

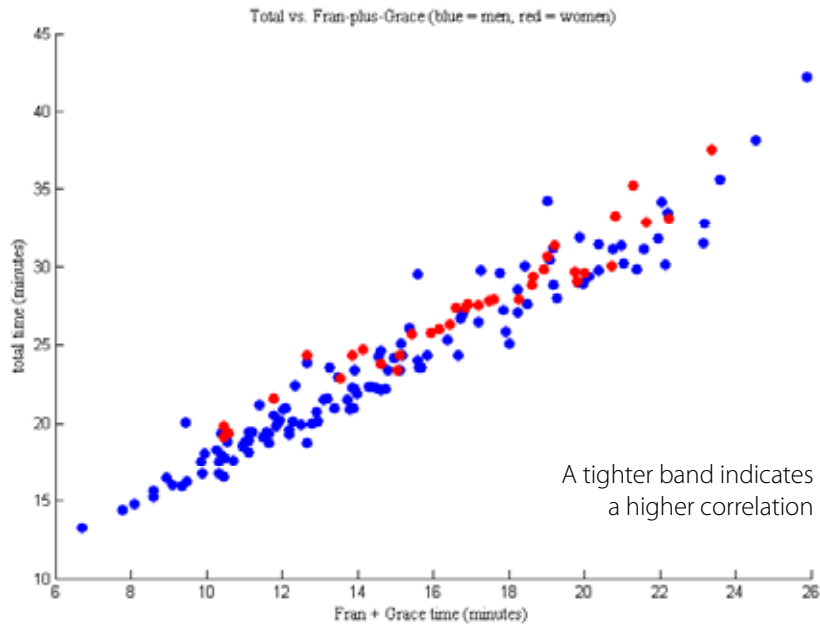
A Statistical Analysis of the CrossFit Games

Robert Novy-Marx



It is no exaggeration to say that the 2008 CrossFit Games tested who could do the fastest high-pull Fran and heavy Grace. This is somewhat surprising, as the deadlift and run events constituted roughly 40% of the average competitor's total time. Nevertheless, these events were largely irrelevant with respect to the Games' final outcome. Fran-plus-Grace times alone explain 94% of the variation in overall times, for both men and women, and ranking the athletes on the basis of Fran and Grace alone would have yielded outcomes very similar to the actual results.

There are two reasons the deadlift and run events mattered so little in the overall results. First, while these events made up almost 40% of the average competitor's final time, they contributed substantially less than 40% of the variation to these times. Simply put, while the difference in Grace times between the top and bottom performers is enormous, the best runners are just not that much faster than the worst runners, even over a course that takes as long on average as Grace (and in the '08 Games it didn't). For example, on the men's side each place in the run was worth on average only 1.2 seconds, and each place in the deadlift about two, compared with more than three seconds for each place in the thrusters and over five seconds in the clean and jerk. That is, a place in Grace was worth, on average, more than four times as much as a place in the run.



The second reason is simply that Fran and Grace explain much of the variation in competitors' deadlift and run times: 42% on the men's side and 31% on the women's. The better athletes, and the ones that have been CrossFitting longer, are not only faster at Fran and Grace, but also tend to be better at deadlifts, burpees, and running.

The thruster and clean and jerk events alone consequently do a tremendous job of predicting the Games' final outcome. The rank correlation of competitors' Fran-plus-Grace times and their total times is over 97%, for both men and women. Fran-plus-Grace alone accurately predicts the top five places on the men's side. It also predicts the top four women, though it switches second and third places, because Tanya Wagner is a dominant runner, but marginally slower (seven seconds) than Gillian Mounsey in Fran-plus-Grace.

Good at one thing, good at everything?

The degree of correlation among times across the four events is surprisingly high. On the men's side, thruster times are 60% correlated with deadlift times, 52% with run times, and 51% with clean and jerk times. Deadlift times are 49% correlated with run times, and 62% with clean and jerk times. Clean and jerk times are 15% correlated with run times. These correlations are all quite high, with the

exception of the run/clean and jerk correlation, reflecting the facts that 1) the better athletes, and the ones that have been CrossFitting longer, tend to be better at everything, and 2) the run favors light athletes while Grace favors bigger athletes.

The correlations are much weaker, however, among the top competitors, arbitrarily defined here as the Games' top 20 finishers. Among these top competitors only the times in the thruster and deadlift events are highly correlated (52%). Times in the thruster and run events are moderately correlated (32%), and times in the clean and jerk event somewhat negatively with times in both the thruster and run events (-15% and -20%, respectively). This correlation structure may reflect the fact that the thruster, deadlift, and run events are basically metcon events for the elite men, while the clean and jerk event favored raw power even among the top athletes.

The correlations on the women's side are weaker, although positive between all events. Thruster times are 28% correlated with deadlift times, 17% with run times, and 31% with clean and jerk times. Deadlift times are 32% correlated with run times and 67% with clean and jerk times. Clean and jerk times are 22% correlated with run times. These relatively weaker correlations suggest that the different events—with the obvious exception of the

deadlift and clean and jerk—are “more different” for the women than for the men. This is hardly surprising. The women’s thruster event tests relatively more for upper body strength and relatively less for lactic tolerance than the men’s, and is consequently more dissimilar to both the deadlift event and the run.

Again, correlations are weaker among the top competitors, defined for the women as the Games’ top 10 finishers. Among these top competitors only the times in the run and deadlift event are highly correlated (42%). Times in the run and thruster events are moderately correlated (25%), while times in the thruster and deadlift events are negatively correlated (-16%).

Cumulative time versus event ranking

While CrossFit, like life, isn’t fair (see “All Other Things Being Equal, [The CrossFit Fairness Doctrine](#),” CrossFit Journal Issue 72 - August 2008), and I am not advocating any particular scoring system, it is interesting to consider how the results could have been different.

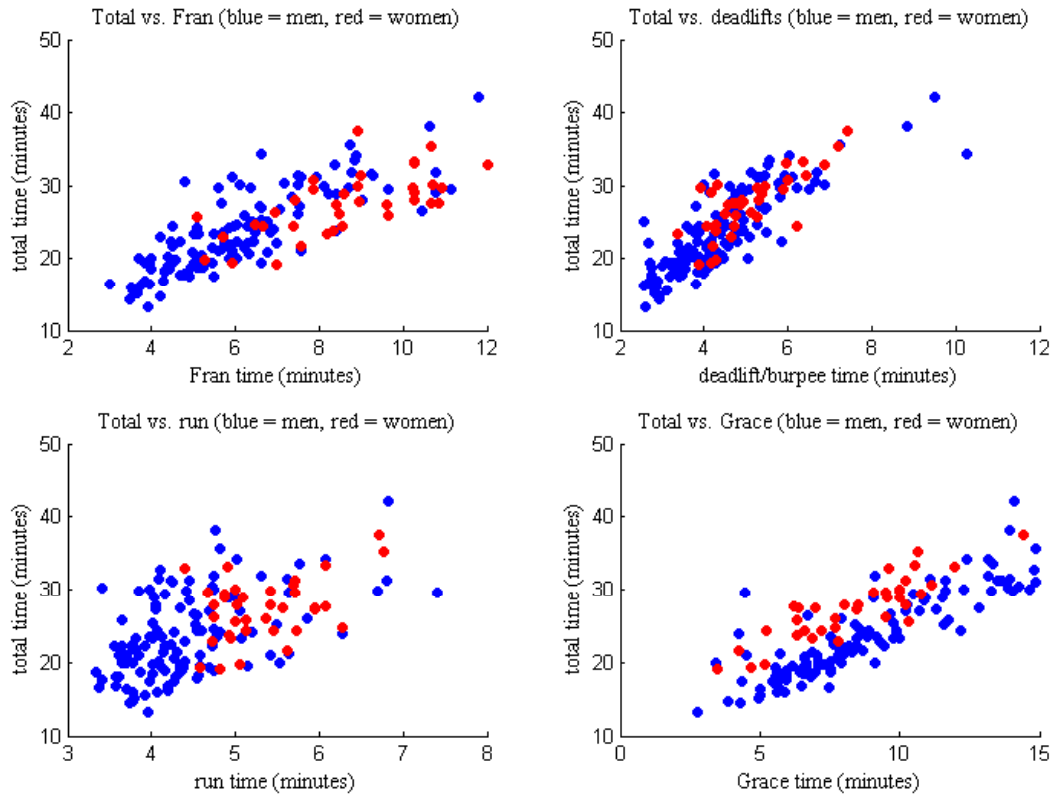
If the ‘08 Games had been scored using the event place system employed in ‘07 (1st = 100, 2nd = 95, 3rd = 90, 4th = 89, 5th = 88...), the big winners on the men’s side would have been Josh Everett (who ends up champion under almost any reasonable scoring system based on event rankings), Speal (who would have finished 3rd instead of 10th) and OPT (7th instead of 12th). The big losers would have been Jason Khalipa (slips to 2nd), Jeremy Thiel (5th instead of 3rd) and Jeff Tincher (8th instead of 5th). The “big winners” tend to be stronger runners, an event that essentially went unrewarded under the ‘08 scoring system, while the “big losers” tend to excel at the clean and jerk, which was vitally important in the ‘08 scoring.

On the women’s side the ‘07 scoring system “does better,” yielding the actual top three finishers from the ‘08 Games. The big winners would have been Libby Dibiase (4th instead of 6th) and Cherie Chan (9th instead of 15th). The big losers would have been Tamara Holmes (6th instead of 4th), Amber Wiggs (13th instead of 8rd) and Robin Lyons (14th instead of 9th). Again, the ‘07 scoring system is relatively favorable to runners (and Ms. Dibiase, who won the deadlift event), and relatively less favorable to the Olympic lifters.

In neither case does the ‘07 scoring do as well predicting the actual Games outcomes as does Fran-plus-Grace alone. That is, simply dropping the deadlift and run events from the total time has less of an impact on the final places than scoring the Games on the basis of event placing. The ‘07 scoring system puts substantial weight on the outcome of the deadlift event than on the run, which was largely irrelevant to the actual Games outcomes.

This analysis gives some guidance on how one could construct a multi-event competition, scored using cumulative time, in which every event “matters.” It would require designing the events in a manner that roughly equates the standard deviations, not the averages, of times across events. This generally demands that strength-limited events are shorter than endurance-limited events. Events that some athletes struggle to just complete, like the relatively heavy clean and jerk in this year’s Grace, generate far more time dispersion per average duration (i.e., standard deviation relative to the mean) than events that are easy when you’re not “on the clock,” like the run.





A hypothetical “balanced” 2008 Games

Holding Fran constant, a roughly balanced schedule of events includes seven rounds (not five) of deadlift/burpees, an 1800m (not 750m) hill run, and 20 (not 30) clean and jerks. Average time for Fran on the men’s side was just over six minutes. Scaling of the other events results in an increase in the average deadlift event time from just under four and a half minutes to just over six, an increase in the average run time from just under four and a half minutes to almost eleven, and a decrease in the average Grace time from just over eight and a half minutes to five and a half. While the run in this hypothetical version of the Games takes almost twice as long, on average, as each of the other events, it actually remains the least important event for determining the variation in competitors’ total times.

In this balanced version of the 2008 CrossFit Games, Jason Khalipa’s dominant performance in Grace remains sufficient to overcome Josh Everett’s more balanced performance, leaving the top two places unchanged from the actual 2008 Games. The margin between first and second is, however, much smaller—22 seconds instead

of 68. Speal’s strong run is rewarded, and the punishment for his relatively poor Grace performance is less brutal, catapulting him all the way up to tied-for-third with Jeremy Thiel. This pushes Pat Barber from fourth back to fifth. Rick Frausto retains sixth place by jumping Jeff Tincher (who is disproportionately strong at Fran and Grace), who falls from fifth to seventh. OPT (the other big winner than Speal from the balanced format, and also a relatively strong runner and weak Olympic lifter) jumps four places to eight, pushing Dutch Lowy, Matt Chan, and Lucas Brainerd back to ninth, tenth, and eleventh, respectively.

On the women’s side, Tanya Wagner jumps ahead of Caity Matter to take the title. In these balanced Games Ms. Wagner’s superior running effectively neutralizes Ms. Matter’s advantage in Grace, while her strong Fran more than offsets her disadvantage in the deadlifts. Gillian Mounsey retains third, even though the balanced format is disadvantageous to her, as her strengths are Fran and Grace. Jolie Gentry (a strong runner) and Libby Dibiase (the deadlift event winner) both jump a spot, to fourth and fifth

ORDINARY LEAST SQUARES REGRESSION RESULTS:
EXPLAINING THE TOTALS WITH INDIVIDUAL WORKOUTS
AND PAIRS OF WORKOUTS

explanatory workout	men's results			women's results		
	intercept	slope	adjusted-R ²	intercept	slope	adjusted-R ²
panel A: single workouts as the explanatory variable						
thrusters	8:39 [8.6]	2.44 [15.5]	66.9%	13:16 [5.0]	1.64 [5.4]	44.2%
deadlift	8:18 [8.6]	3.48 [16.4]	69.4%	10:30 [4.2]	3.33 [6.9]	57.4%
run	7:01 [2.7]	3.73 [6.4]	25.1%	11:46 [1.9]	2.93 [2.6]	13.9%
cj	9:18 [11.6]	1.65 [18.8]	74.9%	14:19 [11.0]	1.56 [10.5]	75.6%
panel B: workout pairs as the explanatory variable						
thrusters + run	3:20 [2.4]	1.93 [15.1]	65.7%	5:04 [1.4]	1.60 [6.3]	52.3%
deadlift + run	1:31 [1.1]	2.50 [15.5]	67.0%	1:18 [0.3]	2.51 [6.9]	56.9%
cj + run	1:52 [2.3]	1.66 [28.1]	87.0%	6:57 [4.0]	1.49 [12.1]	80.6%
thrusters + deadlift	4:52 [6.4]	1.78 [25.4]	84.6%	4:50 [2.2]	1.64 [10.1]	74.4%
cj + deadlift	6:10 [9.8]	1.34 [28.8]	87.5%	10:48 [7.9]	1.23 [12.4]	81.4%
thrusters + cj	4:37 [9.9]	1.28 [42.1]	93.8%	6:38 [7.2]	1.22 [22.8]	93.7%

This table reports the results of regressions of total times from the 2008 CrossFit Games on the times of individual workouts and workout pairs. Only competitors that completed all four events, as prescribed and within the allotted time limit, are included (119 men and 36 women) The numbers in square brackets are the test-statistics for the corresponding parameters.

The slope coefficient tells how many seconds athletes' total times increased, on average, with a one second increase in the time required to perform the explanatory workout or workouts. The adjusted-R² is the percentage of the variation in the athletes' total times explained by the explanatory workout or workouts.

respectively, pushing Tamara Holmes (who secured fourth in large part due to her second-place performance in the clean and jerk event) down two spots to sixth.

These balanced results are not objectively superior to the actual Games results in any way, and I expect that individuals' subjective opinions of the balanced format are highly correlated with their own strengths, with stronger runners embracing the concept and stronger Olympic lifters appalled. The balanced format does, however, share attractive features with both the event place scoring used in the 2007 Games and the cumulative time scoring used in 2008. In the balanced format, every event "matters." At the same time, the balanced format rewards dominant victories (like Jason Khalipa's minute-plus thrashing of the other top competitors in Grace) more than marginal victories (like Matt Chan's four-second advantage over Mr. Khalipa in the deadlift event). Rewarding performance across events and dominant performance more than marginal performance both seem reasonable goals for a multi-event scoring system. Some discussion of balancing the dispersion in times across events should consequently be included in any debate regarding the relative merits of event place scoring versus cumulative time scoring.



About The Author

Robert Novy-Marx joined the University of Chicago Booth School of Business faculty in 2004, and studies asset pricing, real options, general equilibrium, and real estate finance. Novy-Marx earned the American Real Estate and Urban Economics Association Dissertation Award in 2005 and the Western Finance Association's Trefftz Award in 2004 for "An Equilibrium Model of Investment Under Uncertainty."

Novy-Marx earned a bachelor's degree in physics from Swarthmore College in 1991 and a PhD in finance from the Haas School of Business at the University of California at Berkeley in 2003. He joined Chicago Booth that same year.

A former professional triathlete, he received a bronze medal at the 1993 World Triathlon Championships in Manchester, England.

