**A Deeper Dive into the CrossFit 2019 Season**

Anna Ballou

*Bios 669*

Background

Much like most conventional professional athletes, CrossFit athletes train year-round in hopes of competing at the highest level of competition, The CrossFit Games. Akin to the Superbowl or the World Series, the CrossFit Games (informally known as “The Games”) is intended to feature the most elite CrossFit athletes of that season. Throughout its 14-year tenure, The Games has evolved significantly through changes in leadership, location, and protocol. More recently, with more and more players competing at a high level, new pathways to qualify for the games had to be developed. In 2019, the CrossFit leadership instituted a dramatic shift to The Games qualification process. To begin the season, all (100,000+) athletes would compete in an online competition called “The Open.” The Open consisted of five events, with the top 20 athletes in each division qualifying directly for The CrossFit Games. Following The Open, athletes who did not qualify could compete in one of many independently organized CrossFit “Sanctional” tournaments at some point during the season. An additional 100 athletes received an invitation to The Games from one of these events. Importantly, because of this dual qualification method, the 2019 CrossFit games featured 141 male and 129 female athletes, a much larger field than previous years.

This exceptionally large field of athletes forced the directors of The Games to institute several cuts to the field very early on in the competition. More specifically, over 100 athletes were cut over the course of the first half of The Games, with only 10% of the field remaining for the final six events. Athletes, viewers, and analysists were left questioning whether the cut system left an accurate reflection of the sport’s best. This was an especially charged topic since the top finishers received sizable monetary prizes. Therefore, this project attempted to use publicly available data about the 2019 CrossFit Open and Games to assess whether the 2019 CrossFit Games truly crowned the fittest men and women in the sport. More specifically, in using The Open as a benchmark for athletes’ fitness, this project assessed whether those standing on the podium were an accurate reflection of the sport’s best.

Data and Methods

Data from the 2019 CrossFit Open and Games was directly downloaded from the data science hub, *Kaggle*. The data repository consisted of several distinct datasets on athlete characteristics, event scores, and open rankings. The datasets used in this analysis included (1) athletes at The Open, (2) athletes at The Games, and (3) The Games scores. Because over 100,000 individuals competed in the 2019 CrossFit Open, all datasets were subsetted to include only the top 100 ranked athletes in the male and female (non-scaled) divisions using *R* and *The Slurm Workload Manager*. The *R* and *Slurm* files can be found in the zip file. A supplemental file that included event names, numbers, and descriptions was generated using information from the CrossFit Games website. All datafiles were read into *SAS On Demand.* Files were joined using three key identifiers: (1) competitor id, (2) event number, (3) and event ordinal.

**First**[[1]](#footnote-1), overall characteristics of athletes at The Open and The Games were examined, including the average age, weight (lbs.), and height (ft.) of the athletes, and the best finish by each represented country. **Second**, a macro was generated to examine the description of each event at The Games to classify it as either a strength-based, cardiovascular-based, or mixed -style event. The macro (*classify\_events*) scanned each event description for key terms associated with either strength-based or cardiovascular-based exercises. For example, all events that included one of the major CrossFit lifts (i.e. squat, clean, snatch, deadlift, etc.) were flagged as a potential strength event, while events that featured phrases such as “run,” “sprint,” “bike,” and “row” were flagged as potential cardiovascular events. If, at the end of the scanning, only strength terms were flagged in an event description, then that event would be classified as a strength event. Similarly, if only cardiovascular terms were flagged, then that event would be classified as a cardiovascular event. Those event descriptions that included both strength and cardiovascular terms (or neither) were classified as a mixed event type. This macro — which takes an input dataset with an event description column — could easily be extended to include more terms as the sport evolves. Because strength- and cardiovascular-based events tend to favor different types of athletes, this macro was used to examine whether the event types were balanced throughout The Games, especially during the first six events with the most cuts. An analysis dataset including variables for the newly generate event type, an indicator for whether or not an event occurred during the first half of The Games, and the number of athletes cut after each event was created to analyze whether or not a certain event type was disproportionately represented when most athletes were being cut.

The datasets were then separated into the male and female specific divisions. The **third** step of the analysis included generating a formal report of the top 10 male and female finishers in The Open. The division-specific reports included the athlete’s rank in The Open, their name, country of origin, age (in years), height (in ft.), and weight (in lbs.). **Fourth**, a report that examined how the top ten Games finishers ranked in The Open was generated. The report included the competitor’s name, their Open rank, and two Games summary measures (Games finish and mean event finish). Rows that featured athletes who finished in the top ten in both the Games and Open were highlighted.

Finally, the **fifth** section of the analysis included creating a macro that compared the event-by-event performances of two athletes head-to-head. The macro (*head\_2\_head)* allowed the user to pass in two athlete names and then generated a line plot that depicted each athlete’s placement in each event throughout the course of The Games. The generated plots aided in examining visually how and when athletes were cut as well as any trends between similar or dissimilar athletes.

Results

Both male and female Open athletes had an average age of 27. Men were slightly shorter than the typical U.S. male (mean height of 5.8ft.), while women were about average height (mean height of 5.4ft.). The average weight for male and female athletes was 192lbs., and 142lbs., respectively. 18 countries were represented in both the male and female divisions. Iceland and Australia both saw athletes finish in the top four, while Canada and the U.S. were both consistently in the top half of each division.

The *classifying\_events* macro identified five cardiovascular-based events, four strength-based events and three mixed-style events at the 2019 CrossFit Games. Further, the first three events (which resulted in the cutting of 70 athletes) were either solely cardiovascular or included a cardiovascular component. This suggests that strength-based athletes may have been disproportionately cut during the earlier stages of the Games.

The top 10 male finishers in The Open included athletes from five different countries, aged 21 to 31, all under six feet, and mostly under 200 pounds. Importantly, six of the ten athletes had competed at The Games in prior years, and the number one ranked athlete (Mathew Fraser) was the three-time defending champion. In the female division, athletes represented six different countries, were aged 24-33, had heights ranging from 5.3-5.6 feet, and were all under 170 pounds. Nine of the 10 athletes had significant competitive CrossFit experience and two former multi-time winners (Annie Thorisdottir and Tia-Clair Toomey) were featured in the top 10.

In the male division, only three of the top ten Games finishers also finished in the top 10 in The Open. Further, several of the top ten Games finishers ranked well outside the top 50 in The Open, including the second place winner, Noah Ohlsen. Notably, the fifth place Games athlete is missing from the report. This is because he ranked outside the top 100 in The Open, so he was not even included in the initial analysis dataset. In the female division, a more sizable proportion of the top ten Games athletes (five of the ten) also finished in the top ten in The Open. However, as is seen in the Male division, the ninth place female finisher was not included in the Report because she ranked outside of the top 100 in The Open[[2]](#footnote-2). It is important to note, however, that in both the male and female divisions, the overall champion did finish in the top ten in both The Open and The Games, indicating some level of consistency between the two stages.

The *head\_2\_head* macro was first used to compare the performances of the top two male and female athletes (Mathew Fraser vs. Noah Ohlsen, and Tia-Clair Toomey vs. Kristin Holte, respectively). Both plots look relative similar, with sets of both athletes performing well throughout the entire competition. It is important to note that three of the four athletes had bottom-half finishes at least once during the first 6 cut-worthy events, indicating that there was some room for athletes to have a bad event and not risk getting immediately cut from the competition. The *head\_2\_head* macro was then used to compare an athlete who finished inside the top ten at The Games, but outside of the top at The Open versus an athlete who finished inside the top ten in The Open but was cut during The Games. Interestingly, both athletes that were cut had strong finishes in prior games and are both known as two of the stronger athletes in their respective fields. However, both athletes that made it passed the cuts had one sub-par performance, but were able to recover with top 20 finishes in other events.

Conclusions

Despite several controversial and suggestive findings, the analyses of this project in total supported the idea that the 2019 CrossFit Games did, in fact, crown the fittest men and women in the Sport. Two athletes (one per division) who finished in the top ten at The Games failed to finish inside the top 100 in The Open. However, it is important to note that athletes tend to train through The Open and intend to reach their peak fitness months later (at The Games). Therefore, athletes ranking in the top 100-200 (which was the case for the athletes in question), could reasonably improve their fitness enough to reach top-ten Games potential. The findings from the *classifying\_event* macro suggested that strength-based athletes may have been disproportionately cut during the first half of the competition. However, the best CrossFit athletes (i.e. those who should place in the top ten at The Games) are typically well-rounded, and would perform decently in both high-end strength and cardiovascular events. Additionally, the *head\_to\_head* macro demonstrated that athletes could have one sub-par finish as long as they had consistent performances in other events. Finally, the overall Games winners (and most of the podium of both divisions) were largely unaffected by the cuts. Both prior champions, Mathew Fraser and Tia-Clair Toomey successfully defended their titles, and both second and third place female winners were in the top ten in both The Games and The Open. Therefore, despite it being a season of controversy and skepticism, the 2019 CrossFit Games was largely a success.

1. Note: the order described in the methods doesn’t not perfectly match the order of code in the SAS file. [↑](#footnote-ref-1)
2. Also not pictured, the 8th place Games Finisher was later disqualified. [↑](#footnote-ref-2)