All in the family: Anna and Lisa Hahner's finishing times in the 2016 women's Olympic marathon

David Cottrell[1] Michael C. Herron[2]

February 3, 2017

[1]Postdoctoral Research Fellow, Program in Quantitative Social Science, Dartmouth College, 6108 Silsby Hall, Hanover, NH 03755 (david.cottrell@dartmouth.edu).

[2]Visiting Scholar, Hertie School of Governance, Berlin, Germany, and Professor of Government, Dartmouth College, 6108 Silsby Hall, Hanover, NH 03755 (michael.c.herron@dartmouth.edu).

"I invested all I had and 300 meters before the finish line, I was next to Lisa. It was a magical moment that we could finish this marathon together. We did not think about what we were doing." – Anna Hahner

Introduction

At 9:30am on August 14, 2016, the Women's Olympic marathon kicked off in Rio de Janeiro, Brazil, when 156 runners from 80 countries across the world left the starting line en route to their destination 42.195 kilometers away. Two hours, twenty-four minutes, and four seconds later, Jemima Sumgong of Kenya would be the first to cross the finishline and take home gold; Sumgong was just three and one-half minutes slower than her prior personal best time in the marathon. Approximately 21 minutes later, twin marathoners from Germany, Anna and Lisa Hahner, would cross the finishline together, holding hands and celebrating a personal victory. Although the Hahners would finish 81st and 82nd, respectively, well behind the winners of the marathon, Anna Hahner would describe their joint finish as a "magical moment."

The media quickly picked up on the Hanher story as an image of the beaming twins finishing hand-in-hand captured a public audience. While many believed the moment was a reflection of the Olympic spirit, not everyone agreed with this rosy interpretation. The twins' happy facial expressions at the finish were portrayed as a bit contrived—smiling like "Honigkuchenpferde," cookies in the shape of a horse—and the sports director of the German Athletics Federation, Thomas Kurschilgen, stirred up controversy when he suggested that the Hahners' photo-finish was no coincidence. Kurschilgen averred that the twins slowed down so as to finish simultaneously and create a spectacle which would "generate media attention." Kurschilgen justified his charge with the fact that the twins ran in the Rio marathon at least 18 minutes slower their personal best times prior to the Olympics.[1] Not surprisingly, Kurschilgen's accusations were denied by the Hahner twins, who

[1] `https://www.nytimes.com/2016/08/17/sports/olympics/twins-`

claimed that their simultaneous finish was simply an unintended coincidence.

What happened in the women's Olympic marathon, and how might we develop a statistical approach that assesses whether the Hahner twin's finish was coincidental or intentional? These two interpretations are clearly at odds. If the former, then the Hahners are to be celebrated and their finish treated as an expression of the spirit behind the Olympic games. If the latter, though, then the twins may have violated this spirt by not trying hard enough. It is perhaps too easy for us to write such a glib sentence—neither of us can fathom being able to complete a marathon anywhere in the vicinity of two and a half hours—but we nonetheless want to know what the data from the Olympic marathon tell us. Was the Hahner finish in the 2016 women's Olympic marathon a lovely coincidence or something else?

Among female Olympic marathoners, the Hahner twins were not alone in their familial ties. The marathon also featured twins from North Korea, Kim Hye-song and Kim Hye-gyong, who posted identical times and finished 10th and 11th in the race, respectively. The Kim finish, unlike the Hahner finish, appears devoid of post-race controversy. Moreover, three triplets from Estonia competed in the Rio marathon, although only two, Lily Luik and Leila Luik, finished it, in 97th and 114th place, respectively. The third Estonia triplet, Liina Luik, recorded what is known as a DNF—an abbreviation that means did not finish. Although our focus here is the Hahner twins, we touch on the Kim twins and Luik triplets in the course of what follows.

Marathon data and our research design

For each participant who started the women's Olympic marathon, we know several things: personal best marathon time prior to the 2016 Olympic games; split times from the Rio marathon course at 5 kilometers, 10 kilometers, and so forth; and, finishing time. We cannot directly observe the

finish-marathon-hand-in-hand-but-their-country-says-they-crossed-a-line.html

effort that an individual put into the race, and we do not know why some runners have DNF results. Some runners may have injured themselves on the course and accordingly dropped out, and others may have dropped out, uninjured, in anticipation of an unsatisfactory result. Of the 156 marathon starters, 133 completed the race and 23 DNFed at various locations throughout the course. The overall DNF rate was thus $\frac{23}{156} \approx 0.15$, and the relatively small sample size at our disposal means that a 95% confidence for this rate is fairly wide, namely, $(0.098, 0.22)$.

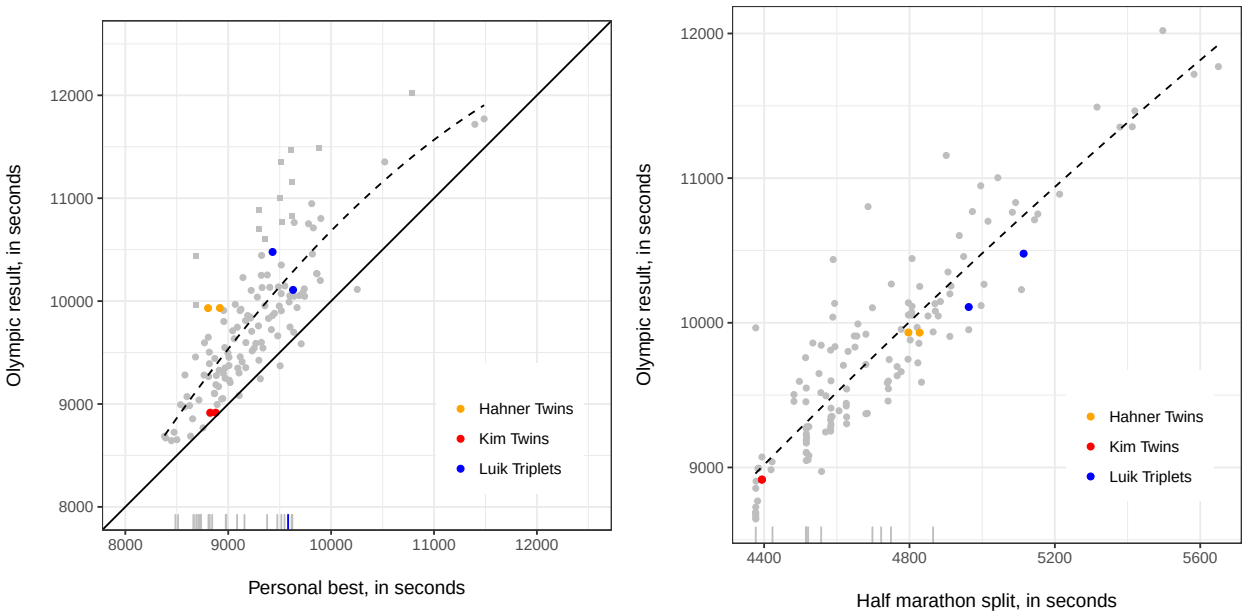
The Kurschilgen accusation against the Hahner twins has two components, that these two women ran slowly *and* that they finished simultaneously. We suspect that Kurschilgen would not have expressed ire at the Hahners had they finished in 1st and 2nd place in Rio, hand-in-hand with wide grins, but of course we do not know this. As such, our investigation of the charges that Kurschilgen offered will distinguish between the idea of slow finish versus a simultaneous finish.

Our research design is twofold. First, we present visualizations that describe various features of the Hahner twin's results, and an important element in our visualizations is the difference between a runner's Rio time and her prior personal best time in the marathon. Overall, our visualizations suggest that the Hahner twins' pace in the marathon was slow, but not excessively so, but that their simultaneous finish was quite unusual given the twins' differences in abilities (and similarly for the Kim twins). We then turn to a regression-based simulation of the marathon, which we develop based on relationships (excluding the Hahner and Kim twins and Luik triplets) between known personal best times and observed Rio finishing times. Our simulations show that the Hahner twins finished suspiciously close to each other given the disparty between their personal best times. Finally, we return to Kurschilgen claims about the Hahners and offer our thoughts about their validity.

Visualizing the Olympic marathon

Figure 1 contains two plots, both of which describe how finishing times from the Rio Olympic women's marathon varied as a function of athletes' personal best times and their Rio half marathon split times. The points in the plots are colored by twin/triplet status, and both plots contain second-order polynomial regression lines. We treat an athlete's prior personal best marathon time as a measure of the athlete's underlying marathon talent, and we interpret an athlete's half marathon split as a measure of the athlete's skill at marathon running on August 14, 2016, the date of the women's marathon in Rio.

Figure 1: Olympic results and measures of marathon ability



(a) Finishing and personal best times

(b) Finishing and half marathon split times

Figure 1b includes a solid 45-degree line as the axes in the figure correspond to full marathons; in addition, the figure's rug marks denote personal best times of runners who DNFed. Given the paucity of points (five of them) below the pictured 45-degree line, the vast majority of Olympic marathoners ran slower in Rio compared to their personal bests. Relative to personal best times, the Hahner twins were definitely on the slow side, but a number of runners had greater differences

between their Olympic times and their personal bests than the Hahners. These runners are denoted with squares in Figure 1a, and there are 13 such symbols, highlighting approximately 10% of the racers who completed the marathon. The figure shows moreover that there were two women who had personal best times slightly faster than the Hahner's and yet finished after the German twins. Although Figure 1a suggests that the Hahner twins were slower than one would have expected given their previous best marathon times, it is not consistent with the accusation that they dramatically slowed down in the Rio marathon.

One limitation of using personal best times as indicators of underlying athletic talent is that these times are potentially confounded by the marathon courses at which they were set (some courses, like the Berlin marathon, are known for fast times) and race conditions like weather. In addition, personal best times may not capture raceday idiosyncrasies that might affect individual runners. With that in mind, Figure 1b plots marathon finishing times against half marathon splits. This figure has a regression line as before but no 45-degree line.

Figure 1b shows that there was nothing abnormal about the Hahner twins' overall finishing times, conditional on their half marathon split times. This is consistent with the previous Figure 1a and inconsistent with accusations made against the Hahner twins, at least the part of the accusation that focused on their overall pace.

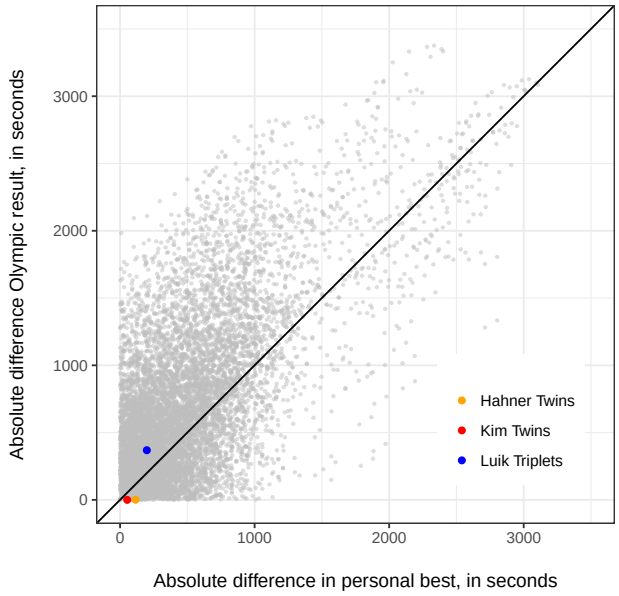
If the Hahner twins did not slow down excessively, might they have run somewhat strategically at the end of the race in order to generate a simultaneous finish? We now offer a visualization that speaks to this question.

The personal best times of the Hahner twins were 115 seconds apart and their official finishing times were separated by one second. Is such a 115 to one compression typical among pairs of runners or, in contrast, was it unusual? Are there other pairs of marathoners who had a difference between personal best times of 115 seconds apart and, if so, how close where their finishing times?

Of the 133 marathon finishers, there are $\binom{133}{2} = 8,778$ pairs of runners. Of these and ignoring the Hahner twins, ten had exactly a 115 second gap in personal best times. Of these ten pairs,

differences in finishing times, in seconds, are as follows: 36, 93, 172, 319, 379, 459, 552, 671, 675, and 739. In other words, of all pairs of runners in the Rio marathon who had a personal best difference that was equivalent to the Hahner twins' difference, the twins had the greatest compression based on finishing time.

Figure 2: Differences in personal best times and finishing times



We can generalize this result by looking at all pairs of runners in the marathon. For all 8,778 pairs of 133 finishers, Figure 2 plots differences in finishing times against differences in personal best times, and pairs of twins/triplets are identified by the same color scheme we used earlier.

Consider first the Hahner twins. They are effectively located on the horizontal axis because their difference in finishing times is one second. However, there are many points about the Hahner's orange dot, and this shows that, conditional on an approximate 115 second difference in personal best times, most marathoners did not have close finishing times like the Hahnes. Some pairs of runners with around 115 second personal best differences had finishing time differences of 1000 seconds, i.e., in excess of 15 minutes. Of course the points in Figure 2 are not independent, but they provide a sense of the dispersion in finishing time differences that one can expect conditional

on differences in personal best times.

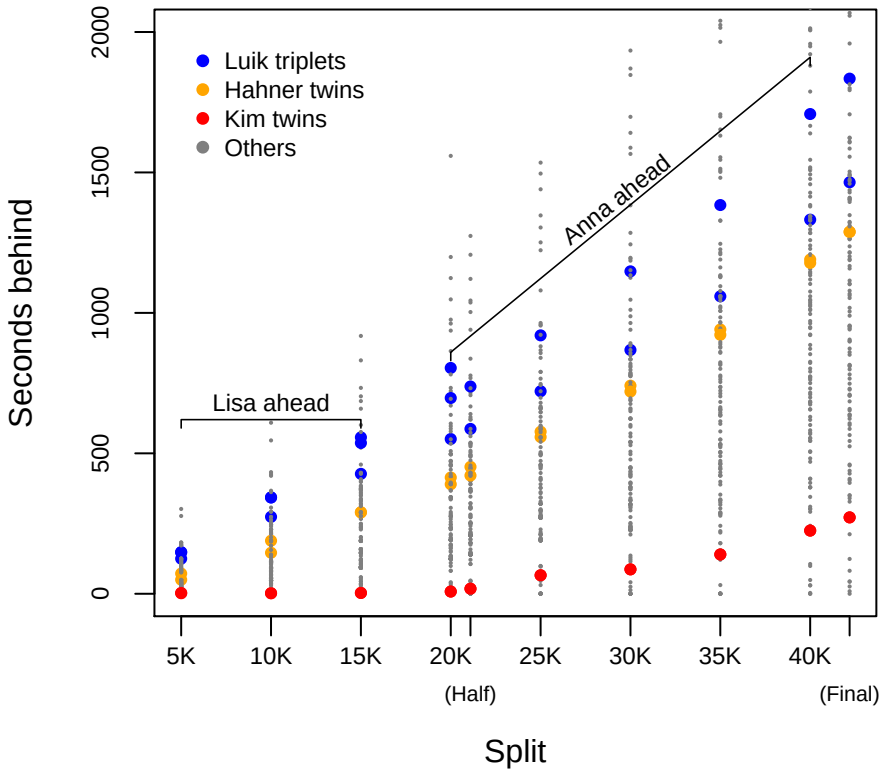
Thinking about the accusations leveled against the Hahner twins, Figure 2 suggests that Anna and Lisa Hahner did indeed run with an eye on each other. In fact, the same can be said of the Kim twins, who ran seemingly in lockstep throughout the entire Rio marathon. The North Korean twins had a personal best difference of 53 seconds and a finishing time difference of literally zero seconds. Beyond these twins, there were eight pairs of Rio runners with a 53 second personal best difference, and resulting finishing time differences are as follows: 9, 51, 228, 340, 352, 571, 662, and 751. As in the Hahner case, the Kim twins compressed their finishing times more than any other pair of runners with a similar personal best difference.

To get a sense of the lockstep nature of the Kim twins race, Figure 3 describes each runner's status at the various split times on the marathon course. Each dot in the figure depicts a recorded split and the number of seconds each runner was behind the race leader at the time. There are more dots at earlier splits due to the accumulation of DNFs. The usual color scheme applies here, and the Estonian DNF prior to the half marathon split is evident. The orange dots in the figure represent the Hahner twins, and as noted in the figure Lisa Hahner was ahead of her sister through 15 kilometers at which point Anna surged. Figure 3 contains two red dots representing the North Korean Kim twins, but this is not visually apparent because the Kim twins had identical split times during the entire race marathon.

Simulating the Marathon

We now consider the following question: if we take into consideration the Hahner twin's similarities as runners as well as natural variation in marathon finishing times, what is the probability that the twins would finish the Rio race at roughly the same time and/or sequentially? To answer this question we need to know the counterfactual distribution of potential marathon finishing times that would have occurred had the Hahner twins independently (in particular, of each other) and repeat-

Figure 3: Runner status by split



Note: each dot represents one runner at a split. DNFs are not pictured, and splits are not to scale.

edly run the Rio marathon, holding constant marathon conditions, the abilities of other runners, and so forth. Access to such a distribution would establish the set of potential outcomes that could have occurred on August 14, 2016, and we could in principle use this distribution to determine the likelihood that a simultaneous finish by the Hahners, or at least a near-simultaneous finish, occurred by chance alone. If these twins rarely finish the marathon together in such a counterfactual world, then one might be skeptical that their observed finish occurred without coordination—as they in fact claimed.

Unfortunately for us but nonetheless fortunately for the race's participants, it is not possible to rerun the women's Olympic marathon in Rio to establish a distribution of potential race outcomes for the Hahner twins. However, we can attempt to simulate this distribution by estimating the distribution of every other runner's finishing time, conditional on prior abilities, and drawing from that distribution in order to observe the likelihood that, for example, Lisa and Anna Hahner finished simultaneously.

To estimate the conditional distribution of each Rio final results, we assume that each runner's marathon time Y_i is distributed normally with a mean that is a quadratic function of her personal best time X_i:

$$Y_i \mid X_i \sim N\left(\beta_0 + \beta_1 X_i + \beta_2 X_i^2 + e_i, \sigma\right).$$

where $N(\cdot, \cdot)$ denotes a normal distribution. Personal best times are all prior to the Rio marathon. We estimate β_1 , β_2, and σ using ordinary least squares and all finishing runners except for the twins and triplets. For a simulated marathon, we then draw a runner's time from an estimated distribution, conditioned on the runner's personal best marathon time. Once a race is simulated for all runners, twins and triplets included, we locate Anna and Lisa Hahner in our set of simulated results and record both the time between their simulated finishes and the difference in their simulated ranks. We then simulate a new race—drawing a new set of finishing times—and record the same quantities. In particular, the steps of the simulation are as follows.

1. Ignoring twins and triplets, estimate with least squares a linear model that predicts a runner's finishing time Y_i based on her pre-Olympic personal best time X_i and her best time squared.

2. Extract the resulting coefficient vector, estimated covariance matrix, and estimated regression variance from this model.

3. For each simulated rate, draw intercept and slope estimates $\tilde{\beta}_0$, $\tilde{\beta}_1$, and $\tilde{\beta}_2$, respectively, from

a multivariate normal distribution with mean equal to the previously estimated coefficient vector and covariance equal to the previously estimated covariance matrix.

4. Given values of $\tilde{\beta}_0$, $\tilde{\beta}_1$, and $\tilde{\beta}_2$, for each runner draw an error $\tilde{e}$ from a normal distribution with mean zero and a standard deviation equal to the standard deviation of the original regression model's residuals.

5. For each runner, predict her final result by combining the randomly generated beta coefficients and error terms, $\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 X_i + \tilde{\beta}_2 X_i^2 + \tilde{e}$.

6. Eliminate each runner from the simulated race with a probability equal to the observed fraction of marathoners who did not finish the marathon.

7. Calculate the difference in time and difference in ranking between Anna Hahner and Lisa Hahner, assuming both finished the race.

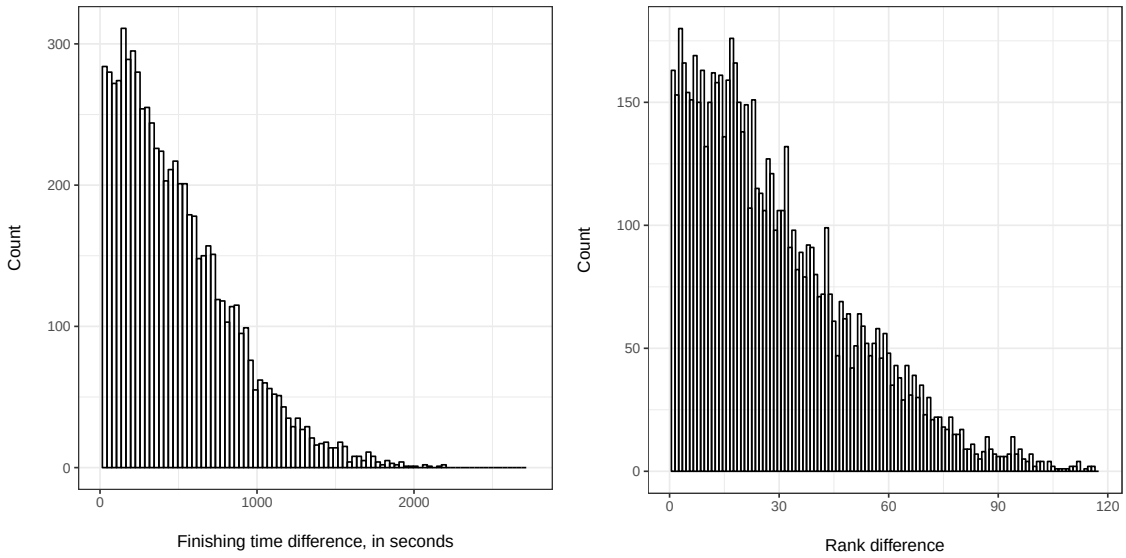
8. Repeat above steps 10,000 times.

From our simulations we can generate intervals which describe the extent of the variability in racer finishing times. In 95% of the simulations in which Anna Hahner completed the marathon, her finishing time was between 9,480 seconds and 10,703 seconds. This is consistent with Anna's observed finishing time, which was 9,932 seconds. In fact, our simulations estimate that Anna should have finished the Rio marathon in 10,092 seconds, which is slightly more than two and one-half minutes slower than her actual time. Lisa Hahner's corresponding interval is 9,427 to 10,619 with an observed finishing time of 9,933 seconds.

The bottom line here is that our simulated marathon finishes are consistent with the Hahner twins' finishing times. And note that the regression model underlying the simulations was estimated without the Hahner twins (and same for the Kim twins and Luik triplets). According to the our simulations, then, neither Hahner twin ran appreciably slowly, conditional on personal best times prior to the Rio Olympics.

With an eye on the matter of simultaneous finishing, Figure 4 contains two histograms based on our simulated race results. Figure 5a is a histogram which shows the distribution of absolute differences in Hahner twin finishing time where differences are grouped in 30 second bins; counts for the various bins are denoted by the vertical lengths of the bars. Figure 5b is similar but depicts the distribution of the absolute differences in Hahner twin rankings. Differences are grouped as single units ranging from no runner between the twins to nearly 120 runners between them.

Figure 4: Distribution of simulated results for the Hahner twins



(a) Simulated differences in finishing times

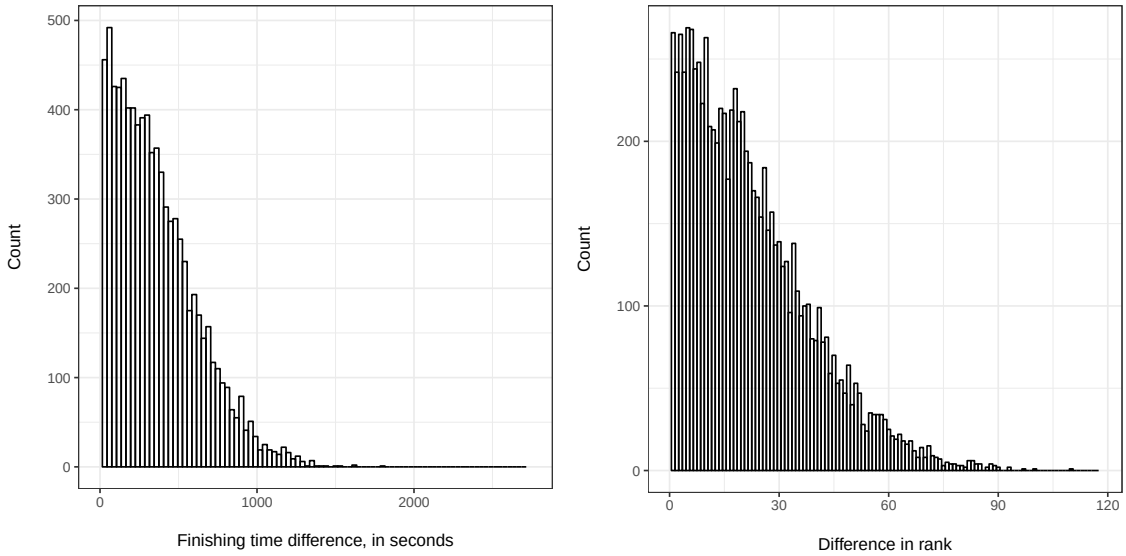
(b) Simulated differences in places

The results plotted in the histograms raise questions about the credibility of Anna and Lisa Hahner's story. In particular, the histogram suggest that a simultaneous finish would be very rare if Anna and Lisa had run independently of each other. For example, in fewer than 300 of the 10,000 simulated races did Anna and Lisa finished within 30 seconds of each other. Moreover, they finished in consecutive rank in fewer than 175 of the 10,000 races. Therefore, the close finish that was observed, where Anna and Lisa crossed the finish line one-after-the-other, would be highly unlikely if the two German twins had raced independently of each other.

Earlier we noted that personal best times might arguably be poor measures of runner abilities.

With this in mind, we repeated our simulations but replaced personal best times with half marathon splits. See Figure 5. While this change reduces the variation in the predicted outcome of each runner and therefore, reduces the expected distance between Anna and Lisa Hahner, even with half marathon split as a measure of ability it is still quite rare for Anna and Lisa to finish simultaneously or consecutively.

Figure 5: Distribution of simulated results for the Hahner twins based on half marathon split



(a) Simulated differences in finishing times

(b) Simulated differences in places

An important point regarding our simulations is the way that we handled DNFs. As our description above indicates, we assumed that DNF probabilities are the same for all runners and that the likelihood of a DNF is a not function of a runner's anticipated marathon finishing time. The rug marks in Figure 1a suggest that runners with better personal best times may be more likely to DNF than other runners, *ceteris paribus*. We suspect that this occurs because some better runners may expend excessive energy trying to achieve a good result in the marathon and in so doing injure or exhaust themselves; lesser runners, in contrast, may be content to finish respectably. Regardless of the validity of this conjecture, Figure 1a shows that the Hahner twins are representative of sort of runners who DNFed in Rio. Since our simulated Hahner statistics are conditioned on both

Hahner twins finishing, it follows that they are conservative. The fact that both women finished the marathon was notable in and of itself, and by discounting the possibility of a Hahner DNF we are giving the benefit of the doubt to the Hahner twins.

Conclusion

Further reading

- http://www.telegraph.co.uk/olympics/2016/08/17/german-twins-criticised-for-finishing-olympic-marathon-fun-run-h
- https://www.nytimes.com/2016/08/17/sports/olympics/twins-finish-marathon-hand-in-hand-but-their-country-says-they-crossed-a-line.html
- https://www.welt.de/sport/olympia/article157669264/Das-falsche-Laecheln-der-deutschen-Lauf-Zwillinge.html