Understanding Correlations Between Short Sprint Distance Ratios

by C. D. Chester

Previously in my research I have modeled all my data off Brian Mackenzie's online calculators. So, I brought it upon myself to make my own research project and create my own regression models.

I started by going to the IAAF website and analyzing 125 unique sprinters PB's and their ratios, if they had any. I selected 100 from the 2018 Men's Senior Best by Athlete section and 25 from the respective 400m version as well.

Together those groups added up to approximately 2600.[1] However, only some of these athletes did all the events I was looking for, so it was necessary to include the lower-bound as well of approximately 1100 athletes. By selecting 125 unique sprinters it achieved a 95% confidence rating and margin of error of 9%. However, of the 125 unique sprinters it was narrowed down to 45 unique sprinters who all had PB's listed for the 100m, 200m, and 400m. This made the margin of error go up to 14%.

As previously stated only 45 of the 125 unique athletes were used in the data regression. This notes that $36 \pm 14\%$ of all 2600 athletes ran the 100m, 200m, and 400m.[2] Thus, at most only half of the 2600 ran the required events, while a minimum 22% of the 2600 ran the events.

Upon taking the ratios between the 400m and the 100m (4/1 R) and the 400m compared to the 200m (4/2 R) as well yielded the only result with high correlation. Thus, using a specific time of any of the results to predict another was not directly successful. However, by manipulating these ratios of high correlation I believe I have found a relevant way to predict times.

The regression models I tried gave a correlation in between 93 and 95%, thus I opted to take one the easier and better regression: the linear model. The linear model gave a correlation of 93.6%[3] by using this formula:

$$\left(\frac{400m}{200m}Ratio\right) \approx 0.4081\left(\frac{400m}{100m}Ratio\right) + 0.4278;\ 4.07 \leq \left(\frac{400m}{100m}Ratio\right) \leq 5.03$$

$$\therefore \left(\frac{400m}{100m}Ratio\right) \approx \left[\left(\frac{400m}{200m}Ratio\right) - 0.4278\right] \div 0.4081;\ 2.11 \leq \left(\frac{400m}{200m}Ratio\right) \leq 2.5$$

---

[1] Actual total was 2,574 total athletes with duplicates counted.

[2] $45 \div 125 = 36\%$

[3] Note that due to it having a 93.6% correlation it has a deviation of 6.4%.

It is also worth noting this formula as it is key to using any time to predict another:

$$\left(\frac{200m}{100m}Ratio\right) = \left(\frac{400m}{100m}Ratio\right) \div \left(\frac{400m}{200m}Ratio\right)$$

$$\therefore \left(\frac{200m}{100m}Ratio\right) \approx \left(\frac{400m}{100m}Ratio\right) \div \left[0.4081\left(\frac{400m}{100m}Ratio\right) + 0.4278\right];$$
$$1.95 \leq \left(\frac{200m}{100m}Ratio\right) \leq 2.03$$

$$\therefore \left(\frac{200m}{100m}Ratio\right) \approx \left[\left(\frac{400m}{200m}Ratio\right) - 0.4278\right] \div 0.4081\left(\frac{400m}{200m}Ratio\right);$$
$$1.95 \leq \left(\frac{200m}{100m}Ratio\right) \leq 2.03$$

Therefore the $\left(\frac{200m}{100m}Ratio\right)$ must be in or at the bounds of 1.95 and 2.03.

To refer to an earlier regression I did imagine that the athlete ran an 11 second 100m. Through my regression of Brian Mackenzie's data, it said that that athlete would run a 200m in 22.14 seconds and a 400m in 49.22. Comparing that to these boundary formulas give that an athlete running an 11 second 100m will run a 200m in between 21.45 and 22.33$\pm$6% seconds and the 400m in between 45.26 and 55.83$\pm$6% seconds.[4] My regression model of Brian's data is in that range so hypothetically the formulas I just found from the IAAF data should be accurate to an extent. To give a better example use the 11 second 100m again, but instead of using the bounds use a direct ratio:

$$\left(\frac{200m}{100m}Ratio\right) \approx 2.013, \left(\frac{400m}{100m}Ratio\right) \approx 4.475, \left(\frac{400m}{200m}Ratio\right) \approx 2.223$$

$$100m = 11\ seconds$$

$$200m \approx 22.14 \pm 6\%\ \ seconds$$

$$400m \approx 49.22 \pm 6\%\ \ seconds$$

This time the results of Brian's were achieved (with a margin of error) and they stayed within the bounds; they obviously were going to be included since Brian's time length was found in our general time estimate. It should be noted that these formulas I just made only work for well-trained men. For instance, someone running a 200m in 25.2 and 100m in 11.9 has a $\left(\frac{200m}{100m}Ratio\right) \approx 2.12$ which exceeds the 2.03 upper bound.[5] This shows that the athlete has yet to reach near optimal potential.

---

[4] The $\pm$6% applies to both bounds, not just the larger bound.

[5] This exact result is from a high school student I coached who was a freshman. The athlete had poor stamina, and this reflects via exceeding the upper bound of 2.03.

To summarize my results:

$$\left(\frac{200m}{100m}Ratio\right) = \left(\frac{400m}{100m}Ratio\right) \div \left(\frac{400m}{200m}Ratio\right)$$

$$\left(\frac{400m}{200m}Ratio\right) \approx 0.4081\left(\frac{400m}{100m}Ratio\right) + 0.4278;\ 4.07 \leq \left(\frac{400m}{100m}Ratio\right) \leq 5.03$$

$$\left(\frac{400m}{100m}Ratio\right) \approx \left[\left(\frac{400m}{200m}Ratio\right) - 0.4278\right] \div 0.4081;\ 2.11 \leq \left(\frac{400m}{200m}Ratio\right) \leq 2.5$$

$$\left(\frac{200m}{100m}Ratio\right) \approx \left(\frac{400m}{100m}Ratio\right) \div \left[0.4081\left(\frac{400m}{100m}Ratio\right) + 0.4278\right];$$

$$1.95 \leq \left(\frac{200m}{100m}Ratio\right) \leq 2.03$$

$$\left(\frac{200m}{100m}Ratio\right) \approx \left[\left(\frac{400m}{200m}Ratio\right) - 0.4278\right] \div 0.4081\left(\frac{400m}{200m}Ratio\right);$$

$$1.95 \leq \left(\frac{200m}{100m}Ratio\right) \leq 2.03$$

- Works with athletes near optimal potential.
- 6% margin of error for final results.
- 22-50% of US track and field sprinters run the 100m, 200m, and 400m.

References

https://www.iaaf.org/records/toplists/sprints/400-metres/outdoor/men/senior/2018?regionType=countries&region=usa&timing=electronic&page=1&bestResultsOnly=true

https://www.iaaf.org/records/toplists/sprints/100-metres/outdoor/men/senior/2018?regionType=countries&region=usa&timing=electronic&windReading=regular&page=1&bestResultsOnly=true

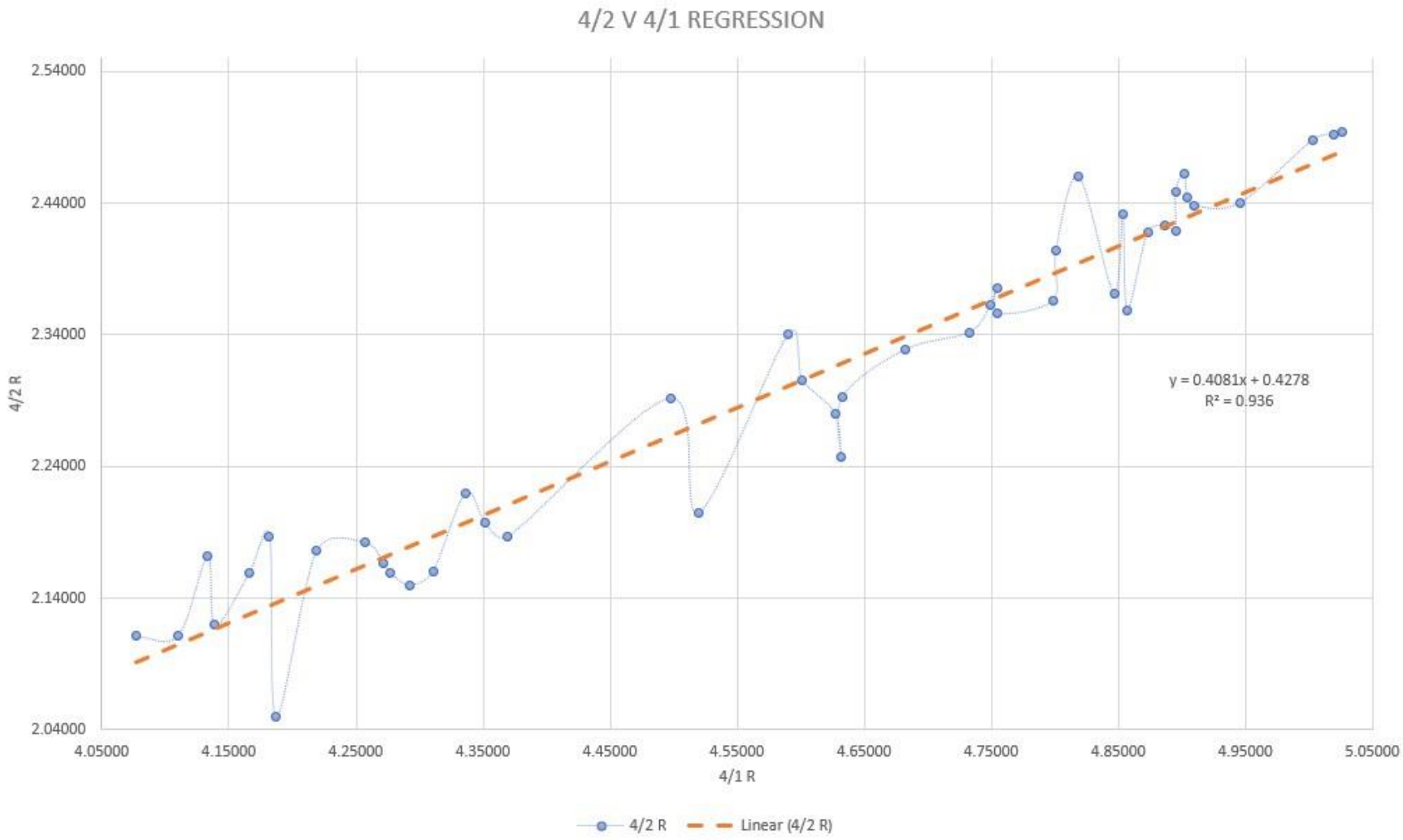
https://www.athletic.net/TrackAndField/Athlete.aspx?AID=9295294#!/L0

https://cdchester.co.uk/wp-content/uploads/2018/05/100m-200m-and-400m-prediction-formulas.pdf

https://cdchester.co.uk/2018/05/08/247/

Data Used

NAME	100M	200M	400M	2/1 R	4/1 R	4/2 R
PATRICK BLAKE LEEPER	11.05	21.34	45.05	1.93122	4.07692	2.11106
WILBERT LONDON	10.82	21.06	44.47	1.94640	4.10998	2.11159
STEWART TREVOR	11.04	21.01	45.63	1.90308	4.13315	2.17182
MICHAEL CHERRY	10.79	21.07	44.66	1.95273	4.13902	2.11960
FRED KERLEY	10.49	20.24	43.70	1.92946	4.16587	2.15909
PAUL DEDEWO	10.67	20.40	44.62	1.91190	4.18182	2.18725
DEVIN DIXON	10.80	22.06	45.22	2.04259	4.18704	2.04986
RICKY MORGAN	10.77	20.88	45.44	1.93872	4.21913	2.17625
CHRISTIAN TAYLOR	10.61	20.70	45.17	1.95099	4.25730	2.18213
CALVIN SMITH	10.49	20.68	44.81	1.97140	4.27169	2.16683
KAHMARI MONTGOMERY	10.50	20.80	44.91	1.98095	4.27714	2.15913
RASHARD CLARK	10.61	21.18	45.54	1.99623	4.29218	2.15014
QUINTAVEON POOLE	10.50	20.95	45.26	1.99524	4.31048	2.16038
MICHAEL NORMAN	10.27	20.06	44.53	1.95326	4.33593	2.21984
TRYELL RICHARD	10.39	20.57	45.21	1.97979	4.35130	2.19786
GIL ROBERTS	10.12	20.22	44.22	1.99802	4.36957	2.18694
CORREION MOSBY	10.34	20.29	46.50	1.96228	4.49710	2.29177
KASAUN JAMES	10.29	21.09	46.51	2.04956	4.51992	2.20531
JARON FLOURNOY	10.32	20.24	47.37	1.96124	4.59012	2.34042
NARON ROLLINS	10.40	20.75	47.85	1.99519	4.60096	2.30602
KOSSI TCHENAWOU	10.40	21.10	48.12	2.02885	4.62692	2.28057
RONNIE BAKER	9.97	20.55	46.18	2.06118	4.63190	2.24720
TYSON GAY	9.69	19.58	44.89	2.02064	4.63261	2.29265
ISAIAH BRANDT-SIMS	10.41	20.93	48.74	2.01057	4.68204	2.32871
KENZO COTTON	10.07	20.35	47.66	2.02085	4.73287	2.34201
DEVON ALLEN	10.26	20.62	48.73	2.00975	4.74951	2.36324
MCKINLEY WEST	10.29	20.59	48.92	2.00097	4.75413	2.37591
JAYLEN BACON	10.00	20.18	47.55	2.01800	4.75500	2.35629
TERRYON CONWELL	10.33	20.95	49.57	2.02807	4.79864	2.36611
AMEER WEBB	9.94	19.85	47.72	1.99698	4.80080	2.40403
TERRANCE LAIRD	10.42	20.41	50.21	1.95873	4.81862	2.46007
KAREEM FAIR	10.35	21.15	50.16	2.04348	4.84638	2.37163
MCCONICO KWANTREYL	10.30	20.56	49.99	1.99612	4.85340	2.43142
DARRYL HARAWAY	10.20	21.00	49.54	2.05882	4.85686	2.35905
KHANCE MEYERS	10.31	20.78	50.24	2.01552	4.87294	2.41771
ANTHONY SCHWARTZ	10.15	20.47	49.60	2.01675	4.88670	2.42306
JEREMY PHILLIPS	10.36	20.96	50.71	2.02317	4.89479	2.41937
MALIK WILSON	10.38	20.75	50.81	1.99904	4.89499	2.44867
DETORRIAN GREEN	10.34	20.58	50.68	1.99033	4.90135	2.46259
MUSTAQEEM WILLIAMS	10.22	20.50	50.12	2.00587	4.90411	2.44488
ERIC HARRISON	10.28	20.70	50.47	2.01362	4.90953	2.43816
KYREE KING	10.00	20.27	49.46	2.02700	4.94600	2.44006
JOSH DAVIS	10.19	20.49	50.98	2.01079	5.00294	2.48804
KENDAL WILLIAMS	10.06	20.26	50.50	2.01392	5.01988	2.49260
MICHAEL TIMPSON	10.33	20.82	51.92	2.01549	5.02614	2.49376


4/2 V 4/1 REGRESSION
2.54000
2.44000
2.34000
2.24000
2.14000
2.04000
4/2 R
4.05000
4.15000
4.25000
4.35000
4.45000
4.55000
4.65000
4.75000
4.85000
4.95000
5.05000
4/1 R
y = 0.4081x + 0.4278
R² = 0.936
4/2 R
Linear (4/2 R)