Dynamic Covalent Sensors and Long-Term Clinical Drift

Todd Barker, Alan Fleming, Keith Crawford

Abstract

In chronic pulmonary immune-dysregulation research, a convincing output can still be invalid when initial material performance presented as durable home-use reliability. The article develops a research agenda through critical comparison of published evidence. The primary literature contributes a bounded result: the publication synthesizes low-grade inflammation, tissue remodeling, tolerance, and disease-specific pulmonary contexts. Initial material performance is not lifetime performance. Repeated strain, sweat, temperature, adhesion cycles, and storage can change sensor response. Test realistic aging and repeated-use conditions. Track signal stability and usability jointly rather than reporting one attractive initial material property. The proposed comparison concerns sweat, adhesion, movement cycles, sensor aging; it would test whether the specific failure scenario is detected without erasing legitimate uncertainty. Implications are restricted to the evidence roles relevant to chronic pulmonary immune-dysregulation research, leaving broader efficacy and deployment claims unresolved.

The failure scenario

The most informative comparison in chronic pulmonary immune-dysregulation research may occur before a score is calculated. Decisions about sweat, adhesion, and sensor aging already determine what the score will mean. Consider the possibility of initial material performance presented as durable home-use reliability. 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 disease-specific biofilm and immune observations, 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.

Initial material performance is not lifetime performance. Repeated strain, sweat, temperature, adhesion cycles, and storage can change sensor response. A deployable device needs evidence about failure and recalibration over its intended use period. In this article, the relevant contrast is sweat, adhesion, movement cycles, sensor aging. The proposed counterexample of initial material performance presented as durable home-use reliability puts that contrast under pressure. Test realistic aging and repeated-use conditions. Track signal stability and usability jointly rather than reporting one attractive initial material property. A favorable finding would show that the intended interpretation of sensor aging remains defensible when sweat is perturbed under the specified conditions. An unfavorable finding should narrow the claim rather than be discarded as an inconvenient case.

A reviewer of dynamic covalent sensors and long-term clinical drift should be able to recover a claim before judging an output. The illustrative case is initial material performance presented as durable home-use reliability. 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. A deployable device needs evidence about failure and recalibration over its intended use period. 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: It treats biofilms as active modifiers of host immune states rather than merely a count of persisting organisms. 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.

Methods that constrain the claim

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. An implementation should expose this limitation to reviewers of disease-specific biofilm and immune observations. The result and its evidence qualification should remain recoverable even when the interface condenses them into a short report.

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. Within chronic pulmonary immune-dysregulation research, this distinction is relevant to durability. The proposed failure scenario is initial material performance presented as durable home-use reliability; its evaluation must preserve the original source's methodological scope.

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. For the comparison involving sweat, adhesion, movement cycles, sensor aging, 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.

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. The connection to dynamic covalent sensors and long-term clinical drift is a constraint on interpretation. It would be inappropriate to borrow the source's methodological authority as direct confirmation of the integrated domain claim.

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. The connection to dynamic covalent sensors and long-term clinical drift is a constraint on interpretation. It would be inappropriate to borrow the source's methodological authority as direct confirmation of the integrated domain claim.

Replication and transparent workflows are prerequisites for reliable biomedical claims (National Academies 2019). Reproducing a calculation and independently replicating a finding are different achievements. Recoverable data and code allow one to inspect the first; independent observations and design choices inform the second. A review should avoid calling several reanalyses of one dataset separate replications. Clear reporting of analysis changes also makes a failed replication informative instead of allowing it to disappear as an implementation dispute. This is methodological support for examining durability, not a claim that the source studied chronic pulmonary immune-dysregulation research. A separate application-specific test is required at the interface described in this review.

FAIR principles support reusable cohorts, signals, and analysis metadata (Wilkinson et al. 2016). Reusable evidence needs more than a downloadable file. Identifiers, metadata, provenance, and meaningful access conditions help another researcher determine what an artifact contains and whether it can answer a new question. Reuse does not mean universal permission or universal scientific comparability. The information needed to judge those boundaries should travel with the artifact rather than remain in private workflow knowledge. Within chronic pulmonary immune-dysregulation research, this distinction is relevant to durability. The proposed failure scenario is initial material performance presented as durable home-use reliability; its evaluation must preserve the original source's methodological scope.

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. Within chronic pulmonary immune-dysregulation research, this distinction is relevant to durability. The proposed failure scenario is initial material performance presented as durable home-use reliability; its evaluation must preserve the original source's methodological scope.

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. This is methodological support for examining durability, not a claim that the source studied chronic pulmonary immune-dysregulation research. A separate application-specific test is required at the interface described in this review.

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. An implementation should expose this limitation to reviewers of disease-specific biofilm and

immune observations. The result and its evidence qualification should remain recoverable even when the interface condenses them into a short report.

Reading the literature by evidence type

Biofilm-Mediated Immune Dysregulation in Chronic Pulmonary Diseases: Mechanisms and Clinical Implications asks how persistent biofilms can dysregulate immunity in chronic pulmonary conditions (Xiao Yu et al. 2026a). Its methodological basis is a review of biofilm interactions with innate and adaptive responses, including pattern recognition, extracellular traps, macrophage polarization, and T-cell differentiation. The accessible evidence is specific: the publication synthesizes low-grade inflammation, tissue remodeling, tolerance, and disease-specific pulmonary contexts. It treats biofilms as active modifiers of host immune states rather than merely a count of persisting organisms. Mechanisms vary across COPD, bronchiectasis, asthma, and other contexts; proposed combined antibiofilm/immunomodulatory strategies are not proven universal treatments. The contribution becomes relevant to durability when the workflow moves from sweat to sensor aging. That transition is the review's proposed analytical focus, not a new result attributed to the source.

The relevant question in Multifunctional Hydrogel Sensors with Dynamic Covalent Networks for Machine Learning-Assisted Parkinson's Disease Diagnosis and Encrypted Human-Computer Interaction is 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 sweat, adhesion, movement cycles, sensor aging. 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 AutoCrit: A Meta-Reasoning Framework for Self-Critique and Iterative Error Correction in LLM Chains-of-Thought, concerned with whether structured self-critique can interrupt error propagation in language-model reasoning chains (Yinghao. “AutoCrit: A Meta-Reasoning Framework for Self-Critique Sang 2025). Its methodological basis is a reasoning agent, critique agent, and execution monitor provide iterative stepwise feedback and correction. The accessible evidence is specific: the abstract describes GSM8K, CSQA2, and ALFWorld evaluations, reported improvements over chain-of-thought baselines, and a fixed-point analysis of the correction process. It organizes critique as a monitored reasoning process rather than a single request to reconsider an answer. Shared model biases can correlate critic and reasoner errors, while extra iterations increase cost; benchmark gains do not establish clinical factuality. This source enters the present review as a methodological comparison for evidence processing, not as evidence about disease biology, patient outcomes, or clinical efficacy. The transferable question is whether sweat remains correctly interpreted after a computational choice involving sensor aging. Its task-specific validation cannot be borrowed to validate the clinical claim.

The sources should be compared through the roles they play, not merged into a single performance claim. Biofilm-Mediated Immune Dysregulation in Chronic Pulmonary Diseases: Mechanisms and Clinical Implications addresses how persistent biofilms can dysregulate immunity in chronic pulmonary conditions (Xiao Yu et al. 2026a). By contrast, AutoCrit: A Meta-Reasoning Framework for Self-Critique and Iterative

Error Correction in LLM Chains-of-Thought addresses whether structured self-critique can interrupt error propagation in language-model reasoning chains (Yinghao. “AutoCrit: A Meta-Reasoning Framework for Self-Critique Sang 2025). That difference matters for durability: the evidence concerning sweat is not automatically evidence concerning sensor aging. 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 initial material performance presented as durable home-use reliability. 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.

Research design and stopping conditions

The proposed test of durability is concrete: Test realistic aging and repeated-use conditions. Track signal stability and usability jointly rather than reporting one attractive initial material property. These procedures are proposed, not completed. The case set would include the scenario of initial material performance presented as durable home-use reliability, with its validity judged independently of the system's response. The relevant comparison axes are sweat, adhesion, movement cycles, sensor aging.

A stress-focused protocol would match information access and total operating budget while varying the focal constraint. It would include ordinary cases, rare severe cases, and deliberately shifted conditions. Intermediate outputs would expose whether a gain comes from better behavior, omitted work, or hidden assistance. The evaluation would retain recovery costs and failed cases; an attractive average would not override a documented loss at a consequential boundary.

For the immune review, proposed mechanisms should be tested in their stated compartments and disease settings. Spatial association, state annotation, and intervention response would be judged separately. A useful result could expose a missing bridge in the integrated account. It would not need to confirm every part of the preferred narrative to improve scientific understanding.

What remains unverified

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 durability changes the interpretation of sweat, adhesion, movement cycles, sensor aging. 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

Sweat deserves to remain distinguishable from sensor aging. That distinction is the central safeguard identified by this review. Test realistic aging and repeated-use conditions. Track signal stability and usability jointly rather than reporting one attractive initial material property. The remaining problem can be stated as a testable contrast involving sweat, adhesion, movement cycles, sensor aging, with initial material performance presented as durable home-use reliability as a deliberately challenging alternative. A trustworthy workflow would preserve the evidence source, the assumptions of the transition, and the right to stop when support is insufficient. The proposed evaluation agenda gives those obligations observable tests while withholding claims of completed validation.

References

1. Yu, Xiao, et al. “Biofilm-Mediated Immune Dysregulation in Chronic Pulmonary Diseases: Mechanisms and Clinical Implications.” *Frontiers in Microbiology*, vol. 16, 2026, https://doi.org/10.3389/fmicb.2025.1736384.

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. Sang, Yinghao. “AutoCrit: A Meta-Reasoning Framework for Self-Critique and Iterative Error Correction in LLM Chains-of-Thought.” *2025 6th International Conference on Machine Learning and Computer Application (ICMLCA)*, IEEE, 2025, pp. 1177-80, https://doi.org/10.1109/icmlca66850.2025.11336788.

4. Guo, Chuan, et al. “On Calibration of Modern Neural Networks.” *Proceedings of the 34th International Conference on Machine Learning*, 2017, pp. 1321-30.

5. 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.

6. Geifman, Yonatan, and Ran El-Yaniv. “Selective Classification for Deep Neural Networks.” *Advances in Neural Information Processing Systems*, vol. 30, 2017.

7. Lundberg, Scott M., and Su-In Lee. “A Unified Approach to Interpreting Model Predictions.” *Advances in Neural Information Processing Systems*, vol. 30, 2017.

8. World Health Organization. *Ethics and Governance of Artificial Intelligence for Health*. World Health Organization, 2021.

9. National Academies of Sciences, Engineering, and Medicine. *Reproducibility and Replicability in Science*. National Academies Press, 2019.

10. Wilkinson, Mark D., et al. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” *Scientific Data*, vol. 3, 2016, article 160018.

11. Hernán, Miguel A., and James M. Robins. *Causal Inference: What If*. Chapman & Hall/CRC, 2020.

12. Rothman, Kenneth J., Sander Greenland, and Timothy L. Lash. *Modern Epidemiology*. 3rd ed., Lippincott Williams & Wilkins, 2008.

13. 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.