I Z A

Do Men Matter to Female Competition Even When They Don't?

Erica G. Birk
Logan M. Lee
Glen R. Waddell

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Do Men Matter to Female Competition
Even When They Don't?

Erica G. Birk
University of Oregon

Logan M. Lee
Grinnell College

Glen R. Waddell
*University of Oregon
and IZA*

ABSTRACT

Do Men Matter to Female Competition Even When They Don't?[*]

A large literature attempts to identify factors that contribute to gender differences in performance and in the decision to compete. We exploit a highly competitive environment in which elite-female athletes are exposed to the presence of men without the element of direct competition, which allows for the identification of psychological effects of competition. Our results suggest that the presence of men affects the performance of female runners differentially across ability, with negative performance effects being concentrated among lower-ability runners.

Corresponding author:

Glen R. Waddell
Department of Economics
University of Oregon
Eugene, OR 97403-1285
USA
E-mail: waddell@uoregon.edu

---

1 Introduction

A large literature suggests that men are more competitive than women with men tending to enter into competition at higher rates (Niederle and Vesterlund, 2007) and perform better conditional on having competition imposed on them (Gneezy et al., 2003). Niederle and Vesterlund (2007) concludes that gender differences in competitiveness are driven by differences in confidence and underlying preferences for entering and performing in competitive environments. Given the implications of this, they call on economists to give greater attention to what contributes to such gender differences, going as far as to suggest that "much may be gained if we can create environments in which high-ability women are willing to compete." To the extent economic gains are achieved through competition—from grades in the classroom to promotion and advancement on the job, competition is largely unavoidable—this is an important task, albeit a difficult one.

We consider the New York City Marathon as an environment in which high-performing females—all of them having selected into this highly competitive environment—are exogenously treated to the potential influence of male competitors, but in a way that has no direct bearing on the competition in which they are engaged. That is, as they are only competing against other females it is a reasonable to anticipate that there should be no influence of these males on female performance. Yet, in this environment, empirical evidence suggests that there are significant declines in the performance of relatively low performing females upon interaction with male competitors.

In particular, elite marathoning presents a limited but unique opportunity for us to con-

sider one aspect of competition and competitiveness that has potentially confounded earlier studies of competitiveness. Namely, we are able to hold constant the return to performance while nonetheless changing the competitive environment with the introduction of men. While we cannot rule out an interpretation of our results as suggestive of feedback effects, we welcome broader notions of the psychology of competitiveness as it relates to the performance of elite-level female competitors.

Unfortunately, there is no similar opportunity to cleanly identify the responsiveness of males to the presence of females, nor the responsiveness of females to other females. Thus, being careful to keep our inference statements conditioned on this, we contribute to the existing literature by identifying a particular sensitivity of female performance to the presence of men, without having the potential typical confoundedness associated with the direct competition with men.[1]

In Section 2 we describe our data and develop our empirical strategy, which we follow with the results in Section 3. We offer concluding remarks in Section 4.

---

[1]There is a large literature on gender differences in competitiveness generally, or the relative performance of men and women in response to competition. See, for example, Altonji and Blank (1999), Goldin and Rouse (2000), Black and Strahan (2001), Campbell (2002), Gneezy et al. (2003), Gneezy and Rustichini (2004), Price (2008), Buser et al. (2014), Anbarci et al. (2014), Flory et al. (2014). There appears to be significant heterogeneity in the size and even the direction of the competitiveness gap based on urban/rural setting (Bjorvatn et al., 2016), geographic region (Cárdenas et al., 2012; Khachatryan et al., 2015), age (Larson, 2005; Samak, 2013), ethnic group (Gong and Yang, 2012), and phase of the menstrual cycle (Buser, 2012; Wozniak et al., 2014). Of particular interest, Cotton et al. (2013) find that the performance gap diminishes with repeated contests with females eventually outperforming men in competitive settings. This may speak to the experience of elite marathons runners who face continuous competition in their careers. Related through their use of long-distance running data, Allen et al. (2014) finds evidence of context-dependent preferences among marathon runners as evidenced by bunching around end times that fall on an hour or half hour. Frick (2011) uses data a cross section of ultra-marathons to argue that men's races are more competitive than women's races. Deaner (2013) argues that the lower time variance observed among the best male runners relative to the best female runners is evidence for a male disposition towards competitiveness.

2 Methods

2.1 Data and identification

For our analysis, we collect data from the 2007 through 2014 New York City (NYC) marathons.[2] Marathons can offer—and the NYC Marathon does—a set of attributes that together facilitate the identification of an important parameter in the potential determinants of competitiveness and the role of gender in competition. In particular, prize money and the rewards associated with marathon performance are often gender specific. Thus, even though both male and female runners can be on the course at the same time, female outcomes are not determined directly by male performance. This restricts the potential determinants of female performance in an important way, offering a rare opportunity (outside of the laboratory) to exploit variation in the presence of men while holding constant the marginal return to female effort or performance. Thus, to the extent female runners systematically respond to male runners, there can be no appeal to within-tournament mechanisms or incentives for explanation. Remaining will be factors that can be characterized as potential contributors to the psychology of female competitiveness. These factors may include pride, shame, or feedback aversion.

However, there are two additional details of the NYC Marathon that allow for clean identification of this potential responsiveness. First, in the NYC Marathon, "elite" runners are managed separately from the general population of entrants, and in a way that we exploit. Specifically, the group of elite women, there are 32 in the typical NYC Marathon,

---

[2]Due to Hurricane Sandy, the 2012 New York City Marathon did not take place.

begin the race approximately 30 minutes ahead of the next wave of runners including a group of (typically 33) elite men. This sets up the prospect that the fastest male runners will begin passing female runners somewhere in the middle of the race. In a typical year, the slowest elite female runner is passed by the fastest male runner between the 20th and 25th kilometers.[3] Second, we observe each runner's split-times—cumulative running times at every five kilometers throughout the race. Constructing a test around differences in cumulative times by whether runners were passed or not is problematic as, of course, slower runners are more likely to be passed. The availability of runner specific split-times allows for the inclusion of runner fixed effects however, we must be careful in considering the identifying variation left behind when we absorb individual unobserved heterogeneity into the error structure.

While a common objective among marathoners is to maintain a constant pace, marathon data often reveal variation in pace throughout races.[4] In the NYC Marathon, for example, average pace declines throughout the race and at higher rates for slower (on average) runners. As such, absorbing only level differences in average pace across runners into the error structure may fall short of protecting against the threat of reverse causality—being passed by a male runner may be endogenous to runner-specific *changes*, and cannot therefore be captured with a fixed component.

Moreover, note that any variation in the ordinal ranking of competitors throughout the

---

[3]Unlike the elite runners, males and females in the general population start together. As such, female runners in the general population are not passed by male runners so much as they are continuously among them, endogenously separating and interacting with them throughout the race. "Treatment" is therefore much-less clear and the conditionally exogenous variation required to identify the causal parameter of interest is not apparent. Thus, we restrict the sample to elite runners, and further exclude from analysis all elite runners who did not finish.

[4]Constant pace is thought to be the strategy of choice in elite marathoning, though there is evidence of some desire for "negative splits," or running the second half of the marathon at a faster pace.

race inadvertently exposes a "runner fixed-effects" approach to identifying across runner-specific observations that vary in the expected returns to performance—if one's declining relative pace implies a declining ordinal rank, for example, any causal effect of being passed on pace will be be confounded with the returns to effort declining as ordinal rank declines. Again, this is threatening to our ability to achieve clean identification through runner fixed effects.

Thus, while we will report specifications inclusive of runner fixed effects, these challenges motivate an alternative source of identifying variation in which we have much more confidence. It is in the second artifact of NYC Marathon data that we find the ability to construct our preferred identification strategy. Namely, using each runner's split-times, we rank-order runners at each split and ask whether, *within* ordinal rank, the pace of those who had been passed at that distance systematically differs from the pace of those who held the same rank at that split in a different year but who had not been passed.[5] Thus, while runner-specific split times facilitate separately identifying the pre- and post-treatment pace of each runner, where "treatment" is then defined as having being passed by the group of fastest male runners since the most-recent split, we restrict identifying variation to that which exists across years

---

[5]It seems evident that a kinked-regression-discontinuity design would be informative in such a case, were we to observe continuous measures of runner pace. Unfortunately, because we observe only average pace over 5km increments we are not able to directly observe the change in pace that occurs at the precise moment a runner is caught. Due to this limitation, it is possible that runners begin to slow down before actually being caught when they first become aware of a male runner approaching. Alternatively, we cannot reject the hypothesis that women speed up to avoid being caught initially and then slow down more dramatically once they have been passed. Thus, we estimated the net effect of any such responsiveness occurring within 5km splits.

within ordinal ranks at each split.[6] We do give up the ability to include unobserved runner heterogeneity, so include among our identifying assumptions that ordinal rank is sufficient to absorb such level effects in ability. Note, however, that a within-split-by-rank estimator does restrict identifying variation to runners who, at the point they are being observed, project the same return to any change in ordinal rank in the tournament.

2.2 Model

Below, we consider whether, in the interval in which she is first passed by male runners, deviations in runner i's pace systematically vary from the predicted pace. We define the unit of analysis as runner i in split s in year t, and—and this is the important consideration we discusses above—we consider variation within the same ordinal rank r at split s. It is this within-rank estimator that restricts our identifying variation to within competitors who project the same return to performance at the point at which we observe them. Further, in establishing a causal relationship between being caught and any resulting change in pace, our identifying assumption is that after controlling for year and distance effects, the within-rank variation across race years at a given distance is independent of average changes in pace. Specifically, we estimate,

$$Pace_{isrt} = \beta Passed_{srt} + \gamma_{sr} + \delta_t + \epsilon_{isrt}, \tag{1}$$

---

[6]For example, those runners who were in tenth place at the 25km split and had just been passed (between the 20th and 25th splits) will be compared to those runners who were also in tenth place at the 25km split but had not been passed. In subsequent heterogeneity analysis, we will separately identify the effect of being passed early in the race versus late, and be careful to identify there how many male runners passed each female runner.

where $Pace_{isrt}$ is the average pace runner i ran the 5km split ending at $s \in [5, 10, ..., 40, 42]$, which we allow to vary with sr-specific effects, γ_{sr} (i.e., we include split-by-rank fixed effects), year-specific effects, δ_t, and by unobservables, ϵ_{isrt}. If being passed by a male runner decreases female-runner pace, $\hat{\beta}$ will be negative in Equation (1). In all specifications, we report robust standard-error estimates, estimates allowing for possible clustering in splits, s, or in split-by-rank, sr, and (where appropriate) in observations of runner, i. All inference statements will be with respect to the most the most conservative standard-error estimates, though qualitative results are not sensitive to any such assumptions.

In identifying the causal parameter associated with being passed by the fastest male runners we assume that, conditional on observables, the five-kilometer split in which the female at rank r is passed by the fastest male is random with respect to the error in predicting her own split time. Important for our analysis is that the 30-minute lead times afforded to elite female runners is sufficiently short that at least some female runners are encroached upon and passed by the fastest male runner. Elite male runners are faster than their female counterparts across all distances. However, a large fraction of elite female runners—almost 40 percent in our sample—are not caught by a male runner. As these competitors are never "treated," they serve well to identify the effect of controls and to fit the average pace within each 5km split.

3 Results

In Table 1 we report the estimates of the effect of being caught in a given 5km split on a female athlete's average pace. We report specifications with only year fixed effects (in Column 1) and year and 5km-split fixed effects (in Column 2). In both models, point estimates are large and negative—our first indication that average female pace may fall when passed by males. In these models, average pace in the 5km split in which females are first passed is roughly between 7- and 10-percent lower relative to the mean pace of 15.68 km/hr.

However, to the extent the fastest males passed females who are slower or are slowing down differentially more within 5km splits, the point estimates in (1) and (2) will be biased down. Adding controls for runners' ordinal ranks in the race in Column (3) cuts the point estimate and effect size considerably, suggesting that with ordinal rank we way well be absorbing the confounding effects that we anticipated imparting negative bias to naive models.

Of course, if slower females are the same as those who are passed, it is possible that bias remains in our Column (3) estimate of β. We therefore introduce individual fixed effects in Column (4), where being passed by the fasted men lowers female pace by just less than one percent (at the mean), explaining roughly 10 percent of a standard deviation in female pace.

In Column (5) we report what is our preferred specification. As we discussed earlier, the identifying variation here comes from differences in pace across similarly ranked runners across race years, some who had been passed between split $s-1$ and split s and some who had not. That is, we include rank-by-split fixed effects. Note, however, that year fixed effects

take on a slightly elevated role, here, as part of the variation in whether females are passed or not (even within specific ranks at specific distances) could arise from whether the marathon was particularly slow or fast in a given year.[7] In this model, being passed by the fasted men leads to a 0.5-percent decline in average pace. As our confidence in the internal validity of the estimator is highest in this model, we are tempted to conclude that there is no significant effect of males on female performance.

However, having arrived at our preferred specification—suggestive of no responsiveness, on average—we consider the potential heterogeneity in treatment falling upon those being passed early and late in the race. While the average effect reported in Table 1 is contributed to by within-rank-distance estimates of treatment, and therefore designed to absorb "rank-ability" into the error structure, to the extent those passed early in the race are lower ability—even within elite competitors there are arguably relative differences in ability—the average effect may be hiding systematic differences across splits. Finding any systematic response among such elite competitors could be surprising, and notable given the very high levels of performance and rigor among such athletes.

In Table 2 we report point estimates from our preferred specification, separately across splits 15km-20km through 40km-42km, which reveals that the null result of Table 1 is hiding important variation in the influence of men on female performance.[8] To the extent that females passed later in the race are stronger performing, heterogenous treatment effects across distance could be suggestive of differential response to feedback generally, or to the direct

---

[7]Of course, to the extent *individual* runners have a bad day, they fall in ordinal rank and we exploit only the within-rank variation for identification.

[8]Note that we are unable to identify heterogeneity across distance using the model of Column (4) in Table 1, as identifying variation is coming from variation across distance for given runners.

influence of male runners, by runner ability. Using data from *Jeopardy!*, Jetter and Walker (2016) suggests that high-ability females are more-positively responsive to the presence of male competitors, suggesting the gender gap may be ability dependent.[9] As anticipated, we see larger falloffs in performance among those passed earlier in the race, with point estimates increases becoming less negative throughout the race—even becoming positive in the 5km split before the final 2km of the race. The heterogeneity we find would also be consistent with finishing the NYC Marathon being increasingly salient toward the end of the race. Though distance fixed effects capture any such general tendency, female competitors are possibly able to enjoy increasing determination when passed—ultimately enough to yield what appears to be boosts in performance when passed late in the race.

In Figure 1, we see that there is variation in the average number of men passing females, by distance. Because male runners tend to spread out throughout the race, women who are passed in the 15km-20km split are passed by roughly 24 men while those passed in the 35km-40km split are passed by only 6. As such, when we estimate treatment effects across distance, we anticipate that point estimates may reflect the net effect of having been passed by different numbers of male runners. Of course, to the extent "better" runners are less responsive to being passed in the first place, a similar decreasing pattern in treatment effect across distance should result, given that it is the better runners who are passed later in the race. In the end, dividing point estimates by the average number of men passing (which we

---

[9]Jetter and Walker (2016) finds that the gender gap in competitive behavior disappears when women are placed in competition with only men. To control for potential selection into competition with two males, they employ an individual fixed-effects specification using only women who competed against two men during one show and against at least one woman in another. As "only the player with the highest dollar score will receive that value in the form of cash and be eligible to compete in the subsequent episode," restricting the sample in this way may identify the effect of gender composition on only higher ability females.

provide in Table 2) reveals a similar pattern of decline.[10]

4   Conclusion

We exploit an opportunity presented in how the New York City Marathon is conducted to consider the potential influence of men on an elite group of female competitors. For identification, we exploit the initial passing of female runners by the fastest male runners, as such passings create a rare opportunity to change nothing about the anticipated rewards to female performance or the returns to effort in the competitive environment, removing an appeal to within-tournament mechanisms or incentives as potential confounders in explaining away how female runners systematically respond to being overtaken by men.

In a runner fixed-effect specification, we find that female pace falls a modest 0.9 percent at the mean (effect size 0.1) with being overtaken by men, on average, which is consistent with negative responses associated with feedback aversion, and with mechanisms such as pride or shame responses to performing, even at this elite level. Our preferred specification— identifying off of similarly ranked females at specific distances with the race, but across years, where in some years they had been overtaken by the men and in other years they had not—suggests that there is no response. This is encouraging, and suggestive that the any gains associated with creating environments in which high-ability women choose to compete

---

[10]We consider the number of runners passing a female runner within a given split, but this is an endogenous regressor, to the extent that treated runners that slow down more are likely to be passed by more male runners in that split. We do not have the ability to separate this endogeneity from the fact that being passed earlier makes it more likely the male runners were still relatively bunched together. When included as a control in our preferred specification results are consistent with Table 2, although the explanatory power is absorbed more by the number of passing males.

will not be dissipated directly through the influence of men. Heterogeneity analysis suggests that there are relative losses evident among lower ability females, even at very high levels of competitiveness.
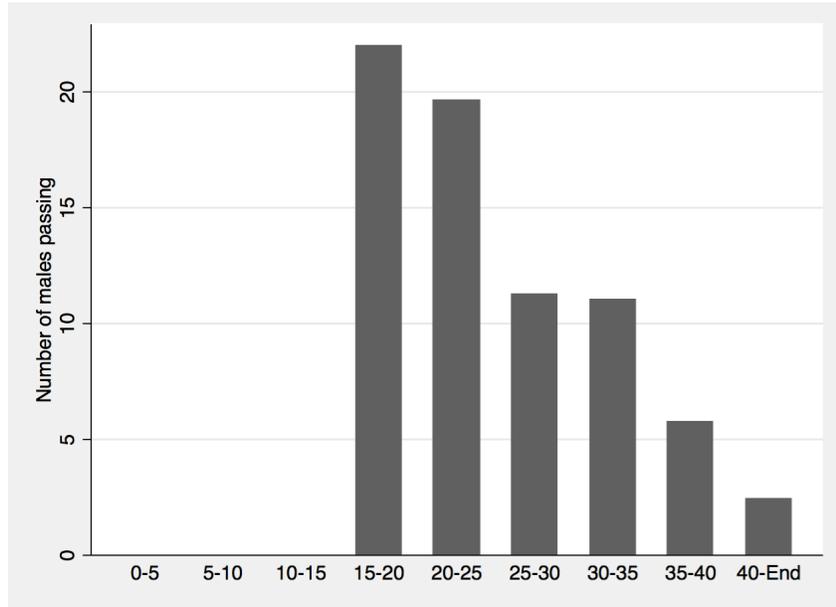
References

Allen, Eric J, Patricia M Dechow, Devin G Pope, and George Wu, "Reference-dependent Preferences: Evidence from Marathon Runners," Technical Report, National Bureau of Economic Research 2014.

Altonji, Joseph G and Rebecca M Blank, "Race and Gender in the Labor Market," *Handbook of Labor Economics*, 1999, *3*, 3143–3259.

Anbarci, Nejat, Jungmin Lee, and Aydogan Ulker, "Win at All Costs or Lose Gracefully in High-Stakes Competition? Gender Differences in Professional Tennis," *Journal of Sports Economics*, 2014, p. 1527002514531788.

Bjorvatn, Kjetil, Ranveig Falch, and Ulrikke Hernæs, "Gender, Context and Competition: Experimental Evidence from Rural and Urban Uganda," *Journal of Behavioral and Experimental Economics*, 2016, *61*, 31–37.

Black, Sandra E and Philip E Strahan, "The Division of Spoils: Rent-Sharing and Discrimination in a Regulated Industry," *American Economic Review*, 2001, pp. 814–831.

Buser, Thomas, "The Impact of the Menstrual cycle and Hormonal Contraceptives on Competitiveness," *Journal of Economic Behavior & Organization*, 2012, *83* (1), 1–10.

_ , Muriel Niederle, and Hessel Oosterbeek, "Gender, Competitiveness, and Career Choices," *The Quarterly Journal of Economics*, 2014, *1409*, 1447.

Campbell, Anne, "A Mind of Her Own: The Evolutionary Psychology of Women.," 2002.

Cárdenas, Juan-Camilo, Anna Dreber, Emma Von Essen, and Eva Ranehill, "Gender Differences in Competitiveness and Risk Taking: Comparing Children in Colombia and Sweden," *Journal of Economic Behavior & Organization*, 2012, *83* (1), 11–23.

Cotton, Christopher, Frank McIntyre, and Joseph Price, "Gender Differences in Repeated Competition: Evidence From School Math Contests," *Journal of Economic Behavior & Organization*, 2013, *86*, 52–66.

Deaner, Robert O, "Distance Running as an Ideal Domain for Showing a Sex Difference in Competitiveness," *Archives of sexual behavior*, 2013, *42* (3), 413–428.

Flory, Jeffrey A, Andreas Leibbrandt, and John A List, "Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job-entry Decisions," *The Review of Economic Studies*, 2014, p. rdu030.

Frick, Bernd, "Gender Differences in Competitive Orientations: Empirical Evidence from Ultramarathon Running," *Journal of Sports Economics*, 2011, *12* (3), 317–340.

Gneezy, Uri and Aldo Rustichini, "Gender and Competition at a Young Age," *The American Economic Review*, 2004, *94* (2), 377–381.

_ , **Muriel Niederle, Aldo Rustichini et al.**, "Performance in Competitive Environments: Gender Differences," *Quarterly Journal of Economics*, 2003, *118* (3), 1049–1074.

Goldin, Claudia and Cecilia Rouse, "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *The American Economic Review*, 2000, *90* (4), 715–741.

Gong, Binglin and Chun-Lei Yang, "Gender Differences in Risk Attitudes: Field Experiments on the Matrilineal Mosuo and the Patriarchal Yi," *Journal of Economic Behavior & Organization*, 2012, *83* (1), 59–65.

Jetter, Michael and Jay K Walker, "Gender in Jeopardy!: The Role of Opponent Gender in High-Stakes Competition," 2016.

Khachatryan, Karen, Anna Dreber, Emma Von Essen, and Eva Ranehill, "Gender and Preferences at a Young Age: Evidence from Armenia," *Journal of Economic Behavior & Organization*, 2015, *118*, 318–332.

Larson, Lia Christine, *When Girls Stop Competing Against Boys: An Experimental Analysis of the Competitive Behavior of Young Children* 2005.

Niederle, Muriel and Lise Vesterlund, "Do Women Shy Away from Competition? Do Men Compete too Much?," *The Quarterly Journal of Economics*, 2007, pp. 1067–1101.

Price, Joseph, "Gender Differences in the Response to Competition," *Industrial & Labor Relations Review*, 2008, *61* (3), 320–333.

Samak, Anya C, "Is There a Gender Gap in Preschoolers Competitiveness? An Experiment in the US," *Journal of Economic Behavior & Organization*, 2013, *92*, 22–31.

Wozniak, David, William T Harbaugh, and Ulrich Mayr, "The Menstrual Cycle and Performance Feedback Alter Gender Differences in Competitive Choices," *Journal of Labor Economics*, 2014, *32* (1), 161–198.

Figure 1

Number of male runners passing an elite female in the 5km in which she was first passed

Notes: Data represent the average number of "Elite" male runners who passed an "Elite" female runner in the 5km-splits among the 2007–2011 and 2013–2014 New York City (NYC) marathons.

16

Table 1

The responsiveness of female pace (km/h) to being passed by the fastest males

	(1)	(2)	(3)	(4)	(5)
Passed	-1.541***	-1.207***	-0.273***	-0.201***	-0.120
	(0.14)	(0.34)	(0.08)	(0.05)	(0.11)
Year FE	Yes	Yes	Yes	Yes	Yes
Split FE	No	Yes	Yes	Yes	Yes
Rank FE	No	No	Yes	Yes	Yes
Runner FE	No	No	No	Yes	No
Split×Rank FE	No	No	No	No	Yes
Observations	2,052	2,052	2,052	2,052	2,052
Mean	15.68	15.68	15.68	15.68	15.68
% change	-9.83	-7.70	-1.74	-1.28	-0.77
Effect size	1.13	0.89	0.20	0.15	0.09

Notes: Observations are at the level of year-runner-distance. Reported coefficients are estimates of the effect of being passed on the average pace between split $k-1$ and split k, where $k \in \{5, 10, .., 40, 42\}$. Robust standard errors are reported in parentheses, allowing for clustering at the level of identifying variation. Percent impacts are relative to the mean. *** significant at 1%; ** significant at 5%; * significant at 10%.

Table 2

The responsiveness of female pace (km/h), by distance

	(1) 15k-20k	(2) 20k-25k	(3) 25k-30k	(4) 30k-35k	(5) 35k-40k	(6) 40k-42k
Passed	-2.421***	-1.333***	-0.508*	-0.038	0.241*	-0.170
	(0.13)	(0.37)	(0.30)	(0.19)	(0.12)	(0.31)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Split×Rank FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	228	228	228	228	228	228
Mean	15.98	15.47	15.92	15.21	15.08	15.05
% change	-15.15	-8.62	-3.19	-0.26	1.60	-1.42
Effect size	2.20	1.09	0.38	0.03	0.15	0.13
Males passing	22.00	19.67	11.28	11.05	5.60	2.42
Effect / pass	-0.11	-0.07	-0.05	-0.00	0.04	-0.09

Notes: Observations are at the level of year-runner-distance. Reported coefficients are estimates of the effect of being passed on the average pace between split $k-1$ and split k, where $k \in \{5, 10, .., 40, 42\}$. Robust standard errors are reported in parentheses, allowing for clustering in split-by-rank bins. Percent impacts are relative to the mean. *** significant at 1%; ** significant at 5%; * significant at 10%.