

Patient Experience and Measurement Discordance

Abstract

Subjective benefit, physiological signals, treatment goals create distinct evidence obligations in subjective and objective sleep evidence. A bounded review of the literature is used to formulate, rather than claim to complete, an evaluation. The focal evidence concerns whether subjective sleep gains in empty nose syndrome are sufficient to support claims about objectively improved sleep. Different outcomes can matter simultaneously. A patient-reported improvement and an unchanged physiological measure can each be real and relevant. Prespecify outcome roles and inspect discordance. Preserve patient priorities while withholding unsupported physiological or efficacy conclusions. The proposed comparison concerns subjective benefit, physiological signals, treatment goals; it would test whether the specific failure scenario is detected without erasing legitimate uncertainty. The original source's limits remain material: Its argument cannot establish the efficacy or inefficacy of a particular treatment, and measurement discordance does not invalidate patient-reported experience.

The interface under examination

An evidence workflow can be fast, reproducible, and still answer the wrong question. In this review, subjective benefit provides the entry point for testing that possibility against treatment goals. Consider the possibility of discordance treated as proof that one outcome is worthless. This is a proposed stress scenario, not an observed event reported by the cited studies. Its purpose is to expose which evidence contract would be violated even if the final output looked plausible. For patient-reported experience and independent measurements, the relevant claim must identify its population or artifact class, its time window, and the action it is supposed to support. A failure at any of these levels may persist even after a local technical improvement. The analysis therefore treats published methods as bounded evidence and asks how their limitations interact with this particular decision. An operational analogy is useful only if it keeps those limitations visible.

Different outcomes can matter simultaneously. A patient-reported improvement and an unchanged physiological measure can each be real and relevant. The claim should specify whose objective is being evaluated and avoid forcing all benefits onto one scale. In this article, the relevant contrast is subjective benefit, physiological signals, treatment goals. The proposed counterexample of discordance treated as proof that one outcome is worthless puts that contrast under pressure. Prespecify outcome roles and inspect discordance. Preserve patient priorities while withholding unsupported physiological or efficacy conclusions. A favorable finding would show that the intended interpretation of treatment goals remains defensible when subjective benefit is perturbed under the specified conditions. An unfavorable finding should narrow the claim rather than be discarded as an inconvenient case.

A reviewer of patient experience and measurement discordance should be able to recover a claim before judging an output. The illustrative case is discordance treated as proof that one outcome is worthless. The input record should retain enough information to decide which part of that scenario violates the intended contract. That includes the original source type, measurement or artifact status, and the reason the proposed interpretation exceeds its support. The evaluator would compare the system's response with a qualified judgment made without seeing that response. Disagreement should be classified rather than erased by rewriting the prompt. The claim should specify whose objective is being evaluated and avoid forcing all benefits onto one scale. A useful negative result would show that the interface cannot preserve this

distinction under realistic use; a useful positive result would demonstrate preservation on independent cases rather than merely on the example used to develop the rule. The primary source remains relevant through its particular contribution: The note separates the importance of patient experience from the evidentiary requirements for physiological claims. That contribution should be retained even if the integrated workflow fails. Conversely, successful integration would not remove the source's limitations or turn a proposed extension into an already observed result. This asymmetric interpretation protects both the literature and the user from an unjustified all-or-nothing verdict.

Evidence without scope inflation

(Xiao Yu et al. 2025) examines whether subjective sleep gains in empty nose syndrome are sufficient to support claims about objectively improved sleep in *Subjective Sleep Gains in Empty Nose Syndrome: A Cautionary Note on the Need for Objective Validation*. Its methodological basis is a brief clinical correspondence emphasizing the need for objective validation alongside reported benefit. The accessible evidence is specific: the official bibliographic record identifies a two-page cautionary note, not a newly completed trial or an independent sleep cohort. The note separates the importance of patient experience from the evidentiary requirements for physiological claims. Its argument cannot establish the efficacy or inefficacy of a particular treatment, and measurement discordance does not invalidate patient-reported experience. The source supplies one part of the comparison involving subjective benefit, physiological signals, treatment goals. Its evidence should retain its original unit and scope; the present synthesis proposes a connection without claiming that the paper directly tested this integrated workflow.

A useful boundary case is supplied by *Multifunctional Hydrogel Sensors with Dynamic Covalent Networks for Machine Learning-Assisted Parkinson's Disease Diagnosis and Encrypted Human-Computer Interaction*, concerned with how a multifunctional dynamic-covalent hydrogel can support movement sensing and machine-learning-assisted Parkinson-related recognition (Siqi Ding et al. 2025). Its methodological basis is a multicomponent hydrogel network combining polyacrylic acid, oxidized carboxymethylcellulose, GelMA, and lignosulfonate-derived functionality with sensor-based classification. The accessible evidence is specific: the abstract reports adhesive, extensible, conductive behavior and movement recognition, including Morse-like finger-amplitude messages. It links material mechanics to usable sensing and interaction rather than evaluating a material solely by one mechanical property. Laboratory classification is not validated clinical diagnosis, durability under sweat and repeated use requires testing, and message encoding alone is not cryptographic security evaluation. The source supplies one part of the comparison involving subjective benefit, physiological signals, treatment goals. Its evidence should retain its original unit and scope; the present synthesis proposes a connection without claiming that the paper directly tested this integrated workflow.

The sources should be compared through the roles they play, not merged into a single performance claim. *Subjective Sleep Gains in Empty Nose Syndrome: A Cautionary Note on the Need for Objective Validation* addresses whether subjective sleep gains in empty nose syndrome are sufficient to support claims about objectively improved sleep (Xiao Yu et al. 2025). By contrast, *Multifunctional Hydrogel Sensors with Dynamic Covalent Networks for Machine Learning-Assisted Parkinson's Disease Diagnosis and Encrypted Human-Computer Interaction* addresses how a multifunctional dynamic-covalent hydrogel can support movement sensing and machine-learning-assisted Parkinson-related recognition (Siqi Ding et al. 2025). That difference matters for plural outcomes: the evidence concerning subjective benefit is not automatically evidence concerning treatment goals. A successful software operation, a plausible mechanistic account, and a validated outcome are different achievements. Their integration may improve traceability, but it cannot create an unmeasured link. A coherent review should identify where the sources complement one another

and where their scopes remain incompatible. Here the proposed failure scenario is discordance treated as proof that one outcome is worthless. The analytical value of that scenario is that it can occur without requiring any cited paper to be wrong: an otherwise valid local finding can be misused at a later interface. Preventing that misuse requires evidence about the interface itself, not simply a larger collection of local results.

Checks for the proposed workflow

Causal inference requires a clearly specified intervention, time order, and exchangeability assumptions (Hernán and Robins 2020). A prediction can be accurate without identifying what an intervention would change. A useful causal analysis begins by defining the hypothetical comparison, eligibility, and time zero. Its assumptions concern the process generating exposure and observation, not merely the flexibility of a fitted model. In this review, causal vocabulary is appropriate only when that comparison can be specified and defended. This is methodological support for examining plural outcomes, not a claim that the source studied subjective and objective sleep evidence. A separate application-specific test is required at the interface described in this review.

Epidemiologic design makes confounding, selection, and measurement error part of interpretation (Rothman, Greenland, and Lash 2008). Confounding, selection, and measurement error are distinct problems and need not move an estimate in the same direction. An analysis should describe them before deciding which adjustment is adequate. A larger dataset can reduce random error while leaving systematic error almost untouched. Precision should therefore be interpreted alongside the quality of the design and the eligibility of its population. Within subjective and objective sleep evidence, this distinction is relevant to plural outcomes. The proposed failure scenario is discordance treated as proof that one outcome is worthless; its evaluation must preserve the original source's methodological scope.

False-discovery control is central when many molecular features are screened (Benjamini and Hochberg 1995). Multiplicity arises from the family of questions actually searched. Controlling false discoveries within a named family does not automatically cover later tissue filters, alternative endpoints, or narrative choices. The proposed audit should retain exploratory branches so the family is visible. Independent confirmation then addresses a different question from whether a discovery threshold was correctly applied during screening. The connection to patient experience and measurement discordance is a constraint on interpretation. It would be inappropriate to borrow the source's methodological authority as direct confirmation of the integrated domain claim.

Medical machine learning must fit data provenance, care pathways, and prospective evaluation (Rajkomar, Dean, and Kohane 2019). The clinical relevance of a model depends on how its output changes a care process. Development data can contain information that becomes available only after the intended decision. A review must identify that availability boundary before comparing algorithms. Prospective use also changes who is measured and who is treated, so evaluation of a frozen historical dataset cannot establish the entire future workflow. Within subjective and objective sleep evidence, this distinction is relevant to plural outcomes. The proposed failure scenario is discordance treated as proof that one outcome is worthless; its evaluation must preserve the original source's methodological scope.

Responsible health AI connects prediction with workflow, monitoring, and patient impact (Wiens et al. 2019). The point at which a model reaches practice is not the end of evaluation. New populations, changing measurements, and feedback can alter failure modes. A responsible deployment needs a route for recognizing deterioration and a person authorized to respond. Monitoring should be linked to feasible corrective actions; an alert that cannot change the workflow provides accountability only in a very limited sense. Within

subjective and objective sleep evidence, this distinction is relevant to plural outcomes. The proposed failure scenario is discordance treated as proof that one outcome is worthless; its evaluation must preserve the original source's methodological scope.

Calibration separates discrimination from confidence used in clinical escalation (Guo et al. 2017). Discrimination and calibration deserve separate evaluation. A score can order cases well yet systematically overstate its certainty. Aggregate calibration can also hide local miscalibration near the threshold where action changes. An audit should retain predicted probabilities and observed outcomes rather than report only correct classifications, and it should evaluate calibration on data not reused to choose the correction. The connection to patient experience and measurement discordance is a constraint on interpretation. It would be inappropriate to borrow the source's methodological authority as direct confirmation of the integrated domain claim.

Uncertainty can deteriorate under the site and cohort shifts common in biomedicine (Ovadia et al. 2019). Uncertainty is most valuable when a system moves beyond familiar conditions, yet those conditions can also weaken its uncertainty estimates. A warning mechanism must therefore be evaluated under shift rather than assumed to generalize because it works on ordinary held-out data. Different shifts can affect ranking, calibration, and error concentration differently. No single in-distribution curve demonstrates protection against them all. An implementation should expose this limitation to reviewers of patient-reported experience and independent measurements. The result and its evidence qualification should remain recoverable even when the interface condenses them into a short report.

Selective prediction formalizes abstention when evidence does not support action (Geifman and El-Yaniv 2017). Abstention creates a tradeoff between coverage and residual error. Comparing systems without matching coverage can make one look safer merely because it leaves more work unfinished. Deferred cases also consume human or computational resources. The evaluation should show the risk-coverage relation and account for the fate of excluded cases, rather than treating them as though they have ceased to matter. This is methodological support for examining plural outcomes, not a claim that the source studied subjective and objective sleep evidence. A separate application-specific test is required at the interface described in this review.

Feature attribution can guide audit but does not establish biological causality (Lundberg and Lee 2017). Attribution decomposes a model output according to a defined explanatory convention. That decomposition can reveal dependence on suspicious features without establishing why the outcome occurs in the world. Reference choices and feature dependence can affect interpretation. A useful audit should compare explanations with known model behavior and should not transform a feature contribution into a treatment or mechanism claim. For the comparison involving subjective benefit, physiological signals, treatment goals, the source supplies an analytical obligation. The review's proposed application remains a hypothesis, with the deliberately difficult scenario serving as a check on scope inflation.

Health-AI governance requires accountability, transparency, inclusion, and protection of autonomy (World Health Organization 2021). Health-AI governance concerns people affected by a system as well as the organization operating it. Transparency without a route to challenge an output is incomplete. Autonomy and inclusion require attention to who can obtain follow-up, who is represented in evaluation, and who bears false alarms. These obligations remain even when an assistance tool is described as advisory rather than decisive. An implementation should expose this limitation to reviewers of patient-reported experience and independent measurements. The result and its evidence qualification should remain recoverable even when the interface condenses them into a short report.

Research priorities and negative results

A research agenda should address the specific negative case of discordance treated as proof that one outcome is worthless. Prespecify outcome roles and inspect discordance. Preserve patient priorities while withholding unsupported physiological or efficacy conclusions. These procedures are proposed, not completed. The case set would include the scenario of discordance treated as proof that one outcome is worthless, with its validity judged independently of the system's response. The relevant comparison axes are subjective benefit, physiological signals, treatment goals.

A measurement-focused protocol would compare the intended construct with at least one independently motivated reference. Discordant cases would be retained rather than averaged away. The analysis would record acquisition conditions, measurement error, and what each instrument or report omits. Success would mean preserving a justified interpretation under those limitations, not forcing every measurement to agree or declaring the most convenient measure the unquestioned ground truth.

The correspondence contribution should remain a source of interpretive questions, not a dataset of treatment outcomes. A source-review study can assess whether summaries preserve that role. A separate clinical or measurement study would be required to test patient outcomes, with professional oversight and appropriate ethical design. No automated-review score would substitute for that independent evidence.

Interpretive boundaries

The cited sources still have design-specific limits even where their abstracts and methods are accessible. This review is a bounded conceptual synthesis rather than a systematic review with an exhaustive search protocol. It cannot establish the effectiveness of the proposed integrated workflow. Its specific contribution is to articulate how plural outcomes changes the interpretation of subjective benefit, physiological signals, treatment goals. Technical sources support computational comparisons only; they do not validate biological mechanisms or patient benefit. Reviews and correspondence retain their publication roles and are not counted as new experiments. Proposed stress cases are illustrative, not observed failures attributed to the papers. Future work should therefore test the weakest inference described here, preserve contrary results, and disclose the cost of checking rather than treating additional automation as intrinsically beneficial.

Conclusion

The most valuable next result would be a discriminating test of plural outcomes, including a realistic way for the preferred approach to fail. Prespecify outcome roles and inspect discordance. Preserve patient priorities while withholding unsupported physiological or efficacy conclusions. The remaining problem can be stated as a testable contrast involving subjective benefit, physiological signals, treatment goals, with discordance treated as proof that one outcome is worthless as a deliberately challenging alternative. The cited studies provide useful but different forms of support. Their integration should increase traceability and expose missing links, leaving clinical, mechanistic, and operational claims at the strength justified by their own evidence.

References

1. Yu, Xiao, et al. "Subjective Sleep Gains in Empty Nose Syndrome: A Cautionary Note on the Need for Objective Validation." *International Forum of Allergy & Rhinology*, vol. 15, no. 7, 2025, pp. 761-62, <https://doi.org/10.1002/alr.23631>.
2. Ding, Siqi, et al. "Multifunctional Hydrogel Sensors with Dynamic Covalent Networks for Machine Learning-Assisted Parkinson's Disease Diagnosis and Encrypted Human-Computer Interaction." *Materials Today Bio*, vol. 35, 2025, <https://doi.org/10.1016/j.mtbio.2025.102524>.
3. Hernán, Miguel A., and James M. Robins. *Causal Inference: What If*. Chapman & Hall/CRC, 2020.
4. Rothman, Kenneth J., Sander Greenland, and Timothy L. Lash. *Modern Epidemiology*. 3rd ed., Lippincott Williams & Wilkins, 2008.
5. Benjamini, Yoav, and Yosef Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B*, vol. 57, no. 1, 1995, pp. 289-300.
6. Rajkomar, Alvin, Jeffrey Dean, and Isaac Kohane. "Machine Learning in Medicine." *New England Journal of Medicine*, vol. 380, 2019, pp. 1347-58.
7. Wiens, Jenna, et al. "Do No Harm: A Roadmap for Responsible Machine Learning for Health Care." *Nature Medicine*, vol. 25, 2019, pp. 1337-40.
8. Guo, Chuan, et al. "On Calibration of Modern Neural Networks." *Proceedings of the 34th International Conference on Machine Learning*, 2017, pp. 1321-30.
9. Ovadia, Yaniv, et al. "Can You Trust Your Model's Uncertainty? Evaluating Predictive Uncertainty under Dataset Shift." *Advances in Neural Information Processing Systems*, vol. 32, 2019.
10. Geifman, Yonatan, and Ran El-Yaniv. "Selective Classification for Deep Neural Networks." *Advances in Neural Information Processing Systems*, vol. 30, 2017.
11. Lundberg, Scott M., and Su-In Lee. "A Unified Approach to Interpreting Model Predictions." *Advances in Neural Information Processing Systems*, vol. 30, 2017.
12. World Health Organization. *Ethics and Governance of Artificial Intelligence for Health*. World Health Organization, 2021.