This past July we had a special opportunity. The 88 fittest men and women in the world competed against each other at the 2010 CrossFit Games and were ranked from first to last. Furthermore, they spilled their guts about their benchmarks and PRs, self-reporting everything from Fran times to 1RM deadlifts. Wondering if we could learn from that data, I set out to answer a simple question: Can you predict if someone will place well at the CrossFit Games? For example, did the top athletes all share some common characteristic? Did they have fast Helen times? Were they taller than average? How important was age?

The answer is rather complex, but I like to start with no-bullshit simple answers and add detail from there. First, age, height, and weight don’t really matter. Second, strength to body weight ratio is the most powerful predictor of success. Finally, endurance was a startling predictor for both men and women, but in vastly different ways. If you’d like more detail, then read on, but I’ll warn you in advance that things get pretty geeky from this point forward.

# Before We Begin

Everyone loves to know averages, so let’s start with averages for the 2010 CrossFit Games Competitors:

## Women

Forty-three women competed. The average female competitor was 29 years old, weighed 138 pounds, and was a little over 5’4” tall. Her Fran time was 4:02. She could clean and jerk 158 pounds and perform 32 consecutive pullups. Her Fight Gone Bad score was 344. Her 5K run time was 22:20.

## Men

Forty-five men competed. The average male competitor was 28 years old, weighed 188 pounds, and was almost 5’11” tall. His Fran time was 2:42. He could clean and jerk 272 pounds and perform 61 consecutive pullups. His Fight Gone Bad score was 409. His 5K run time was 19:46.

# Assumptions

My analysis relies on a few assumptions. First, all athlete self-reported data is assumed to be accurate. Next, the only data considered from the 2010 CrossFit Games was final placement. No data was considered from the nine events that comprised the Games, but only the fact that Kristan Clever finished first, Annie Thorisdottir finished second, and so on.

# Limitations

First, I am not a statistician. However, I have a working knowledge of some useful statistical methods and I am an active CrossFitter and affiliate owner. I think this perspective allows me to form some conclusions that would be missed by the either the Ph.D. statistician or the CrossFitter who doesn’t have a hard-on for math.

Next, this analysis is technically limited to predicting competitor placement in the 2010 CrossFit Games. Since we already know final placement for the 2010 Games, that’s rather worthless. However, if we know our bounds then we can apply what we learn here to other arenas. No CrossFit competition will ever mimic the 2010 Games exactly, but if we assume the 2010 Games were a thorough test of fitness by all its definitions, then other thorough tests of fitness might be similarly predictable.

Finally, all athlete self-reported PRs were not complete, leaving a few holes in the data.

# Method

I first analyzed the data while keeping male and female data separate and then repeated the analysis with all athletes combined. The centerpiece of this is a correlation analysis. Correlation tells you whether two things are related. In this case, one of the things examined is final placement in the 2010 CrossFit Games. The other thing examined was every single self-reported metric: body weight, age, height, Grace time, 1RM snatch, Filthy Fifty time, etc. A correlation analysis looks at all those things and tells you how much they affected the first thing: placement in the 2010 CrossFit Games. Did a faster Fran strongly correlate to a better finish? Did a faster Fran mean a worse finish? Did Fran time matter at all, or was something else more important? These are the questions correlation helps you answer. The variables examined for correlation to final placement are listed below. You’ll recognize them as all the self-reported data from the 2010 CrossFit Games site.

* Age
* Height
* Weight
* Fran
* Helen
* Grace
* Filthy Fifty
* Fight Gone Bad
* Clean and Jerk 1RM
* Snatch 1RM
* Deadlift 1RM
* Back Squat 1RM
* Pullups, max effort
* 400m Run
* 5K Run

There’s also a twist: you can make up your own variables. Wait, it’s more legit than it sounds. If it makes sense to take some of the existing data and add it up, divide it by something, etc, then you can see if it correlates to placement. I did this with a few of the self-reported data. I examined the below combinations of data and assigned them names for convenience:

* Strength: the sum of Clean and Jerk, Snatch, Deadlift, and Back Squat
* Strength to Body Weight Ratio: Strength divided by the athlete’s Weight
* Sprint: the sum of Fran, Helen, and Grace
* Sprint to Body Weight Ratio: Sprint divided by the athlete’s Weight
* Endurance: the sum of 5K Run and Filthy Fifty
* Endurance to Body Weight Ratio: Endurance divided by the athlete’s Weight

The above combinations aren’t magic. I devised them using my own judgment. Someone else might choose different combinations or perform different operations on the data. My intent was to group variables that shared a common time domain or metabolic pathway. I examined body weight ratios to see if normalizing the results by general size would prove relevant. These variables would help reveal whether the 2010 CrossFit Games favored specialists in any particular time domain, metabolic pathway, skill, body size, or body composition.

# Results

## Women

First, let’s look at demographic information. Height and weight were essentially uncorrelated to finishing well. Older athletes tended to place worse, but the correlation is fairly weak.

The variables that most closely predicted final placement were

1. Endurance
2. Strength to Body Weight Ratio
3. Filthy Fifty
4. 5K Run
5. Clean and Jerk

It’s not very descriptive to count Filthy Fifty and 5K Run again, since they together comprise Endurance. Similarly, Clean and Jerk is part of Strength to Body Weight Ratio. Therefore, let’s simplify the list to the top two, which also showed significantly more correlation than #4 and #5:

1. Endurance
2. Strength to Body Weight Ratio

A seat of the pants look at the data confirms these correlations.

* The five fastest 5K Run times belong to ladies in the top 16 finishers.
* Four of the five fastest Filthy Fifty times belong to ladies in the top 15.
* Most strikingly, four of the five top Stength to Body Weight Ratios belong to ladies in the top 11.
* The highest Strength to Body Weight Ratio was Kristan Clever at 7.19, 24% higher than the average 2010 CrossFit Games female competitor.

## Men

Age and weight were essentially uncorrelated to finishing well. Taller athletes actually tended to place worse, but the correlation is fairly weak.

The variables that most closely predicted final placement were

1. Clean and Jerk
2. Grace
3. Sprint
4. Fran
5. Strength
6. Strength to Body Weight Ratio

Correlation was much fuzzier for the men than the ladies. The ladies’ data showed three clear variables that correlated more than any others. The men’s data, however, is much more closely grouped. Clean and Jerk (#1) differs from Strength to Body Weight Ratio (#6) by only about 10%, depending on how you measure the difference. The important point is that all of the above factors carry similar importance in predicting final placement for the male competitors.

So let’s perform the same exercise as with the ladies and whittle these variables down to the most important. Grace and Fran are both part of Sprint, so let’s keep Sprint and axe the other two. Clean and Jerk is part of Strength and Strength is part of Strength to Body Weight Ratio, so let’s keep Strength to Body Weight Ratio. That’s also a valid move because Strength to Body Weight Ratio was an important predictor for the ladies, leading me to believe it might be important for the men as well. That leaves us with the following:

1. Sprint
2. Strength to Body Weight Ratio

Once again, we can glance at the data and confirm these conclusions. Remembering that Sprint is composed of the sum of Grace, Helen, and Fran, consider that

* The five fastest Grace times all finished in the top 11
* Conversely, the five slowest Grace times placed no better than 27th
* While a fast Fran time didn’t ensure a top finish, a slow Fran time ensured a poor finish: The five slowest Fran times placed no better than 32nd
* Four of the five highest Strength to Body Weight Ratios placed in the top 15.

Additionally, the men showed one other rather bizarre correlation: A faster 5K Run correlated to finishing *worse*. This correlation isn’t nearly as strong as the correlations to placing well, but it exists nonetheless. For example

* Four of the five slowest 5K times placed in the top 18.
* Four of the five fastest 5K times placed no better than 19th.
* The only top 18 competitor with one of the fastest 5K times was Mikko Salo.

## Combined

Combining the men and women didn’t yield any spectacular results. No self-reported data showed correlations as strong as the gender separated data. However, the single strongest correlation in the integrated data was Strength to Body Weight Ratio, which further supports its importance as a predictor of success.

Similar to the gender segregated data, weight was uncorrelated to success. As with the ladies, older athletes placed worse, but the correlation was somewhat weak. As with the men, taller athletes also placed worse, again with weak correlation compared to the segregated data. In short, integrating the data turned it into a relatively incomprehensible mess.

# Conclusions

What conclusions can we draw from the raw data above?

First, height and weight don’t really matter. Sure, the guys and gals at either extreme end of the spectrum may face a disadvantage, but the top finishers propel themselves to the top with bodies of all shapes and sizes.

Second, age matters, but not very much. As anyone might expect, it’s difficult for older athletes to keep up with younger ones. No surprise there. But older (30+) athletes certainly don’t face an insurmountable challenge simple because of age.

Next, men and women are different--even beyond basic anatomy. As far as the 2010 CrossFit Games are concerned, the two genders sometimes showed different predictors of success. For example, Endurance was the #1 predictor of a top finisher among women, but a sure predictor of a poor finish for men.

Strength to Body Weight Ratio was an important predictor for both men and women. The data reveal that the more strength you can pack into a compact frame, the more likely you are to place well. This ratio was especially important for women. And since most women at the games weighed approximately 125-150 lbs, a far narrower range than the men, this means that strength is paramount for the CrossFitting woman.

Capability in short metcons like Grace, Fran, and even Helen is vitally important to men. This may be partially due to the inherent nature of CrossFit competitions. Even before performing this analysis I had observed that it’s just logistically difficult to run a workout at a CrossFit competition that requires more than 15 minutes to complete. If a large number of events is desired, many events must be 3-10 minutes in duration. Does proficiency at short metcons lead to success by nature of competition programming? Yes, I think that plays a part.

However, the strong Endurance correlation from the women says otherwise. Further compounding the problem is the fact that a fast 5K Run was directly correlated to finishing poorly among men. Given all this information, one of the following must be true

* The correlations aren’t very relevant at all, and this analysis was just an inconclusive mental exercise.
* Endurance events were important predictors for women while Sprint events were important predictors for men.
* More data is required to see the full picture.

Tony Budding, CrossFit Media Director, commented, “…remember that this is a new sport, and the women's competition is less mature than the men's. As the sport matures, we should expect to see fewer aberrations in athlete stats. And we have very small data for any of it (relative to what we will have going forward for sure). Differences like Endurance as a positive or negative predictor will most likely work themselves out.” I tend to agree.

What can potential competitors learn from this to increase their odds of winning the 2011 CrossFit Games? Well, we have no guarantee the 2011 CrossFit Games will be anything like the 2010 CrossFit Games, so any advice is just an educated guess. So at the risk of being reckless, here’s the loosely supported advice that anyone looking to cover their ass and never be wrong would never tell you, because it might be full of holes:

 Being strong and lean is a recipe for success for everyone. Women must supplement that strength with strong capability in stamina and the oxidative pathway in general. Men must supplement that strength with strong capability in the glycolytic pathway. Another simple tactic for success is to have completed most benchmark workouts and know your PRs. Not surprisingly, competitors that left significant blanks in their self-reported data tended to place poorly. This was also manifested as many competitors suffered from inexperience with high-skill movements like pistols, ring handstand pushups, and rope climbs that are part of many benchmark WODs. Beyond that, work capacity across broad time and modal domains appears to be the key, as it has always been…and as it should be.

Both the raw data and the Excel spreadsheet I used for the analysis are available from the article summary on the 2010 CrossFit Games website. Feel free to play with the numbers and tell me why I’ve got it completely wrong.