Week 1 — The reliability of the evidence base

How well is the field applying the principle that confidence should be calibrated to evidence, and that single studies are provisional?

Objective evidence review · companion to Week 1 · prepared July 2026

Scope & method. This is a rapid, narrative evidence review, not a systematic review. It draws on a structured search of the metascience and sport-science methodology literature (PubMed, Google Scholar, PLOS ONE, and the Sports Science Replication Centre) for empirical estimates of research reliability, statistical practice, and evidence use in sport and exercise science. Findings are reported as the cited studies state them; where field-specific evidence is limited this is noted.

The principle under review

WEEK 1 · ERROR-REDUCTION AND CALIBRATED CONFIDENCE

Week 1 argues that scientific claims should be believed in proportion to the evidence supporting them, that a single study is one estimate measured with error rather than a settled result, and that reliable knowledge accumulates across replications and syntheses. The empirical questions for the field are therefore: how reliable is the published evidence base, how sound are the statistical inferences behind it, and do researchers and practitioners weight evidence accordingly?

1 · How well is it currently applied?

THE EVIDENCE

28%
of applied findings robustly replicated (7 of 25) — Murphy et al. 2025

81%
positive-result rate in kinesiology journals (300 articles) — Twomey et al. 2021

232
papers using a statistically invalid inference method (MBI) — Lohse et al. 2020

Direct replication evidence

The field's first systematic replication project (Murphy, Caldwell, Mesquida et al., 2025) attempted to replicate 25 findings from applied studies in first-quartile journals (2016–2021). Only 28% (7 of 25) robustly replicated against all three pre-set criteria; 88% of original effect sizes were smaller on replication, with a median magnitude reduction of 75%. This is a direct, quantitative estimate that individual published effects are frequently non-robust and over-stated.

Publication bias and positive-result rates

Two independent lines of evidence show the literature is skewed toward positive findings. Twomey and colleagues (2021), examining 300 articles from flagship kinesiology journals, found that roughly 81% of studies testing a hypothesis reported positive (supportive) results — a rate the authors describe as implausibly high — alongside heavy reliance on significance testing (~92%) and limited trial registration and data availability. Mesquida, Murphy, Lakens and Warne (2022) applied a z-curve analysis to applied studies in the *Journal of Sports Sciences* and reported evidence of publication bias and of statistically underpowered designs. Caldwell and colleagues (2020), writing for the Consortium for Transparency in Exercise Science, identify the same publication bias toward positive findings and associated questionable research practices as central problems. Publication bias inflates the average published effect and is inherited by meta-analyses built on the biased record.

Statistical rigour of the inferences

Beyond selective publication, the inferential methods themselves have been unreliable in parts of the field. Sainani (2018) demonstrated that magnitude-based inference (MBI) — widely adopted in sports science — inflates the Type I (false-positive) error rate; a systematic review by Lohse, Sainani and colleagues (2020) found 232 papers using MBI, with very small median sample sizes (10–14) and Type I error rates of roughly 12–45%, and concluded the method should be discontinued. Independently, Knudson (2017) documented that biomechanics research commonly combines small samples with incorrect statistical analyses, producing inflated effect magnitudes. Halperin, Vigotsky, Foster and Pyne (2018) catalogue the same underlying weaknesses across the field: inadequate validation of surrogate outcomes, few longitudinal and replication studies, limited reporting of null or trivial results, and insufficient transparency.

Evidence use in practice

Evidence-based practice is widely endorsed but unevenly applied. In a worldwide survey of 107 basketball practitioners (79% strength-and-conditioning coaches), Pernigoni and colleagues (2022) found the most-cited barriers to applying evidence were practical — devices/facilities (51%), cost (51%), time (27%), players' perceptions (25%) — with only 16% citing insufficient evidence. The research-to-practice gap is therefore driven more by implementation constraints than by evidence availability, which places the onus on appraising the evidence that is used.

Source	Finding	Implication
Murphy, Caldwell, Mesquida et al. (2025) *Sports Medicine*	28% of 25 applied findings robustly replicated; 88% of effects shrank (median −75%)	A single published finding is frequently non-robust and inflated
Twomey et al. (2021) *Communications in Kinesiology*	~81% positive-result rate across 300 kinesiology articles; ~92% used significance testing; limited registration/data	The published record is skewed toward positive results
Mesquida, Murphy, Lakens & Warne (2022) *R. Soc. Open Sci.*	Evidence of publication bias and underpowered designs (z-curve)	Published effects are, on average, over-estimates
Caldwell et al. (2020) *Sports Medicine*	Publication bias toward positive findings; associated questionable research practices	Structural bias in what gets published
Lohse, Sainani et al. (2020) *PLOS ONE*	232 papers used MBI; median n=10–14; Type I error ~12–45%	An invalid inference method produced over-interpreted results at scale
Sainani (2018) *Med. Sci. Sports Exerc.*	MBI inflates the false-positive rate versus standard inference	Some widely used conclusions rest on a flawed method
Knudson (2017) *Sports Biomechanics*	Small samples + incorrect analyses → inflated effect magnitudes	Reliability problems extend across sub-disciplines
Halperin et al. (2018) *IJSPP*	Few replication/longitudinal studies; poor null reporting; weak transparency	Structural weaknesses limit accumulation of reliable knowledge
Pernigoni et al. (2022) *Front. Physiol.*	Barriers to evidence use are mainly practical; only 16% cite insufficient evidence	The research–practice gap is largely about implementation

2 · Is there cause for concern?

OBJECTIVE ASSESSMENT

ASSESSMENT **Substantial concern**

The evidence converges from several independent directions — a low direct replication rate, an implausibly high positive-result rate, documented publication bias and low power, and the large-scale use of at least one statistically invalid inference method. Together these indicate that a substantial proportion of published applied findings would not replicate at their reported magnitude, and that treating individual studies — particularly striking, novel ones — as reliable guides to practice is not supported by the current state of the evidence.

Three qualifications keep the assessment proportionate. First, the direct replication estimate rests on 25 studies and should be read as an initial figure with its own uncertainty. Second, MBI use has declined following the critiques above, so part of this evidence describes a problem now being corrected. Third, the existence of these analyses reflects the field measuring its own reliability, which is a precondition for improvement. The evidence does not imply that most conclusions in the field are false; it implies that confidence in any single unreplicated finding should be modest, and that the field's average published effect is likely inflated.

3 · Immediate steps to improve

ACTIONABLE

1. **Weight evidence by its strength.** Give replicated and meta-analytic evidence more weight than any single study, and treat a lone novel finding as provisional until independently reproduced.
2. **Check the inference method, not just the p-value.** Be sceptical of results based on magnitude-based inference; look for standard effect sizes with confidence intervals (Weeks 7–8).
3. **Check power and precision before believing an effect** — note the sample size, the confidence interval, and whether the design could plausibly detect the claimed effect.
4. **Discount single striking findings pending replication**, particularly large effects from small samples, which are the most likely to be inflated.
5. **Prefer preregistered and registered-report evidence** where available, as it is less subject to publication bias and selective reporting (Week 6).
6. **When applying evidence in practice, appraise it explicitly** rather than adopting findings because they are recent or highly cited; separate implementation barriers from evidential ones.

References

Caldwell, A. R., Vigotsky, A. D., Tenan, M. S., et al. (2020). Moving sport and exercise science forward: a call for the adoption of more transparent research practices. *Sports Medicine, 50*(3), 449–459. https://doi.org/10.1007/s40279-019-01227-1

Halperin, I., Vigotsky, A. D., Foster, C., & Pyne, D. B. (2018). Strengthening the practice of exercise and sport-science research. *International Journal of Sports Physiology and Performance, 13*(2), 127–134. https://doi.org/10.1123/ijspp.2017-0322

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. (2022). Replication concerns in sports and exercise science: a narrative review of selected methodological issues in the field. *Royal Society Open Science, 9*(12), 220946. https://doi.org/10.1098/rsos.220946

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

Pernigoni, M., Conte, D., Calleja-González, J., Boccia, G., Romagnoli, M., & Ferioli, D. (2022). The application of recovery strategies in basketball: a worldwide survey. *Frontiers in Physiology, 13*, 887507. https://doi.org/10.3389/fphys.2022.887507

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

Twomey, R., Harlley, S., Romero Medina, C., et al. (2021). The nature of our literature: a registered report on the positive result rate and reporting practices in kinesiology. *Communications in Kinesiology, 1*(3). https://doi.org/10.51224/cik.v1i3.43