decisions]

So in terms of the applicant characteristics of the overall program rank order lists across specialties, we observed that applicant characteristics differ widely across different specialties. For example, some specialties ranked almost exclusively US MD applicants while other specialties ranked a variety of applicant types.

We also saw some specialties ranked predominantly female applicants, while others ranked fairly low percentages of female applicants. And last in some specialties, almost no programs ranked anyone any applicants who had ever failed Step 1 or Step 2 CK, while in other specialties, a fairly large percentage of programs did so.

We also compared the characteristic of applicants ranked competitively and those ranked non-competitively and we found that certain characteristics are more commonly found on the non-competitive segment. For example, US DO applicants and applicants, whoever failed Step 1 or Step 2 CK tended to be ranked less competitively while female applicants and those who had higher Step 2 CK scores tended to be ranked more competitively.

As we mentioned early in our presentation, determining how or why those applicants end up on a rank order list is complex to unpack, given the multitude of factors that may be contributing to those ranking behaviors. We review these at the beginning and here is the list now.

*On-screen:* [Next Steps. Continue data analysis across all specialties:

- Only a snapshot of 5 specialties was presented, but data are available for 21 specialties
- Explore additional demographic variables

Collect additional data in subsequent Match years:

- Data was insufficient to examine cross variable effects (e.g. applicant type and USMLE score)
- Characterize trend changes]

Again, what we've presented today is part of a larger collaborative project between MNRMP and NBME. So, we want to take a moment to touch on some of our next steps.

Most imminently, we will continue our data analysis across all 21 specialties that met our inclusion criteria. We'll also take a deeper dive into some of the other applicant characteristics that we did not include in our presentation today. For example, being a first generation college graduate.

Eventually we hope to expand this collaboration to include additional data in subsequent match years. This would hopefully enable us to take a more nuanced approach to examine, for example, USMLE scores of ranked applicants by applicant type, which we were unable to do given the sample sizes in our current data. It will also allow us to look at trends to identify how changes are occurring over time with respect to applicant characteristics.