


Workshop Resources



Excel worksheet

Download this worksheet to learn how to calculate reliability (CV%), the standard deviation, and the minimum meaningful change for your athletes over time.

Download

Measuring Metrics that Matter and Meaningful Change:
Statistics for the Applied Practitioner

Prof. Anthony Turner | a.n.turner@mdx.ac.uk

1

Presentation aim. How to:

1

Choose a test that informs your practice

2

Filter out the noisy, pointless metrics

3


Determine how much you trust the remaining metrics

4

Analyse each athlete's data

5

Set targets for each athlete



2

Step 1. Choosing a test

...that informs your practice

3

Identifying what to test and train through a needs analysis							
Coach's Physical KPI's	Cover lots of distance	Be fast	Be agile	Multiple sprints	Win aerial challenges	Win tackles (protect ball)	Be robust
Physical quality	Aerobic capacity	Speed & acceleration	CoDS	RSA	Power	Strength	Symmetry & ROM
Test	MAS	10 m & 30 m, RSI	Pro-agility (inc. decel)	30 m x 6, 20 s rest	CMJ (inc. Loaded jumps)	IMTP	OHS & Nordboard
Exercises	HIIT, SIT, SSG	SPD and Accel drills, plyometrics	Deccel and agility	HIIT, SIT, SSG	Power training (ballistics)	Strength training (squats)	Hams, adductors, glutes, eccentrics, unilateral

4

Choosing a test

Biological Basis

- Is there a justifiable link between the metric of interest and athletic performance?
- Does a theoretical cause and effect relationship exist?

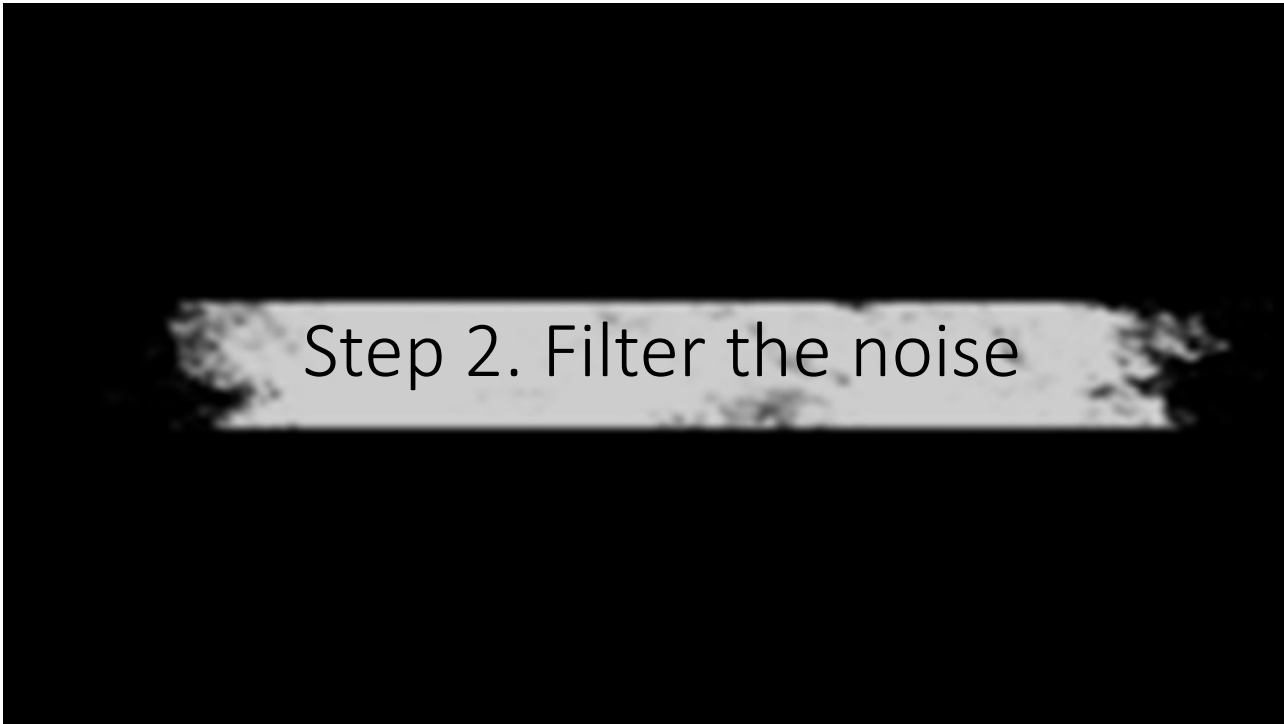
Feasibility

- Logistics surrounding its implementation including: cost, time and staffing.
- How long does it take to produce a report for coaches?
- Is the right culture in place?

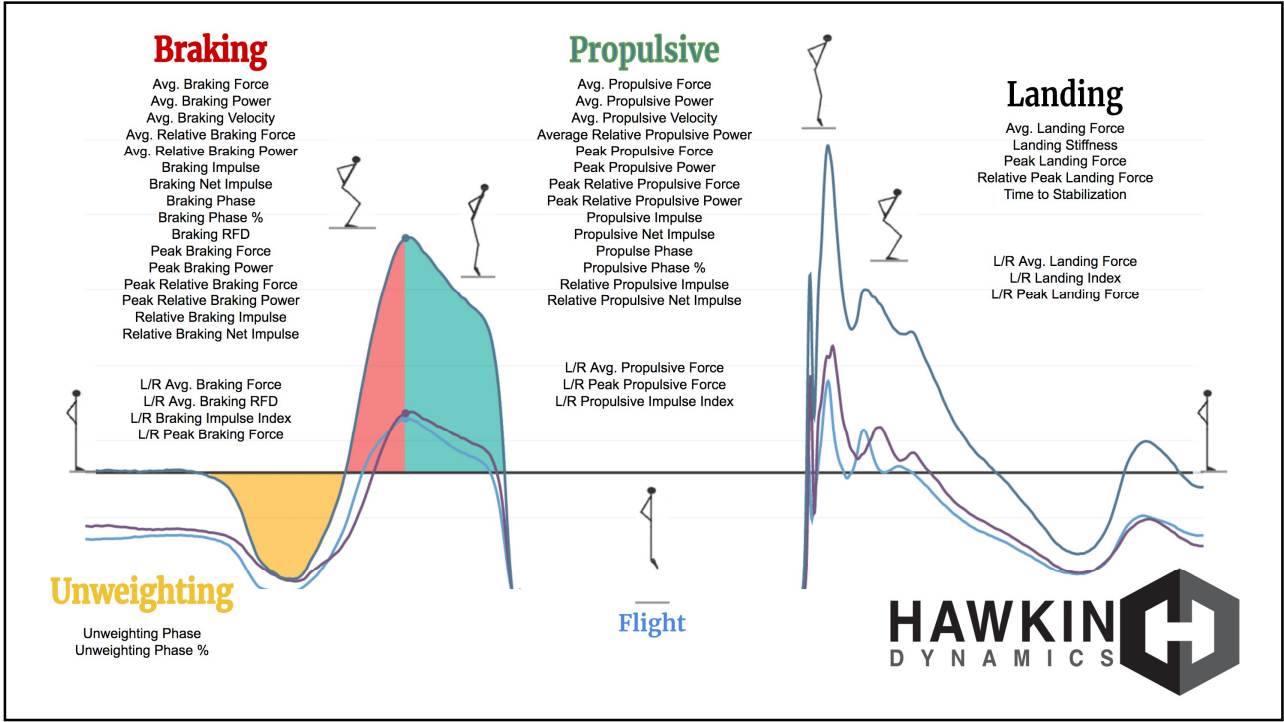
Sensitivity

- To what accuracy can it detect true changes?
- Realistically, can you actually inform practice off the back of this measure?

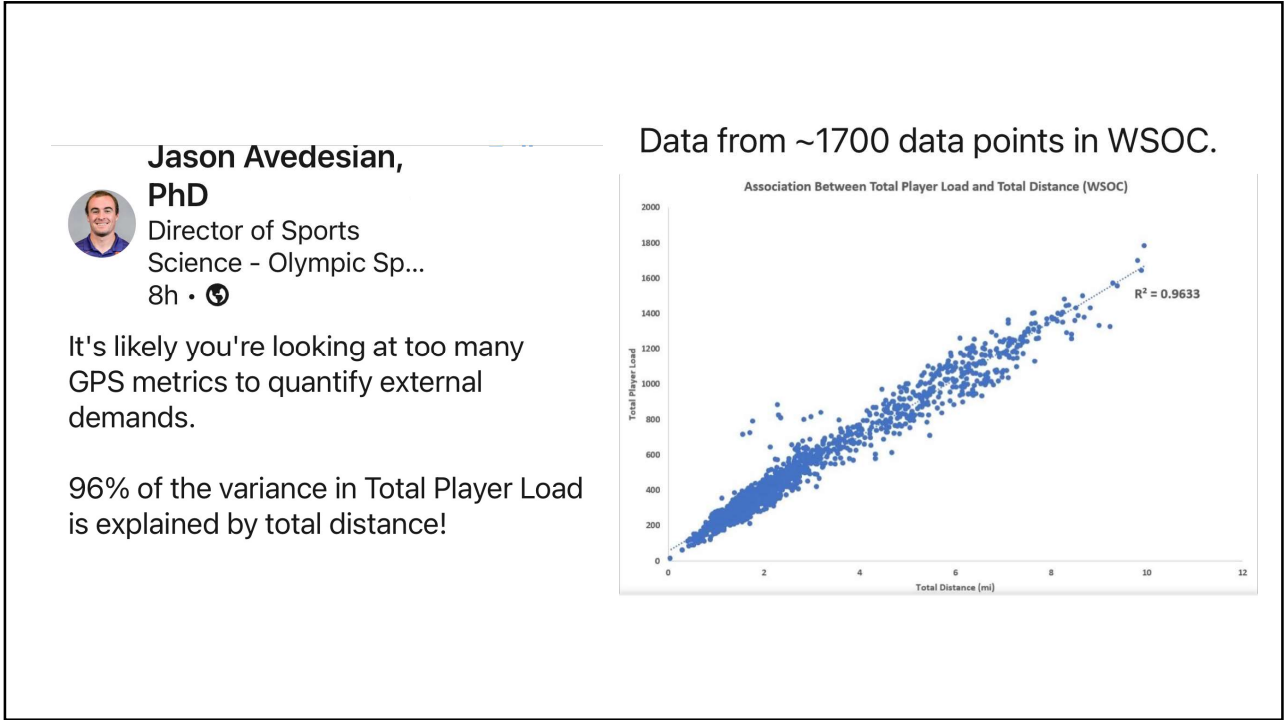
5



6



7



8

Original Research

Journal of Strength and Conditioning Research

Intra- and Interday Reliability of Weightlifting Variables and Correlation to Performance During Cleans

Angela M. Sorensen,¹ Shyam Chavda,¹ Paul Comfort,² Jason Lake,³ and Anthony N. Turner¹

¹London Sports Institute, Middlesex University, London, United Kingdom; ²Human Performance Laboratory, University of Salford, Salford, United Kingdom; and ³Department of Sport and Exercise Sciences, University of Chichester, Chichester, United Kingdom

Level one: Which variables are reliable?

Level 2: Which variables are highly correlated (multicollinearity)?

Level 3. Of the correlated variables which one statistically or logically best explains the performance outcome

Results

Sixteen of the 70 variables analyzed were found to have good to excellent intra- and interday ICC (0.779–0.994 and 0.969–0.996, respectively) and CV (0.64–6.42 and 1.14–6.37, respectively) values (30,36). Using the Pearson’s correlation coefficients ($r = 0.5–1.0$ at $p < 0.005$), these 16 variables were also shown to have strong correlations ($r = 0.880–0.988$) to cleans performed at 90% 1RM. From these 16 variables, bar work variables that were used to calculate bar power variables were then excluded because they are derived from the same force and displacement data and represented duplicate data. The resulting variables were further assessed for multicollinearity, which can be seen in Table 3. This system of filtering resulted in a total of 11 variables exhibiting “good to excellent” ICC with a CV of $\leq 10\%$ for both intraday and interday reliability measures and with correlations to clean performance as reported in Table 2.

9

Sportsmith

Learn Events Premium Courses Listen

ARTICLE

Building the 5-2-180 change of direction speed test

Anthony Turner

10

16th Annual Strength and Conditioning Student Conference

5

Just because you can, doesn't mean you should!

Complexity Bias.

The tendency to prefer more complex or sophisticated options over simpler ones, often because they sound more important or impressive.

11



If you start with the wrong metric, or a valid yet noisy one, there is no form of analysis that can save you from rubbish data and meaningless inferences.

12

Step 3. Is the metric reliable

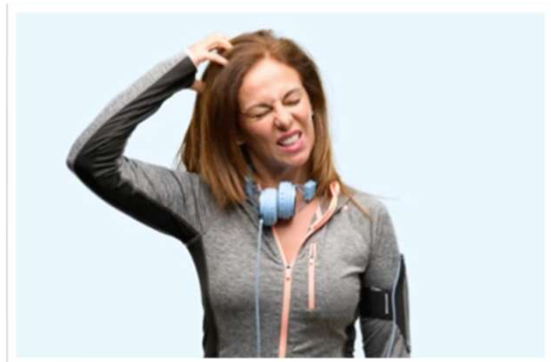
...How much should I trust it?

13

Explain this to your athlete

You bench press 3 times in a week

- In session 1 you bench 70 kg
- In session 2 you bench 72 kg
- In session 3 you bench 69 kg



14

What about this...

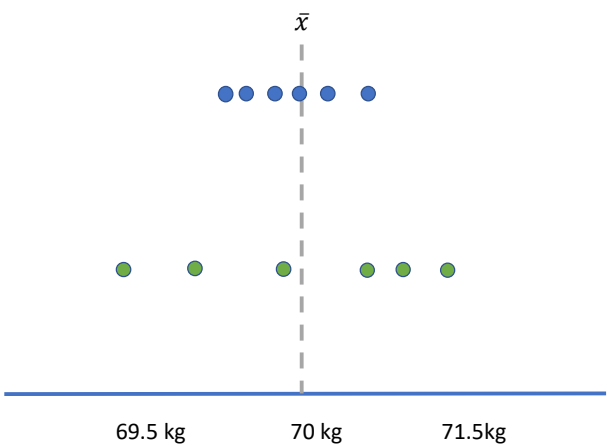
You weigh yourself everyday for 5 days

- On day 1 you weigh 70 kg
- On day 2 you weigh 70.5 kg
- On day 3 you weigh 69.9 kg
- On day 4 you weigh 70.1
- On day 5 you weigh 70.3

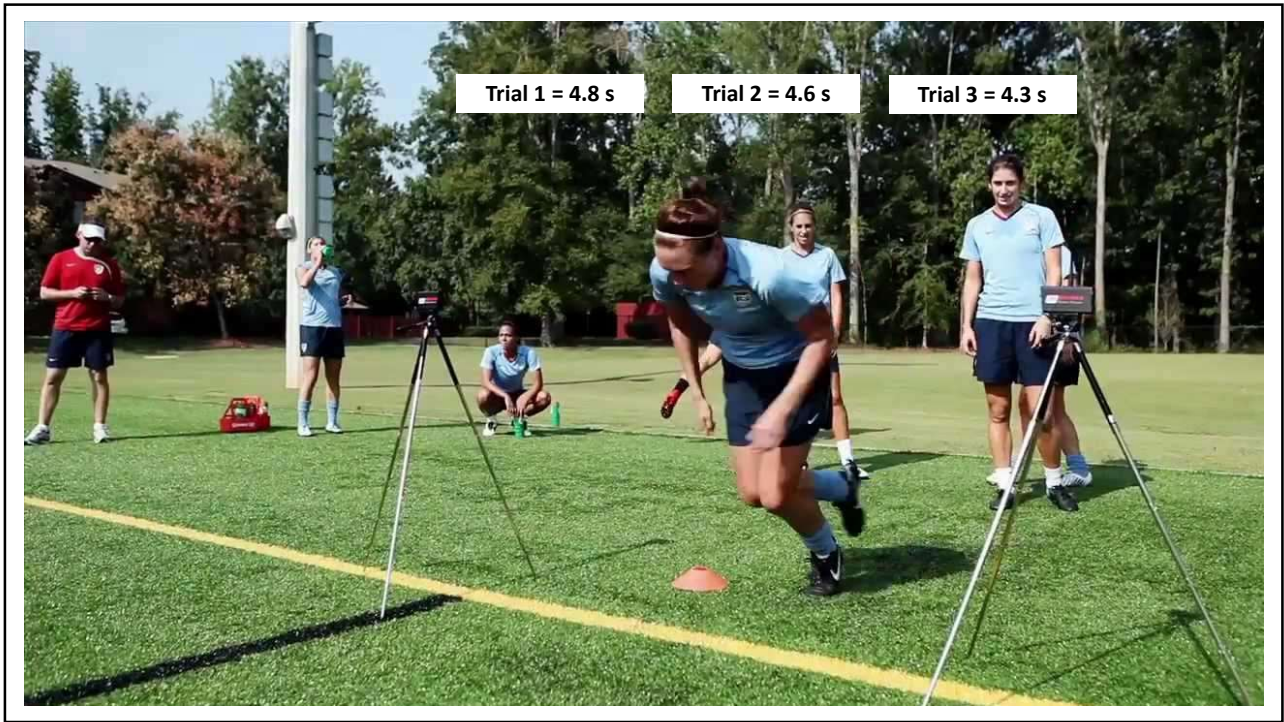


15

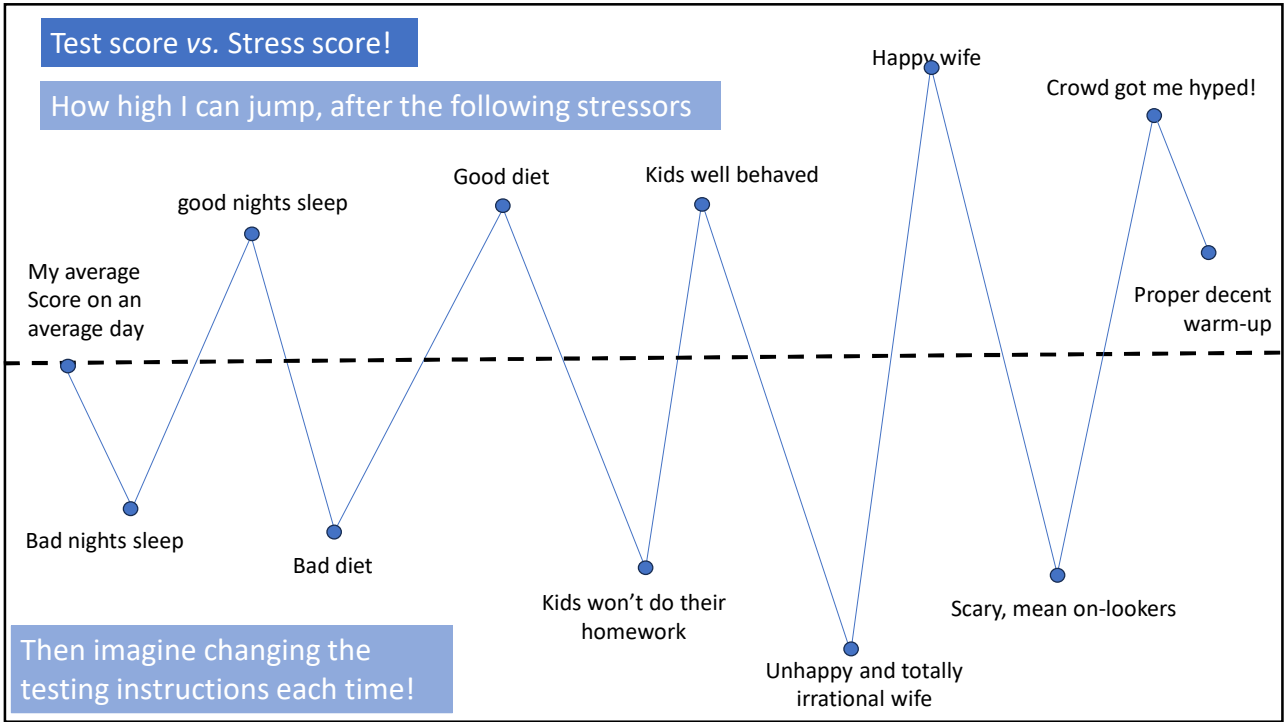
Which weighing scale would you buy?



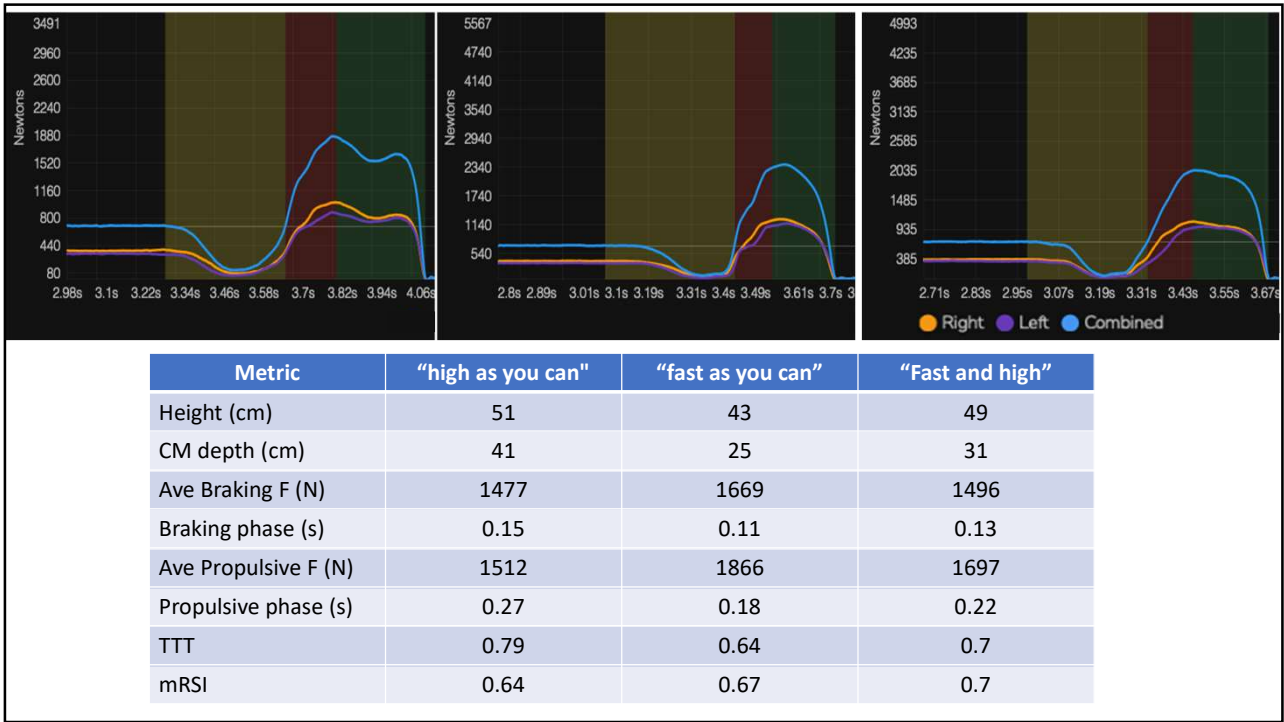
16



17



18



19

Coefficient of variability (CV)

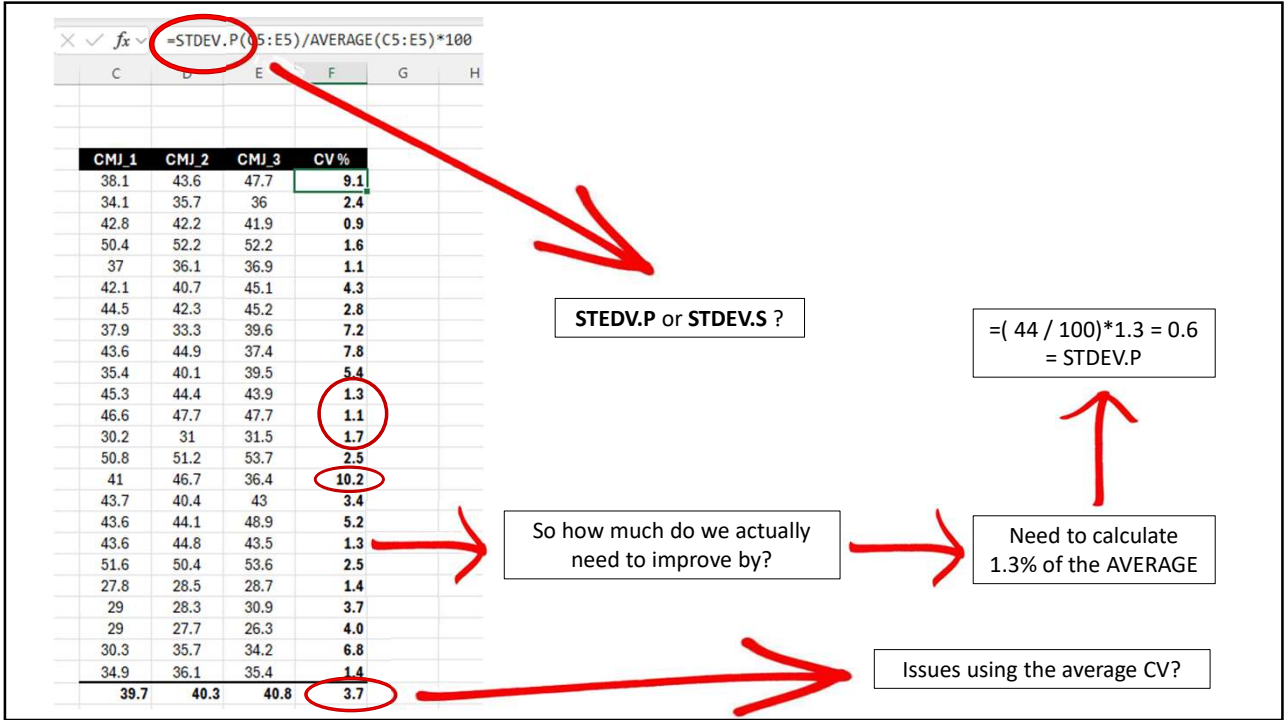
CV % = (SD/mean) *100

- CV of 10% suggests that the SD is 10% of the mean. The higher the CV, the less consistent the data points
- CV best measure of reliability if comparing tests with different units
- E.g., which is more reliable, jump height system with an SD of **3 cm**, or peak force system with an SD of **100 N**?
- Mean score = **40 cm** and **2000 N** respectively. Therefore:

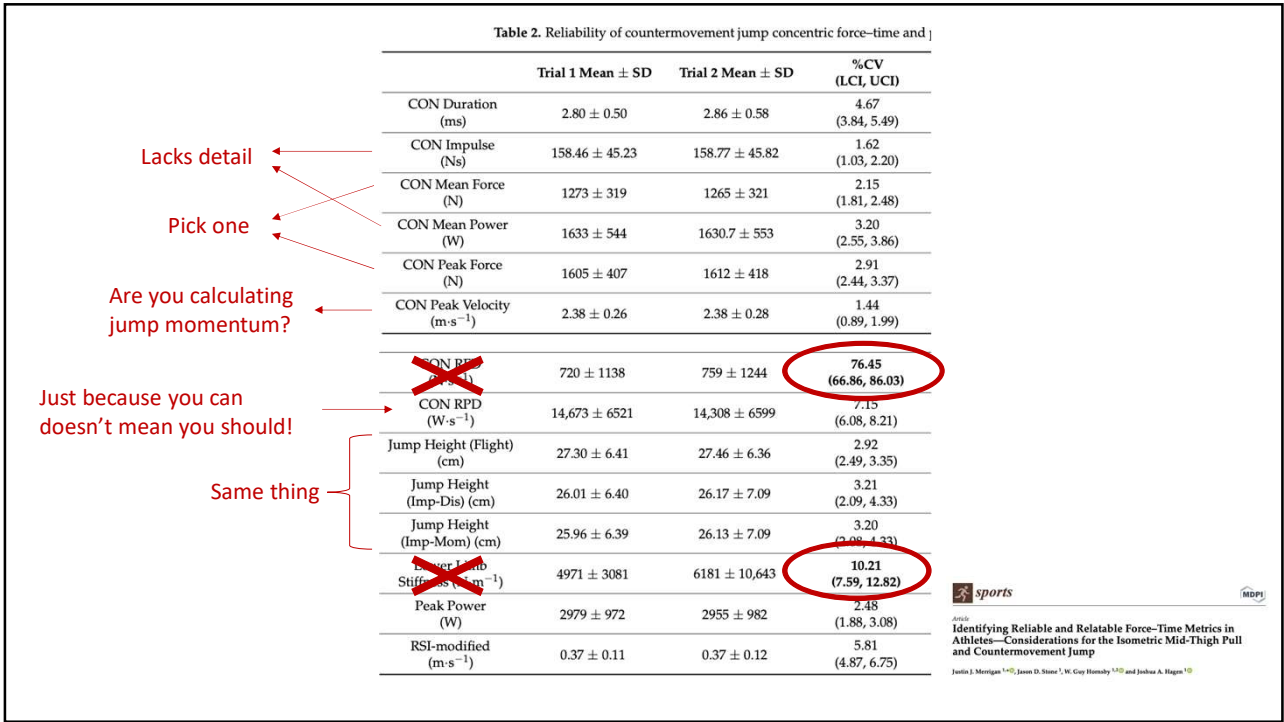
Jump height system
3/40 = 0.075 *100 = 7.5 %

Peak force system
100/2000 = 0.05 *100 = 5 %

20



21



22

If this changes, so too must braking F

Lacks detail

Same thing?

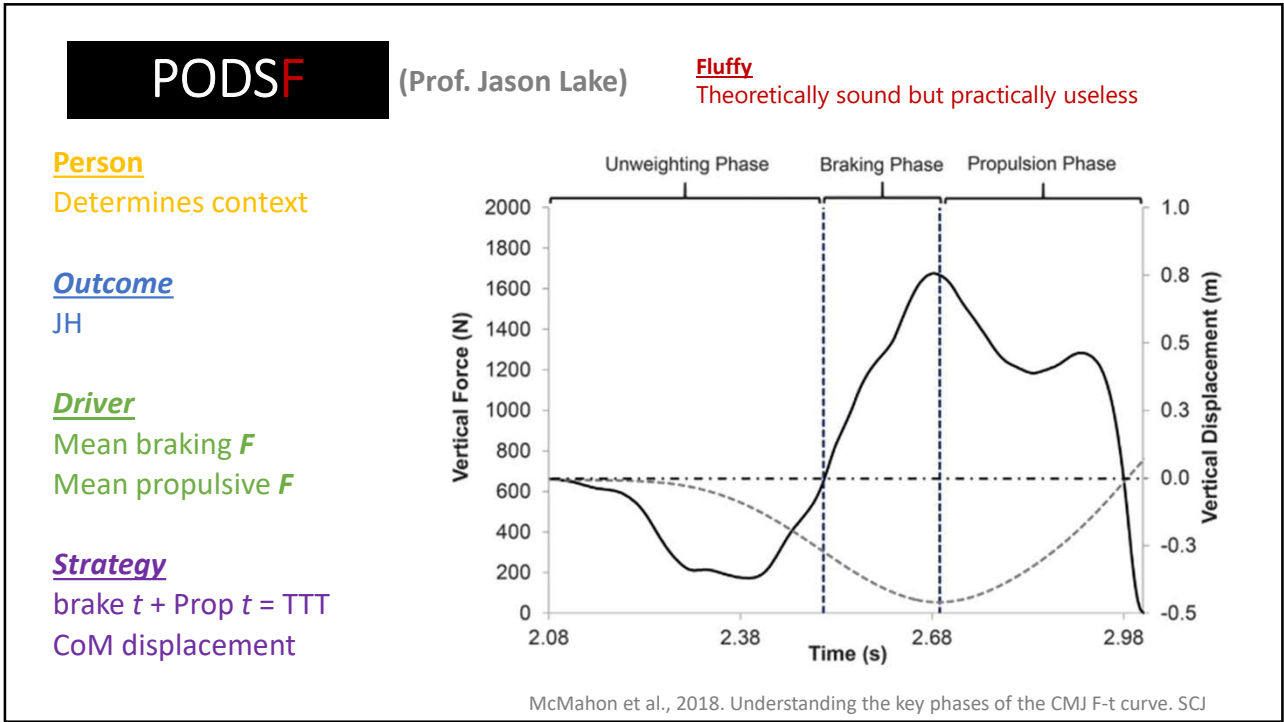
Good case to choose this one!

	Trial 1 Mean \pm SD	Trial 2 Mean \pm SD	%CV (LCI, UCI)
Dip Depth (cm)	-30.57 \pm 8.48	-31.23 \pm 9.82	7.66 (5.00, 10.32)
ECC Braking Impulse (Ns)	60.84 \pm 72.54	60.11 \pm 67.76	14.53 (11.36, 17.70)
ECC Braking RFD (N.s)	4780 \pm 2060	4924 \pm 2446	10.86 (9.26, 12.45)
ECC Decel. Impulse (Ns)	95.56 \pm 50.31	95.26 \pm 51.46	6.34 (4.13, 8.55)
ECC Decel. RFD (N.s)	5661 \pm 2890	5841 \pm 3256	10.38 (8.56, 12.20)
ECC Duration (ms)	479.3 \pm 76.0	479.2 \pm 93.7	6.41 (4.50, 8.31)
ECC Mean Braking Force (N)	843.8 \pm 193.2	846.6 \pm 199.2	3.38 (2.71, 4.05)
ECC Mean Decel. Force (N)	1222 \pm 309	1224 \pm 329	3.43 (2.71, 4.15)
ECC Mean Force (N)	689.9 \pm 149.7	689.8 \pm 149.6	0.06 (0.05, 0.08)
ECC Mean Power (W)	438.0 \pm 142.1	442.5 \pm 151.1	6.83 (4.74, 8.92)
ECC Peak Force (N)	1573 \pm 406	1584 \pm 429	3.34 (2.70, 3.97)
ECC Peak Power (W)	1223 \pm 525	1239 \pm 602	9.24 (6.74, 11.74)
ECC Peak Velocity (m.s ⁻¹)	-1.23 \pm 0.30	-1.23 \pm 0.33	6.26 (4.04, 8.47)

Identifying Reliable and Relatable Force-Time Metrics in Athletes—Considerations for the Isometric Mid-Thigh Pull and Countermovement Jump

Justin J. Morrison ^{1,2,3}, Jason D. Stone ¹, W. Gray Houshky ^{1,2,3} and Joshua A. Hagan ^{1,2}

23

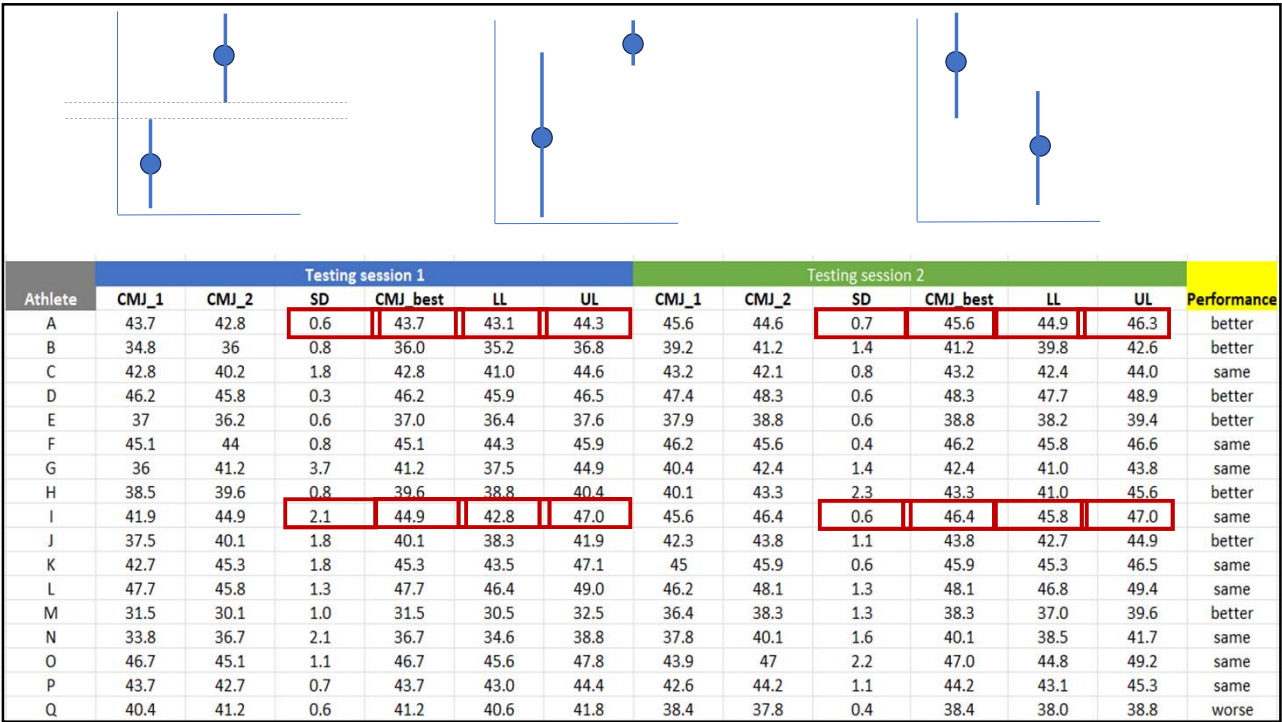


24

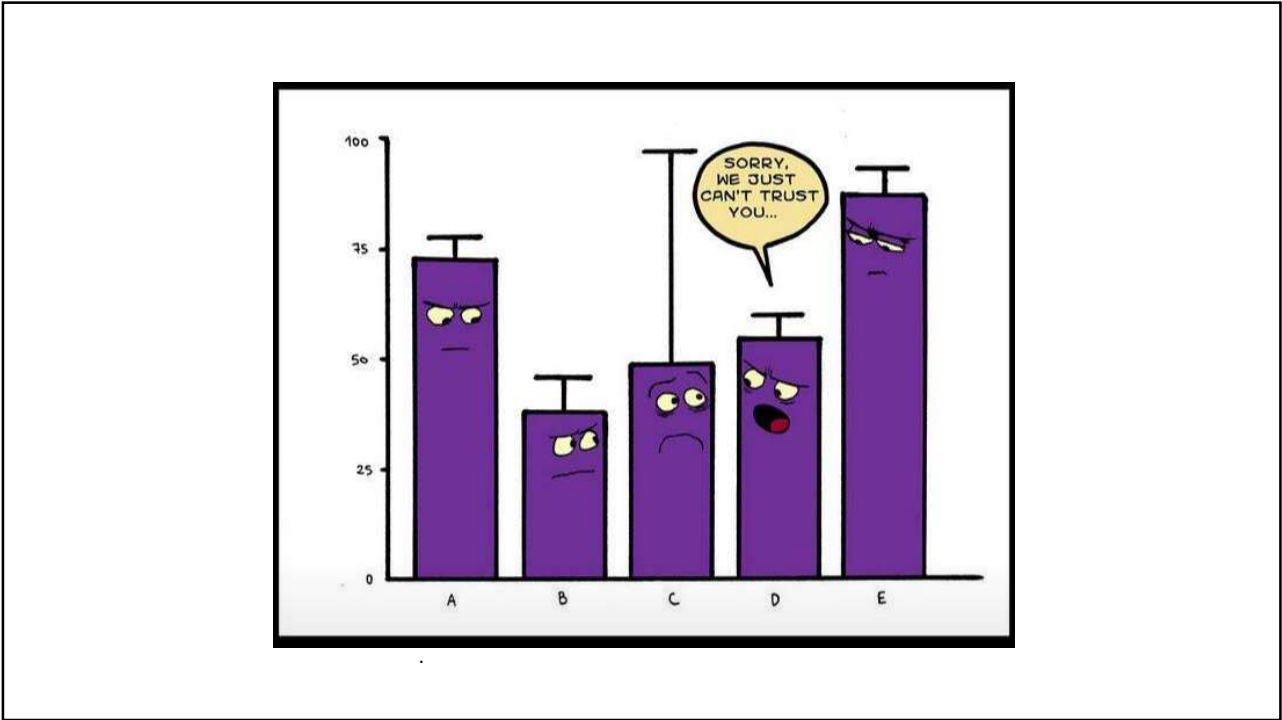
Step 4. Individual athlete analysis

...determining meaningful change

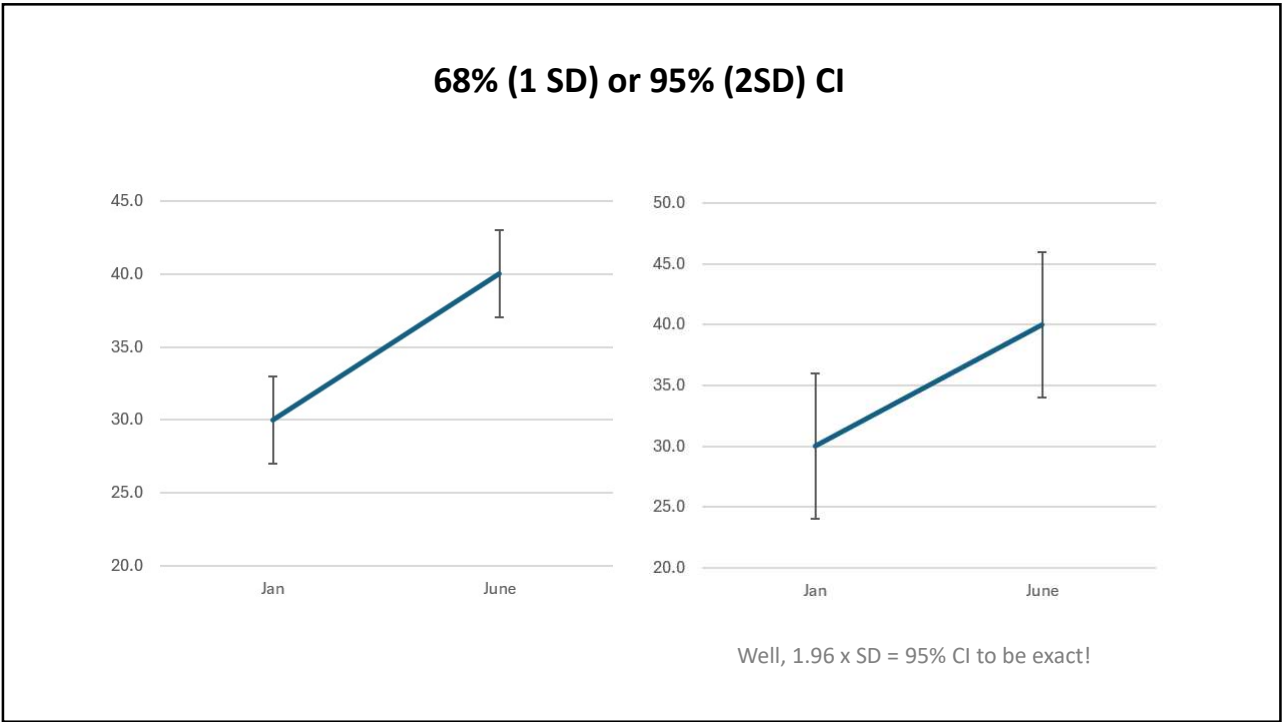
25



26



27



28

Type I or Type II error? That is the question

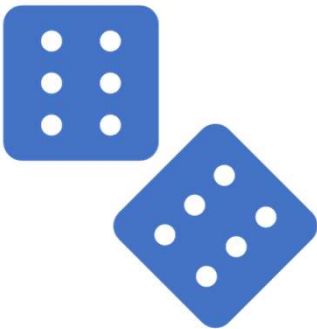
- A **Type I error is a false-positive** – you claim a difference when there is none
- **Type II error is a false-negative** – you claim no difference when there was one



29

Club philosophy: Risk vs. Reward

- Do you prefer to play it safe or be sensitive to smaller changes? Perhaps a philosophical question.
- There is no right or wrong answer. Sometimes you'll be right, sometimes you'll be wrong.
- Therefore, need to focus on the consequences of each scenario to help you choose.



COACHING WISDOM

But did my athlete improve?
Assessing performance changes when N = 1

OVERVIEW

The paper reviews the methods used to assess the risk of making incorrect conclusions about the effectiveness of an intervention when only one subject is available for comparison. It discusses the limitations of such studies and the importance of considering the consequences of Type I and Type II errors. It also provides a checklist for researchers to help them decide whether a study with N = 1 is justified.

INTRODUCTION

When athletes are engaged in training, they are often asked to perform a task that is slightly different from what they are used to. This is done to see if the new task is better than the old one. However, it is often difficult to tell if the new task is really better or if the difference is just due to chance. This is especially true when only one athlete is available for comparison. In such cases, the risk of making incorrect conclusions is high. This paper reviews the methods used to assess the risk of making incorrect conclusions about the effectiveness of an intervention when only one subject is available for comparison. It discusses the limitations of such studies and the importance of considering the consequences of Type I and Type II errors. It also provides a checklist for researchers to help them decide whether a study with N = 1 is justified.

CONCLUSIONS


The paper concludes that studies with N = 1 are often justified, but they must be designed carefully. Researchers should consider the consequences of Type I and Type II errors and should use appropriate statistical methods to assess the risk of making incorrect conclusions. The paper also provides a checklist for researchers to help them decide whether a study with N = 1 is justified.

30

Reducing the noise (*SD*)

The tester

- Expert
- Strict
- Coaching cues



The athlete:

- Homogenous group
- Technique
- Motivation
- Biological variability

Post CMJ – Pre CMJ
Variability

The environment

- Temperature
- Audience
- competition
- Music

The equipment

- Recording frequency
- Calibration
- Unobtrusive


31

Step 5. Comparison with teammates
...to set realistic targets

32

z scores and the TSA

33



Available at: <https://www.nscacertification.com/quiz/>

Total Score of Athleticism: Holistic Athlete Profiling to Enhance Decision-Making

Anthony N. Turner, PhD,¹ Ben Jones, PhD,¹ Perry Stewart, MSc,¹ Chris Bishop, MSc,¹ Nimal Pinar, PhD,¹ Shyam Chavda, MSc,¹ and Paul Reed, PhD²
¹London Sports Institute, Middlesex University, Allianz Park, London, United Kingdom; ²Carnegie Applied Rugby Research (CARR) Centre, Institute for Sport, Physical Activity and Leisure, Leeds Beckett University, Leeds, United Kingdom; ³Neural Performance and Research Team, Arsenal Football Club, United Kingdom; and ⁴Aspetar Orthopaedic and Sports Medicine Hospital, Doha, Qatar

ABSTRACT

Often, the various coaching staff, sport scientists, and medical practitioners of a sports club require a single, holistic indication of an athlete's athleticism. Currently, there is no consensus on how this is best defined, and thus, a total score of athleticism (TSA) may provide one such method. The TSA is derived from the average of Z-scores for 10 fitness measures (small samples) from a sport-specific testing battery, ensuring athletes are judged across all the relevant fitness capacities that best define the physical demands of competition. To aid readers in using the TSA, this article also details how it is computed in EXCEL.

INTRODUCTION

As strength and conditioning coaches, we routinely put our athletes through a variety of fitness assessments to determine their physical capability, so that we can tailor the design of their training program and adapt accordingly. Similarly, the psychologist, physiotherapist, and technical coaches also assess the athlete, with the results equally used to inform future interventions and team selection. But, with so much data collected and thus available for discussion, athlete review meetings, for example, where all staff attend, can often see each practitioner providing more discrete detail than is necessary. For example, although jump height may be informative to the strength and conditioning coach, this score, in this context, may not prove overly helpful to discussions led in by the coaches and other members of the sport science disciplines. These situations, therefore, lend themselves to the strength and conditioning coach providing a single score for the athlete's physical fitness, rather than separately discussing each individual test result. Such an approach can streamline collaborative communication, maximizing the time available for planning and practical delivery. Furthermore, coaches may not be as concerned in the raw score of each athlete, as much as where the score ranked among their teammates, especially when there is competition for places. For example, a coach may have no concept as to what is deemed a good jump height or back squat, with this information only becoming apparent through some analysis that reveals the score is among the highest or lowest in the squad. Also, it can be rare to have the athlete who scored highest on the bench press, also score the highest on a change of direction speed test or Yo-Yo score, for example, suggesting that there is some compromise among the different components of fitness that collectively

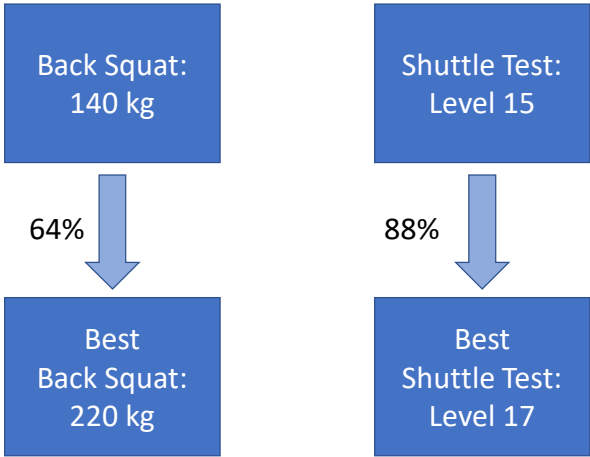
KEY WORDS: statistics, excel, data analysis, testing, feedback

Copyright © National Strength and Conditioning Association
Strength and Conditioning Journal | www.nscajournal.com 01
Copyright © National Strength and Conditioning Association. Unauthorized reproduction of this article is prohibited.



34

Is that score any good and which test did they do best on?



- But maybe the team is fit and they all scored well on the shuttle test...
- Level 15 may have been one of the lowest
- Conversely, there may only be a few strong athletes, so 140kg is really good!

35

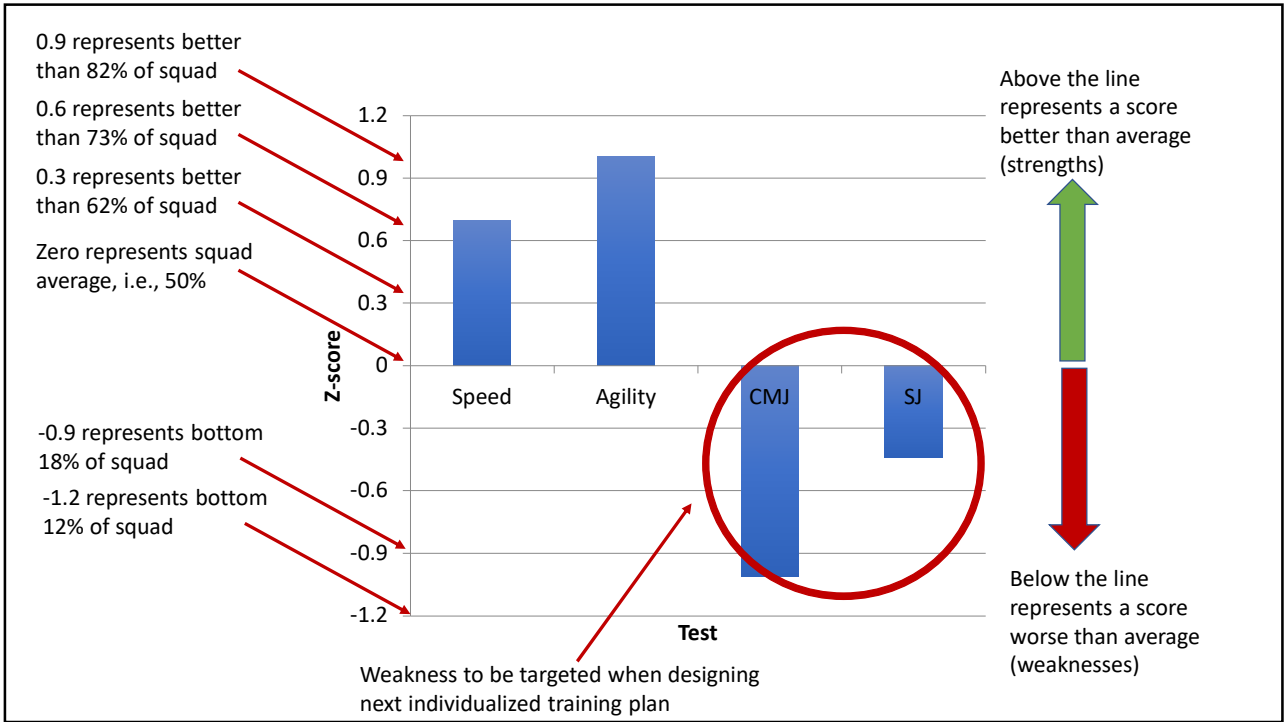
Turn test scores into a z-scores

- Z-scores tell you how many SD's a score is from the mean
- If a z-score = 0, it is identical to the mean score
- If a z-score = 1, it is 1 SD above the mean
- If a z-score = -1, it is 1 SD below the mean

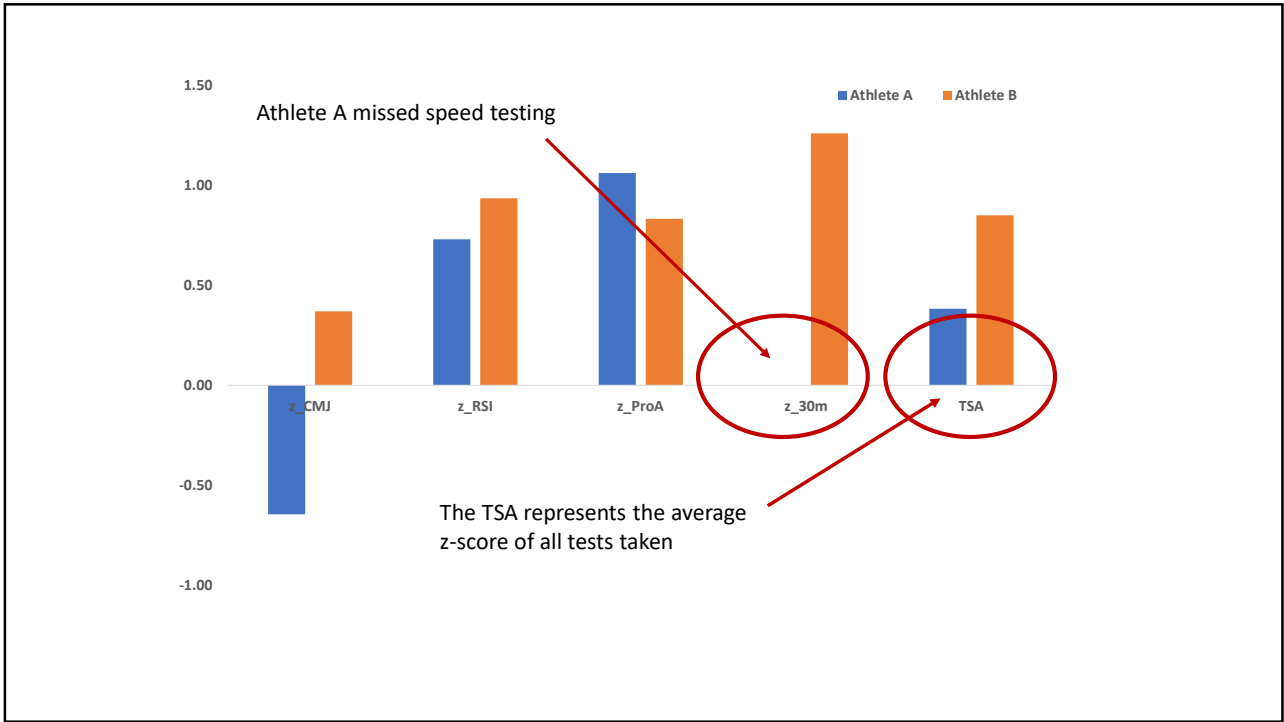
VLOOKUP ✖ ✔ fx =(S2-S\$26)/\$S\$27

	S	T	U	V	AE	AF	AG	AH	AI	AJ
1	Best_CMJ	Best_RSI	Best_ProA	Best_30m	z_CMJ	z_RSI	z_ProA	z_30m	TSA	Rank
2	47.7	1.6	5	4.5	SS27	-0.23	0.15	-1.49	-0.23	18
3	36	1.63	4.9	4.46	-0.77	-0.16	0.38	-1.25	-0.45	19
4	42.8	1.72	6.4	4.19	0.06	0.05	-3.06	0.41	-0.64	20
5	52.2	2.5	4.4	4.34	1.21	1.83	1.52	-0.51	1.01	2
6	37	2.02	4.6	4.25	-0.65	0.73	1.06	0.04	0.30	10
7	45.1	1.81	4.8	4.16	0.34	0.25	0.60	0.59	0.45	7
26	42.29	1.70	5.1	4.26						
27	8.16	0.44	0.44	0.16						

36



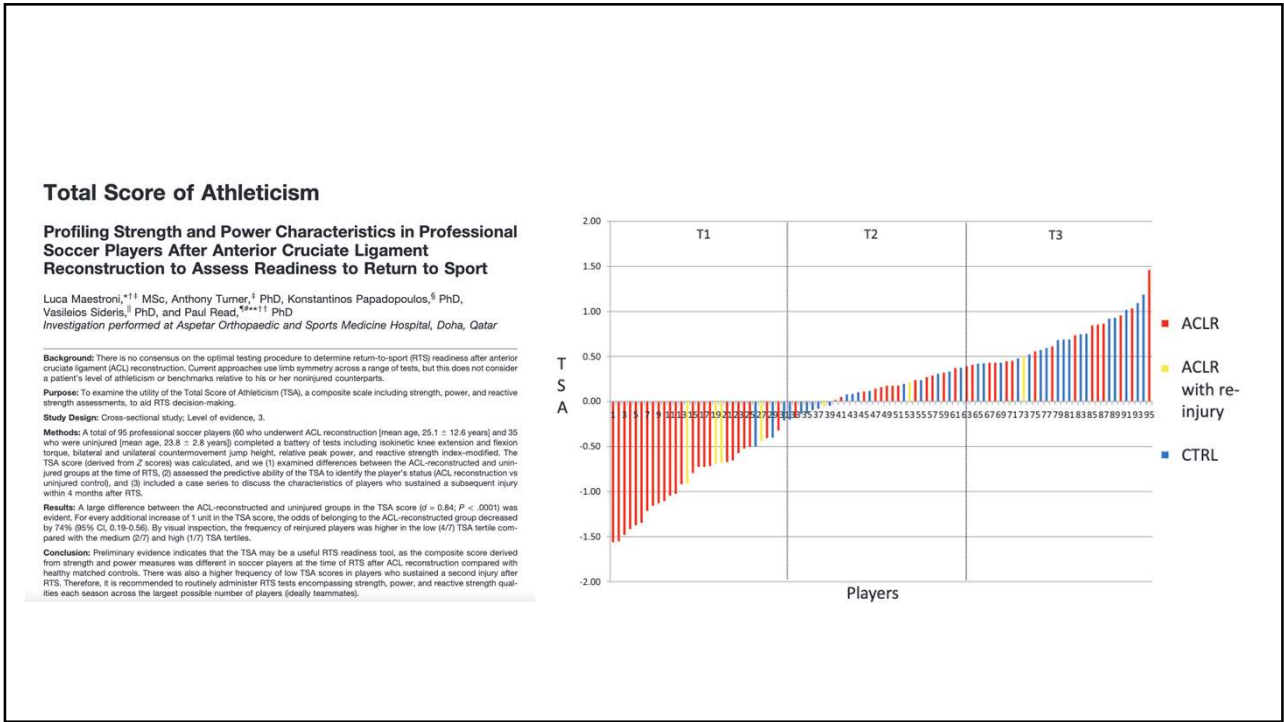
37



38



39



40

