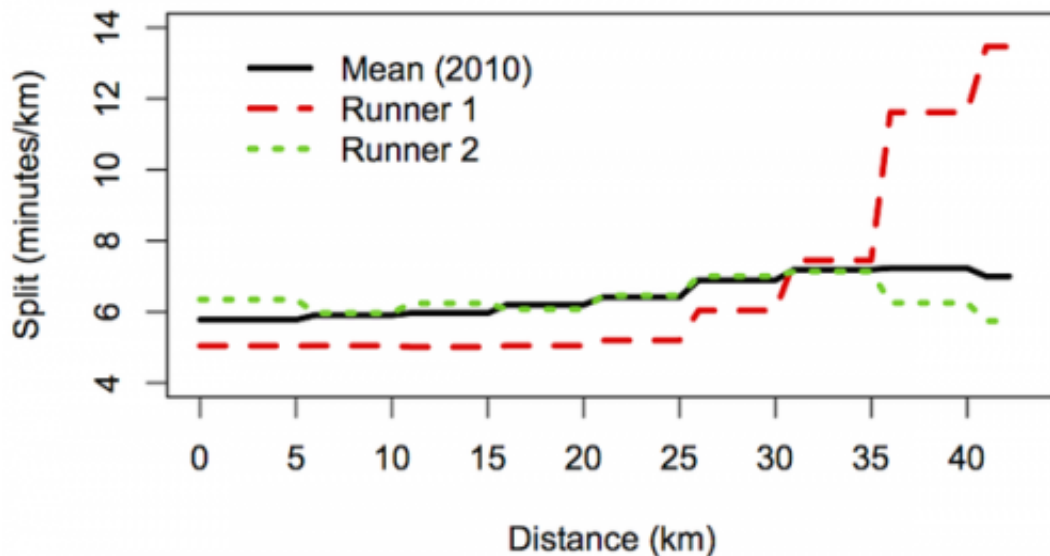


Researchers help Boston Marathon organizers plan for 2014 race

April 15 2014



2010 Split profile comparisons. Credit: *PLoS ONE* 9(4): e93800. doi:10.1371/journal.pone.0093800

After experiencing a tragic and truncated end to the 2013 Boston Marathon, race organizers were faced not only with grief but with hundreds of administrative decisions, including plans for the 2014 race – an event beloved by Bostonians and people around the world.

One of the issues they faced was what to do about the nearly 6,000 [runners](#) who were unable to complete the 2013 race. The Boston Athletic Association, the event's organizers, quickly pledged to provide official

finish times for these runners. Thinking ahead, they also had to consider how to provide these runners with an opportunity to qualify for the 2014 race.

To seek advice on these issues, they contacted Richard Smith, a statistician and marathon runner at the University of North Carolina at Chapel Hill, and director of the Statistical and Applied Mathematics Sciences Institute (SAMSI) based in Research Triangle Park, N.C. They asked Smith to come up with a statistical procedure for predicting each runner's likely finish time based on their pace up to the last checkpoint before they had to stop.

"Once I got their email," said Smith, "of course I knew I had to help them." Smith already knew the organizers, as a result of a previous occasion when he provided advice related to the event's qualifying times.

Smith quickly assembled a team of fellow analysts that included Francesca Dominici and Giovanni Parmigiani at Harvard School of Public Health, and Dorit Hammerling, postdoctoral fellow at SAMSI, who were in the 2013 race and finished uninjured. The team also included Matthew Cefalu, Harvard School of Public Health; Jessi Cisewski, Carnegie Mellon University and Charles Paulson, Puffinware LLC. The results, and the method the researchers developed, were published in the April 11 edition of *PLOS ONE*.

With the help of the Boston Athletic Association, the researchers created a dataset consisting of all the runners in the 2013 race who reached the halfway point but failed to finish, and all the runners from the 2010 and 2011 Boston marathons. The data consist of "split times" from each of the 5 km sections of the course (from the start up to 40 km), and the final 2.2 km. The research team was tasked to predict the missing split times for the runners who failed to finish in 2013.

The researchers adapted techniques used in such contexts as computing missing data in DNA microarray experiments and estimating ratings which Netflix subscribers would have given to movies they had not seen. They proposed five prediction methods and created a validation dataset to measure the runners' performance by mean squared error and other measures. Of the five, the method that worked best used local regression based on a K-nearest-neighbors algorithm (KNN method), though several other methods produced results of similar quality.

The KNN method looks at each of the runners who did not complete the race (DNF) and finds a set of comparison runners who finished the race in 2010 and 2011 whose split times were similar to the DNF runner up to the point where he or she left the race. These runners are called "nearest neighbors."

"We had to come up with a method to compare the runners based on the split points up to a certain point of the race and then had to decide how many of the nearest neighbors to examine in order to develop a prediction for the DNF runner that would be based on the different finishing times of these nearest neighbors," said Smith, who has run the Boston Marathon in the past and will run this year's race. "We decided to choose 200 nearest neighbors. We also tried 100 and 300 nearest neighbors, but the results changed only slightly and didn't make them better."

The Boston Athletic Association decided to grant entry to the 2014 race to anyone who was stopped from completing the 2013 event, so they will have a chance to complete the Boston Marathon after all. But in the course of developing the method, Smith and his colleagues realized there were other uses for the technique.

"We have found that using the KNN method looking at a runner's intermediate split-time will also be useful in predicting the person's

completion time while the [race](#) is in progress," said Smith. "This can be helpful for relatives and friends to be able to meet the person at the finish line."

More information: Hammerling D, Cefalu M, Cisewski J, Dominici F, Parmigiani G, et al. (2014) Completing the Results of the 2013 Boston Marathon. *PLoS ONE* 9(4): e93800. [DOI: 10.1371/journal.pone.0093800](#)

Provided by University of North Carolina at Chapel Hill

Citation: Researchers help Boston Marathon organizers plan for 2014 race (2014, April 15) retrieved 27 April 2024 from <https://phys.org/news/2014-04-boston-marathon.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.