Making Cautious & Responsible Claims

Companion to the Week 11 episode. Say exactly what your evidence supports — no more, no less.

1. The one idea

The last place good research turns into irresponsible communication: the claim quietly outgrows the data. **Calibrate the claim to the evidence** — four disciplining questions.

2. The four-question calibration

Question	The trap → the responsible claim
1. Practical vs statistical significance	**"It was significant" ⇒ it matters.** With large N a trivially tiny effect is highly significant. → Claim importance only when the effect (and the *lower end of its CI*) clears your **SESOI**.
2. External validity	**Studied trained young men → "this works" (for everyone).** → **Bound the claim**: "in trained young men, over 4 weeks, in a lab…". Extending it is a *hypothesis*.
3. Provisionality (regression)	**"This study demonstrates…".** A single study's effect is, on average, *larger* than its replication (winner's curse). → **Probabilistic** language: "preliminary evidence that X *may*… needs replication."
4. Causal language ↔ design	**Observational data → "X improves/causes Y".** → **Associational** language ("associated with", "predicted"); reserve causal verbs for *randomised* designs.

3. Why provisionality? — a field-specific number

> The SSRC's own project (Murphy, Caldwell, Mesquida et al., *Sports Medicine*, 2025) found only **~a quarter** of attempted replications of applied sport-science findings robustly held (significant, same direction, compatible size). Default attitude to any single fresh finding — **including your own** — is *humility*: it's probably an over-estimate.

4. Why causal language matters in practice

A causal claim licenses an **intervention** ("do X to get Y"). If the causation isn't real, you send coaches & athletes chasing a mirage. **Never let a causal verb attach to non-causal evidence.**

5. Do this — this week

Calibrate a claim

Take a main claim (yours or one you're reading) and run the four questions:

1. Big enough to matter (effect + lower CI vs SESOI), or just significant?

2. Bounded by the sample/conditions, or over-generalised?

3. Provisional (one study, will likely shrink), or definitive?

4. Causal verb matched to a randomised design, or attached to association?

Rewrite it to say *exactly* what the evidence supports — usually quieter, more bounded, more trustworthy.

References

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values. *The American Statistician, 70*(2), 129–133.

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.

Gelman, A., & Carlin, J. (2014). Beyond power calculations. *Perspectives on Psychological Science, 9*(6), 641–651.

Mesquida, C., Murphy, J., Lakens, D., & Warne, J. (2022). Replication concerns in sports and exercise science. *Royal Society Open Science, 9*(12), 220946.