Creative Commons Licence: Attribution-Noncommercial License. See: https://creativecommons.org/licenses/by-nc/4.0/. This is a pre-copyedited, author-produced version of an article accepted for publication in Medicine & Science in Sports & Exercise. The published version of record Abt, Grant; Tenan, Matthew S.. Research Design, Data Interpretation, and Adherence to Existing Recommendations in Medicine & Science in Sports & Exercise. Medicine & Science in Sports & Exercise ():10.1249/MSS.0000000003441, April 10, 2024 is available online at: DOI: 10.1249/MSS.00000000003441

Dear Editor-in-Chief,

We would like to comment on the paper by McKay et al. (1). Although we appreciate the constraints in conducting field studies with elite sporting populations, our comments here focus on fundamental aspects of research design, statistical reporting, and statistical interpretation. We also want to make readers aware of the recently published *MSSE* Statistical Reporting Recommendations (SRR) (2).

- 1. No sample size justification. As recommended in the SRR (2), justification of the sample size is important if researchers are to avoid underpowered studies, imprecise studies, and inflated effect sizes. The sample size reported by McKay et al. (1) was 18, which is typical in sport and exercise science (3). Small samples without prespecification are unlikely to detect a true effect, regardless of statistical significance (4).
- 2. No preregistration. As recommended in the SRR (2), study preregistration is important if researchers are to demonstrate that they haven't engaged in questionable research practices such as HARKing and *p*-hacking. While McKay et al. (1) describe the study as 'exploratory', the Discussion and Conclusion sections of their paper treat the results as if they confirm prespecified hypotheses with appropriate sample sizes. We also note that their clinical trial registration (NCT04783207) raises questions about selective outcomes reporting and does not contain a statistical analysis plan for confirmatory analyses to be tested.
- 3. **Nonsignificant 'trend'**. McKay et al. (1) write "Despite a lack of significant difference between the Iron PRE (7.3%) and Iron POST (6.2%) trials, our data does show a *nonsignificant trend* (P = 0.058)...". Numerous authors have outlined the statistical error of using this statement (5). This flawed interpretation of p-values is particularly perplexing in an exploratory study, where it has been argued p-values do not have a clear evidentiary interpretation (6).
- 4. Ignoring confidence interval values. McKay et al. (1) write "However, this [a 1.1% difference] is likely of limited practical benefit in the context of iron supplementation". As many do, McKay et al. (1) focus on the point estimate while ignoring the plausible values within the 95% confidence intervals [Iron PRE 4.4%–12.1%; Iron POST 3.1%–12.5%]. As suggested by Fidler et al. (7), simply reporting confidence intervals does not guarantee that they will be interpreted and discussed.
- 5. Confusing P > 0.05 as evidence of equivalence. "...in an acute setting, preexercise and postexercise iron supplementation appears to be *equally as effective*..." This statement is based on a *p* value of 0.058. As outlined in the SRR (2), a *p* value greater than the *a priori* alpha should not be used to suggest equivalence, which should be examined using equivalence tests.

Reporting guidelines only increase the quality of science if we follow them. Furthermore, the tendency for our field to continually treat small exploratory studies without prespecification as a high or even moderate level of evidence remains troubling. It is probably worthwhile for MSSE to consider separate Original Research classifications and guidelines for exploratory and confirmatory studies, particularly when exploratory work relies on p-values with an unclear interpretation (6,8).

References

- McKay AKA, Anderson B, Peeling P, Whitfield J, Tee N, Zeder C, et al. Iron Absorption in Highly Trained Male Runners: Does it Matter When and Where You Eat Your Iron? Med Sci Sports Exerc [Internet]. 2024 Jan;56(1):118–27. Available from: https://journals.lww.com/10.1249/MSS.00000000003272
- 2. Medicine and Science in Sports and Exercise. Medicine and Science in Sports and Exercise. [cited 2024 Jan 23]. Statistical Reporting Recommendations. Available from: https://journals.lww.com/acsm-msse/Pages/Statistical_reporting_recommendations.aspx

- 3. Abt G, Boreham C, Davison G, Jackson R, Nevill A, Wallace E, et al. Power, precision, and sample size estimation in sport and exercise science research. J Sports Sci [Internet]. 2020 Sep 1;38(17):1933–5. Available from: https://doi.org/10.1080/02640414.2020.1776002
- Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci [Internet].
 2013 May 10;14(5):365–76. Available from: http://www.nature.com/articles/nrn3475
- 5. Wood J, Freemantle N, King M, Nazareth I. Trap of trends to statistical significance: likelihood of near significant P value becoming more significant with extra data. BMJ [Internet]. 2014 Mar 31;348(mar31 2):g2215–g2215. Available from: http://dx.doi.org/doi:10.1136/bmj.g2215
- 6. de Groot AD. The meaning of "significance" for different types of research [translated and annotated by Eric-Jan Wagenmakers, Denny Borsboom, Josine Verhagen, Rogier Kievit, Marjan Bakker, Angelique Cramer, Dora Matzke, Don Mellenbergh, and Han L. J. van der Maas]. Acta Psychol (Amst) [Internet]. 2014 May;148:188–94. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0001691814000304
- Fidler F, Thomason N, Cumming G, Finch S, Leeman J. Editors Can Lead Researchers to Confidence Intervals, but Can't Make Them Think. Psychol Sci [Internet]. 2004 Feb 6;15(2):119–26. Available from: http://journals.sagepub.com/doi/10.1111/j.0963-7214.2004.01502008.x
- Forstmeier W, Wagenmakers E, Parker TH. Detecting and avoiding likely false-positive findings – a practical guide. Biological Reviews [Internet]. 2017 Nov 23;92(4):1941–68. Available from: https://onlinelibrary.wiley.com/doi/10.1111/brv.12315