

Memo

To: Scott Evans, editor, *CHANCE*
From: David Cottrell and Michael C. Herron, Dartmouth College
Date: February 24, 2017
Re: Revisions to women's marathon submission

This memorandum is a response to your February 16, 2017, email regarding our manuscript, "All in the family: German twin finishing times in the 2016 women's Olympic marathon." Thank you very much for inviting us to submit a revision of our manuscript to *CHANCE*, and this memorandum explains how we have updated it in light of comments from you and the Editorial Board.

- "Am not convinced of quadratic model. Two points in upper right may have influenced this. What is p -value for linear vs quadratic? What happens if those 2 points are removed? Could also simply plot differences from personal best as frequency dist'n. Could also plot projected from half vs actual."

The t -value for the quadratic term is -1.88 and the corresponding p -value is 0.062, which is not particularly compelling. We have accordingly dropped the quadratic terms in the scatter plots on p. 4 and, implicitly, from the Studentized residual plots on p. 6. In terms of the latter suggestion about, Figure 1b contains a plot of half marathon split times versus actual finishing times.

- "Marathon finish may be linearly related to the fastest previous marathon with perhaps age in a model. An interesting though experiment may be, if they were blindfolded, would they have finished together?"

We have added age to our simulation model, and the results are qualitatively equivalent. In addition, if one adds age to the linear specification depicted in our basic scatter plot (now without a quadratic term, as noted above), age is a significant predictor of marathon finishing time, even controlling for personal best ($t \approx 2.426$, $p \approx 0.017$). Older runners are slower, all things equal.

- “I printed this out in black and white. Would help see them if Hahnners were darkest color.”

We have changed the color scheme for the dots in our figures. The German Hahner twins are now shown in black, the North Korean Kim twins in red, and the Estonian Luik triplets in blue. Each color is present in its country’s respective flag.

- “a similar conclusion is reached?”

We changed the language in the paper so that it uses this phrasing.

- “pairs are not completely independent as same runner can be in different pairs. Not as though runners are drawn with replacement.”

We agree with the thrust of this comment, and our manuscript now emphasizes the lack of independence. Insofar as dependence in sampling typically diminishes variance, our lack of independence in the pair plots should make our results conservative. This means that that the observed Hahner and Kim compression in final results, conditional on prior best marathon times, is even more unusual.

- “due to subsequent DNF’s?”

We changed the language in the paper so that it uses this phrasing.

- “I don’t get this part. If Y_i is function of X_i , personal best, won’t estimates be same for each simulation? What I would try for an analysis would be look at distribution of finishing times, expecting it to be normal, as most human performance is fairly normal (most are average, some above, some below). I’d use the observed mean and variance of the finishers plus a random error, and see what this theoretical dist’n looks like. How often are finishers within x seconds of each other? Could also use historical marathon data from other races.”

We fully agree with this comment, and indeed the intuition offered by the referee was the motivation behind our simulation. To make this clear, we added a paragraph after the simulation steps, and this paragraph explains why finishing times (and order) will vary both within and across simulations. Then, the plots we generated (see Figures 5 and 6) describe how often the Hahner twins finish in close proximity of each other.

We have thought about historical marathon data but are hesitant to go down that route. The set of runners who competed in Rio is unique to this race, and marathon finishes are confounded by course and local conditions.

- There are errors in figure labels “5a” and “5b”

We apologize for this oversight and have fixed the associated labeling errors.

- “Would help to expand left side bins, as these are the finishes of main interest.”

We agree with this suggestion. Accordingly, in the histograms that describe simulation results, we have highlighted left side bins in red and annotated the plots with descriptions of the sizes of key areas. This emphasizes the finishes that are most important for our results.

- “what are interval widths in seconds? Hard to see.”

We agree that this is hard to see. The text now points out that bin width is 30 seconds.

- “Suddenly Latin?”

We have changed the Latin text to, all things equal.

- “likely”

We apologize for the confusion here, but we are not sure what the referee is requesting. In the marked-up PDF document, the word “likely” appears over the following sentence: “Their finishone second between the two womenwas an extremely low probability event.” We are not sure how to incorporate the referee’s request but are happy to do so if given more information.