

Week 8 — Statistical analysis & error control

Is the field interpreting p-values correctly, reporting effect sizes and intervals, and controlling error across multiple tests?

Objective evidence review · companion to Week 8 · prepared July 2026

Scope & method. Rapid narrative review, not a systematic review. Structured search of the sport-science statistical-methods literature (PubMed, PLOS ONE, J. Science and Medicine in Sport, Journal of Sports Sciences). Findings reported as the cited studies state them; foundational statistical sources (ASA; estimation) are flagged as adjacent.

The principle under review

WEEK 8 · P-VALUES, ESTIMATION, AND MULTIPLICITY

Week 8 argues that a p-value is only the probability of the data under the null, not the probability the effect is real or important; that the .05 threshold is an arbitrary convention rather than a switch; that effect sizes with confidence intervals interpreted against a smallest effect of interest are more informative than significance alone; and that testing many outcomes inflates the error rate. The empirical questions are whether the field interprets significance correctly, reports estimation, uses valid inference methods, and controls multiplicity.

1 · How well is it currently applied?

THE EVIDENCE

just <.05

an excess of results falling just below p = .05 across 1,599 SEM articles — Borg et al. 2023

12–45%

Type I error rate produced by magnitude-based inference — Lohse et al. 2020

232

papers that used that statistically invalid inference method — Lohse et al. 2020

The literature is skewed toward statistical significance

Direct evidence shows the field is organised around the .05 threshold. Borg, Barnett, Caldwell, White and Stewart (2023) examined 3,819 confidence intervals from 1,599 articles across 18 sport and exercise medicine journals (2002–2022) and found a marked under-representation of results corresponding to p-values just above .05 and a corresponding excess just below it — the statistical signature of significance-chasing and publication bias. The authors note that an exclusive focus on statistical significance distorts the literature and exaggerates effect magnitudes, and recommend transparent practices such as Registered Reports.

An invalid inference method was used at scale

Beyond misinterpreting valid tests, part of the field adopted an inference method that is not valid. Sainani (2018) demonstrated that magnitude-based inference (MBI) inflates the Type I error rate relative to standard inference, and a systematic review by Lohse, Sainani and colleagues (2020) identified 232 papers using MBI — typically with very small samples — producing Type I error rates of roughly 12–45%. This is a concrete instance of the field reaching statistical conclusions its methods could not support; MBI use has since declined following these critiques.

Reporting and interpretation practices lag best practice

Estimation is still not the norm. Mesquida, Murphy, Lakens and Warne (2023) found the reporting of test statistics, effect sizes and confidence intervals in the *Journal of Sports Sciences* to be suboptimal, and Knudson (2017) documented recurrent statistical and reporting errors in biomechanics, including reliance on significance testing with small samples and inflated inferences. A further error-control problem is multiplicity: Gorman and Warne (2025) note that testing multiple dependent variables in several ways — without accounting for the number of tests — inflates the true false-positive rate well above the nominal 5%.

Source	Finding	Implication
--------	---------	-------------

Borg, Barnett, Caldwell, White & Stewart (2023) <i>J. Sci. Med. Sport</i>	Excess of results just below $p = .05$ across 1,599 SEM articles	The literature chases statistical significance
Sainani (2018); Lohse, Sainani et al. (2020) <i>MSSE; PLOS ONE</i>	MBI inflates Type I error (12–45%); used in 232 papers	An invalid inference method was applied at scale
Mesquida, Murphy, Lakens & Warne (2023) <i>J. Sports Sciences</i>	Suboptimal reporting of test statistics, effect sizes and CIs	Estimation is not yet standard practice
Knudson (2017) <i>Sports Biomechanics</i>	Recurrent statistical/reporting errors; inflated inferences from small samples	Misinterpretation extends across sub-disciplines
Gorman & Warne (2025) <i>J. Sports Sciences</i>	Multiple dependent variables tested without correction inflate the error rate	Multiplicity is under-controlled
Wasserstein & Lazar (2016) <i>Amer. Statistician (adjacent)</i>	p-values do not measure effect size or the probability the null is true	Defines the correct interpretation being missed

2 • Is there cause for concern?

OBJECTIVE ASSESSMENT

ASSESSMENT **Substantial concern**

Several independent problems compound. The published literature is skewed toward just-significant results; an invalid inference method (MBI) was used in hundreds of papers; effect sizes and confidence intervals are still under-reported relative to bare significance; and multiplicity is frequently uncontrolled. Together these mean that a “significant” result in the field is often over-interpreted — treated as evidence of a real, important, replicable effect when it may be none of these.

The corrective knowledge is, however, well established and increasingly visible. The ASA statement on p-values and the estimation (“new statistics”) approach are widely available; the MBI critiques have measurably reduced its use; and sport-specific meta-research (Borg et al., 2023; Mesquida et al., 2023) is now quantifying these problems and pushing estimation and transparency. The concern is substantial but actively being addressed.

3 • Immediate steps to improve

ACTIONABLE

1. **Report effect sizes with confidence intervals and interpret the whole interval** against a smallest effect of interest, rather than reporting a bare p-value or a “significant/not” verdict.
2. **Treat p as continuous evidence**, not a threshold that flips at .05; do not equate statistical significance with importance or replicability.
3. **Avoid magnitude-based inference** and other non-standard methods that inflate the error rate; use standard estimation and inference.
4. **Pre-specify one primary analysis and control for multiplicity**, reporting all other analyses as exploratory.
5. **Interpret a precise null as informative**: a tight confidence interval around zero is evidence that any effect is small, which a bare “non-significant” cannot convey.

References

- Borg, D. N., Barnett, A. G., Caldwell, A. R., White, N. M., & Stewart, I. B. (2023). The bias for statistical significance in sport and exercise medicine. *Journal of Science and Medicine in Sport*, 26(3), 164–168. <https://doi.org/10.1016/j.jsams.2023.03.002>
- Gorman, B. T., & Warne, J. (2025). Powering a study for more than one dependent variable: a letter to the editor regarding the editorial “sample size estimation revisited”. *Journal of Sports Sciences*. Advance online publication. <https://doi.org/10.1080/02640414.2025.2541432>
- Knudson, D. (2017). Confidence crisis of results in biomechanics research. *Sports Biomechanics*, 16(4), 425–433. <https://doi.org/10.1080/14763141.2016.1246603>

- Lohse, K. R., Sainani, K. L., Taylor, J. A., Butson, M. L., Knight, E. J., & Vickers, A. J. (2020). Systematic review of the use of "magnitude-based inference" in sports science and medicine. *PLOS ONE*, 15(6), e0235318. <https://doi.org/10.1371/journal.pone.0235318>
- Mesquida, C., Murphy, J., Lakens, D., & Warne, J. (2023). Publication bias, statistical power and reporting practices in the Journal of Sports Sciences: potential barriers to replicability. *Journal of Sports Sciences*. <https://doi.org/10.1080/02640414.2023.2269357>
- Sainani, K. L. (2018). The problem with "magnitude-based inference". *Medicine & Science in Sports & Exercise*, 50(10), 2166–2176. <https://doi.org/10.1249/MSS.0000000000001645>
- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: context, process, and purpose. *The American Statistician*, 70(2), 129–133. <https://doi.org/10.1080/00031305.2016.1154108>
- Cumming, G. (2014). The new statistics: why and how. *Psychological Science*, 25(1), 7–29. <https://doi.org/10.1177/0956797613504966>

Improving Research Practices in Sport & Exercise Science — a free short course for the Sports Science Replication Centre. Current State Review, Week 8. Figures are reported as stated in the cited sources; adjacent-field sources flagged. For education.