Week 4 — Hypothesis generation & HARKing

How often are hypotheses genuinely specified in advance, rather than constructed after the results are known?

Objective evidence review · companion to Week 4 · prepared July 2026

Scope & method. Rapid narrative review, not a systematic review. Structured search of the sport-science and meta-research literature (PubMed, BJSM, Sports Medicine, PLOS ONE) for evidence on hypothesis pre-specification, supported-hypothesis rates, and questionable research practices (QRPs). Findings reported as the cited studies state them; two adjacent-field prevalence surveys (psychology; ecology/evolution) are flagged, as direct self-report data are scarce in sport science.

The principle under review

WEEK 4 · FALSIFIABLE, PRE-SPECIFIED PREDICTIONS

Week 4 argues that a hypothesis should be a falsifiable, directional prediction, tied to a mechanism and specified before the data are seen, and that presenting a post-hoc hypothesis as if it were predicted in advance (HARKing) invalidates the statistics that assume one prediction tested once. The empirical questions are how often hypotheses are genuinely pre-specified, and whether the field shows the signatures of HARKing and related practices.

1 · How well is it currently applied?

THE EVIDENCE

82%
of sport & exercise medicine studies reported supported hypotheses — Büttner et al. 2020

51%
of researchers self-reported HARKing at least once (ecology/evolution) — Fraser et al. 2018

10%
of sports-science articles were preregistered (2024) — Bennett et al. 2026

Supported-hypothesis rates are implausibly high

The clearest field-specific signal is indirect but striking. Büttner and colleagues (2020) analysed 129 sport and exercise medicine studies and found 82.2% reported hypotheses supported by their results (70.8% full support of the primary hypothesis). Because a well-powered, unbiased literature testing genuinely uncertain predictions could not plausibly confirm hypotheses at this rate, the authors interpret it as evidence of widespread questionable research practices — including HARKing, p-hacking and cherry-picking. Consistent figures appear elsewhere: Twomey and colleagues (2021) reported an ~81% positive-result rate across 300 kinesiology articles, and Schulz and colleagues (2022) a 61% hypothesis-support rate in sports-medicine trials.

Self-reported HARKing is common where it has been measured

Direct prevalence data are scarce in sport science, but adjacent fields are informative. In a survey of 807 ecology and evolution researchers, Fraser and colleagues (2018) found 51% admitted presenting an unexpected finding as though it had been hypothesised from the start (HARKing), 64% selective reporting, and 42% optional stopping. John, Loewenstein and Prelec (2012), surveying over 2,000 psychologists with incentives for honest reporting, concluded that some QRPs may approach being the prevailing norm. There is no strong reason to expect sport science to differ, and its supported-hypothesis rates are consistent with similar behaviour.

Vague hypotheses and analytic multiplicity make the practice productive

Work from the SSRC identifies the mechanism that makes HARKing and related practices effective at producing false positives. Gorman and Warne (2025) note that studies are frequently powered and analysed around *multiple* dependent variables, tested in several ways — which, knowingly or not, gives researchers the best opportunity to find a significant effect and inflates the Type I error rate. In a related working paper, Gorman and colleagues (in preparation) argue that vague hypotheses which fail to specify effect direction or magnitude, combined with this analytic multiplicity, undermine deductive clarity: a result can be presented as confirming a prediction that was never precise enough to be wrong. Both point to the same corrective as HARKing — a single, directional, pre-specified hypothesis with a defined outcome and analysis. That the field recognises these problems is shown by a survey of 511 researchers (Murphy, Mesquida & Warne, 2023) in which 42% perceived a significant crisis of reproducibility or replicability and a further 36% a slight one.

Pre-specification, which prevents HARKing, is rare but growing

HARKing is prevented by committing to a hypothesis in advance, yet preregistration remains uncommon. Bennett and colleagues (2026), examining 2,006 sports-science articles from 19 journals published in 2024, found only 10.0% were preregistered and 60.0% stated a hypothesis; their analysis evaluates preregistration specifically as a means of reducing the inflated proportion of supported hypotheses. Schulz and colleagues (2022) similarly found only 20% of sports-medicine trials were prospectively registered. Low pre-specification leaves little protection against HARKing, but the emergence of these studies — and of preregistration itself — indicates the problem is being addressed.

Source	Finding	Implication
Büttner, Toomey, McClean, Roe & Delahunt (2020) *Br. J. Sports Med.*	82.2% of 129 SEM studies reported supported hypotheses (70.8% full primary support)	Rate is implausibly high; consistent with widespread QRPs
Twomey et al. (2021) *Comms. in Kinesiology*	~81% positive-result rate across 300 kinesiology articles	Corroborates the inflated positive/supported rate
Fraser et al. (2018) *PLOS ONE (adjacent)*	51% HARKing, 64% selective reporting, 42% optional stopping (self-report)	HARKing is common where directly measured
John, Loewenstein & Prelec (2012) *Psychological Science (adjacent)*	QRPs surprisingly prevalent among >2,000 psychologists	Some QRPs may be a prevailing norm
Bennett et al. (2026) *Sports Medicine*	10.0% of 2,006 sports-science articles preregistered; 60.0% stated a hypothesis	Little pre-specification to prevent HARKing
Schulz et al. (2022) *BMJ Open*	20% of sports-med trials prospectively registered; 61% hypothesis-support rate	Corroborates low pre-specification
Gorman & Warne (2025) *J. Sports Sciences*	Powering/analysing around multiple dependent variables inflates the Type I error rate	Analytic multiplicity is a mechanism for false positives
Gorman et al. (in prep.) *SSRC working paper — unpublished*	Vague hypotheses + multiplicity undermine deductive clarity	Imprecise predictions cannot be genuinely disconfirmed
Murphy, Mesquida & Warne (2023) *Comms. in Kinesiology*	Of 511 researchers, 42% perceive a significant crisis, 36% a slight crisis	The field itself recognises the problem
Mesquida et al. (2022) *R. Soc. Open Sci.*	Publication bias and underpowered designs in the field	Same conditions that make HARKing productive of false positives

2 · Is there cause for concern?

OBJECTIVE ASSESSMENT

ASSESSMENT **Substantial concern**

Multiple field-specific indicators (supported-hypothesis rates of 61–82%, ~81% positive results) are difficult to reconcile with genuinely pre-specified tests of uncertain predictions, and adjacent-field self-report data show HARKing and related practices are common. Combined with very low preregistration, this indicates that a substantial fraction of "predictions" in the literature are unlikely to have been genuine a-priori predictions, which inflates apparent evidence and reduces replicability.

Two qualifications apply. First, high supported-hypothesis rates are indirect evidence — they are consistent with HARKing but also with publication bias and low power (Weeks 1, 7), which act together. Second, direct HARKing prevalence in sport science is inferred from adjacent fields rather than measured, so the exact rate is uncertain. The direction of the evidence is nonetheless clear, and the corrective — preregistration — is well defined and being adopted.

3 · Immediate steps to improve

ACTIONABLE

1. **Write the hypothesis before collecting data,** as a falsifiable, directional prediction derived from a stated mechanism, and record the result that would count as disconfirmation.
2. **Preregister the primary hypothesis and analysis** so the prediction is time-stamped and cannot be revised after the results (Week 6).
3. **Label post-hoc hypotheses as exploratory.** Findings discovered in the data are legitimate as hypotheses for future confirmatory tests, provided they are reported as such.
4. **Set a smallest effect of interest in advance,** so the prediction specifies not just direction but a magnitude that would matter (Week 7).
5. **As reviewers and editors, do not request post-hoc hypotheses** presented as a priori; this is the passive form of HARKing.

References

Bennett, H., Blake, H., d'Unienville, N., et al. (2026). The uptake of preregistration, and its impact on the proportion of supported hypotheses, in sports science. *Sports Medicine*. Advance online publication. https://doi.org/10.1007/s40279-026-02451-2

Büttner, F., Toomey, E., McClean, S., Roe, M., & Delahunt, E. (2020). Are questionable research practices facilitating new discoveries in sport and exercise medicine? The proportion of supported hypotheses is implausibly high. *British Journal of Sports Medicine, 54*(22), 1365–1371. https://doi.org/10.1136/bjsports-2019-101863

Fraser, H., Parker, T., Nakagawa, S., Barnett, A., & Fidler, F. (2018). Questionable research practices in ecology and evolution. *PLOS ONE, 13*(7), e0200303. https://doi.org/10.1371/journal.pone.0200303

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

Gorman, B. T., et al. (in preparation). Testing vague hypotheses: how statistical multiplicity undermines deductive clarity in sport and exercise science. *Sports Science Replication Centre working paper. Unpublished.*

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Murphy, J., Mesquida, C., & Warne, J. (2023). A survey on the attitudes towards and perception of reproducibility and replicability in sports and exercise science. *Communications in Kinesiology, 1*(5), article 53. https://doi.org/10.51224/cik.2023.53

Kerr, N. L. (1998). HARKing: hypothesizing after the results are known. *Personality and Social Psychology Review, 2*(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4

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

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

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