Week 7 — Sample size, power & precision

Are studies in the field large enough to answer their questions, and are sample sizes justified in advance?

Objective evidence review · companion to Week 7 · prepared July 2026

Scope & method. Rapid narrative review, not a systematic review. Structured search of the sport-science meta-research literature (PubMed, Journal of Sports Sciences, Sports Medicine, BMJ Open, SportRxiv), including several Sports Science Replication Centre (SSRC) papers. Findings reported as the cited studies state them; foundational statistical sources (neuroscience, psychology) are flagged as adjacent.

The principle under review

WEEK 7 · ADEQUATE POWER, JUSTIFIED IN ADVANCE

Week 7 argues that a small study gives a distorted answer, not merely a less precise one; that under-powered significant findings are systematically inflated (the winner's curse); and that sample size should be justified in advance against a smallest effect of interest. The empirical questions are how large studies in the field actually are, whether power is adequate, and whether sample sizes are justified.

1 · How well is it currently applied?

THE EVIDENCE

53%
average statistical power in the *Journal of Sports Sciences* (z-curve) — Mesquida et al. 2023

n = 19
median sample size in the same journal — Abt et al. 2020

−75%
median shrinkage of effect sizes on replication (the winner's curse) — Murphy et al. 2025

Studies are small and frequently under-powered

Direct field-specific evidence shows chronic under-powering. Mesquida, Murphy, Lakens and Warne (2023), applying a z-curve analysis to 89 studies in the *Journal of Sports Sciences*, estimated average observed power at 53% across all findings (61% for significant findings only) and found some indication of publication bias — well below the conventional 80% target, and lower still for the modest effects typical of the field. This sits on top of a median sample size of about 19 in the same journal (Abt et al., 2020). Editorial experience concurs: Impellizzeri, Murphy, Mesquida, Warne and colleagues (2025) note that most intervention studies submitted to their journal are under-powered, motivating a dedicated "Preliminary Report" submission category for small-sample studies.

The winner's curse is measurable in the field

Under-powering does not merely cause missed effects; it inflates the effects that do reach significance. In the field's replication project, Murphy and colleagues (2025) found that 88% of original effect sizes shrank on replication, with a median reduction of 75% — the pattern expected when significant results from small studies are inflated by chance (Type-M errors; Button et al., 2013; Gelman & Carlin, 2014). A literature built substantially on small, significant studies therefore over-states effect sizes systematically.

Sample sizes are rarely justified in advance

Pre-specification of sample size is uncommon. Mesquida and colleagues (2023) found the usage and reporting of a-priori power analyses to be suboptimal, with sample sizes often set by heuristics ("20 per condition") rather than justified against a target effect. Schulz and colleagues (2022) similarly found that while 60% of sports-medicine trials reported a sample-size calculation, only 32% justified the expected effect size on which such a calculation depends. A further, subtler problem is that studies are frequently powered for a single outcome yet analysed across several: Gorman and Warne (2025) show that powering and testing across multiple dependent variables inflates the Type I error rate unless the multiplicity is accounted for.

Source	Finding	Implication

Mesquida, Murphy, Lakens & Warne (2023) *J. Sports Sciences*	Average power 53% (61% for significant findings); publication bias; suboptimal a-priori power use	Designs are under-powered and rarely justified
Abt et al. (2020) *J. Sports Sciences*	Median sample size of 19	Typical studies are small
Impellizzeri, Murphy, Mesquida, Warne et al. (2025) *Sci. Med. Football*	Most submitted intervention studies are under-powered	Under-powering is the norm, not the exception
Murphy et al. (2025) *Sports Medicine*	88% of effects shrank on replication (median −75%)	Winner's curse is measurable in the field
Schulz et al. (2022) *BMJ Open*	60% reported a sample-size calculation; only 32% justified the effect size	Justification is often absent even when a calculation is reported
Gorman & Warne (2025) *J. Sports Sciences*	Powering/testing across multiple dependent variables inflates Type I error	Single-outcome power under-states the true error rate
Button et al. (2013) *Nat. Rev. Neurosci. (adjacent)*	Low power inflates significant estimates and reduces reliability	Mechanism behind field-specific findings

2 · Is there cause for concern?

OBJECTIVE ASSESSMENT

ASSESSMENT **Substantial concern**

The evidence is direct and field-specific: typical studies are small (median n ≈ 19), average power is around half rather than the conventional 80%, sample sizes are seldom justified against a target effect, and the predicted consequence — inflated significant effects — is measurable in the field's own replication data (88% of effects shrank on replication). Under these conditions, a significant finding from a typical study is likely to over-state the true effect, and a non-significant one is uninformative.

Two mitigating points apply. First, small samples are often genuinely unavoidable in elite sport; the failure is under-powering that is not acknowledged, not the small sample itself. Second, the field is developing concrete responses — sample-size justification frameworks, multi-site collaboration, and new journal formats such as the Preliminary Report for small-sample studies — which address the problem without pretending the constraint does not exist.

3 · Immediate steps to improve

ACTIONABLE

1. **Run an a-priori power analysis against a smallest effect of interest,** not against a heuristic sample size, and report it.
2. **Justify the sample size honestly** using one of the recognised approaches (a-priori power, resource-constrained plus sensitivity analysis, planning for precision, or an explicit statement of no justification).
3. **Account for multiplicity** when powering and analysing more than one dependent variable, rather than powering for a single outcome and testing many.
4. **When the sample must be small, say so,** report the power and a sensitivity analysis, and flag the winner's-curse risk on any significant finding rather than presenting it as definitive.
5. **Pool across sites or reframe as a preliminary report** when a single lab cannot reach adequate power, rather than publishing an under-powered study as conclusive.

References

Abt, G., Boreham, C., Davison, G., et al. (2020). Power, precision, and sample size estimation in sport and exercise science research. *Journal of Sports Sciences, 38*(17), 1933–1935. https://doi.org/10.1080/02640414.2020.1776002

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365–376. https://doi.org/10.1038/nrn3475

Gelman, A., & Carlin, J. (2014). Beyond power calculations: assessing Type S (sign) and Type M (magnitude) errors. *Perspectives on Psychological Science, 9*(6), 641–651. https://doi.org/10.1177/1745691614551642

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

Impellizzeri, F. M., Murphy, J., Mesquida, C., Warne, J., Hecksteden, A., Batomen, B., Wang, C., Meyer, T., & Lakens, D. (2025). Introducing a new "Preliminary Report" submission category for small-sample intervention studies: rationale and instructions. *Science and Medicine in Football.* Advance online publication. https://doi.org/10.1080/24733938.2025.2580319

Lakens, D. (2022). Sample size justification. *Collabra: Psychology, 8*(1), 33267. https://doi.org/10.1525/collabra.33267

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

Murphy, J., Caldwell, A. R., Mesquida, C., et al. (2025). Estimating the replicability of sports and exercise science research. *Sports Medicine, 55*(10), 2659–2679. https://doi.org/10.1007/s40279-025-02201-w

Schulz, R., Langen, G., Prill, R., Cassel, M., & Weissgerber, T. L. (2022). Reporting and transparent research practices in sports medicine and orthopaedic clinical trials: a meta-research study. *BMJ Open, 12*(8), e059347. https://doi.org/10.1136/bmjopen-2021-059347
