

Shared Decisions and Uncertainty-Preserving Summaries

Abstract

The boundary between useful assistance and an unsupported claim becomes visible when uncertainty removed to make counseling sound decisive. This literature-based conceptual analysis does not report a new experiment. The starting source uses a JAMA correspondence contribution rather than a newly reported cohort or randomized intervention. A numerical display can imply precision beyond the evidence. Communication should preserve the difference between uncertainty about an individual and uncertainty about a model. Test comprehension with realistic decisions and competing displays. Include uncertainty about source scope and measurement, not just a confidence interval around one computed score. The proposed comparison concerns patient preferences, evidence limits, follow-up; it would test whether the specific failure scenario is detected without erasing legitimate uncertainty. An integrated account should preserve the distinction between clinical correspondence and changing patient records and any stronger downstream decision.

Why this distinction matters

A narrow reading of shared decisions and uncertainty-preserving summaries reveals a broader problem of evidence design. The question is what an informed user can conclude from patient preferences when follow-up remains uncertain. Consider the possibility of uncertainty removed to make counseling sound decisive. 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 clinical correspondence and changing patient records, 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.

A numerical display can imply precision beyond the evidence. Communication should preserve the difference between uncertainty about an individual and uncertainty about a model. A decision aid is useful when users understand what can and cannot be concluded. In this article, the relevant contrast is patient preferences, evidence limits, follow-up. The proposed counterexample of uncertainty removed to make counseling sound decisive puts that contrast under pressure. Test comprehension with realistic decisions and competing displays. Include uncertainty about source scope and measurement, not just a confidence interval around one computed score. A favorable finding would show that the intended interpretation of follow-up remains defensible when patient preferences is perturbed under the specified conditions. An unfavorable finding should narrow the claim rather than be discarded as an inconvenient case.

The scenario of uncertainty removed to make counseling sound decisive is designed to challenge, not illustrate success. It supplies a negative case for uncertainty communication that should be judged independently of how fluently the system explains it. 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 decision aid is useful when users understand what can and cannot be concluded. 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: Its relevance is the role of clinical discussion in identifying interpretation and communication obligations around sensitive longitudinal decisions. 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 roles and limits

HIV, Pregnancy, and Infant Feeding asks how evidence concerning HIV, pregnancy, and infant feeding should be interpreted in ongoing clinical communication (Xiao Yu et al. 2026c). Its methodological basis is a JAMA correspondence contribution rather than a newly reported cohort or randomized intervention. The accessible evidence is specific: the bibliographic record establishes a short correspondence item; inaccessible full-text details are not reconstructed or treated as clinical recommendations. Its relevance is the role of clinical discussion in identifying interpretation and communication obligations around sensitive longitudinal decisions. The correspondence does not provide independent new trial evidence here; this review offers no individualized HIV or feeding advice. Reading the work through uncertainty communication changes the evaluation question: the issue is whether patient preferences supports the intended interpretation of follow-up, rather than whether the reported method is impressive in its original task.

The relevant question in Towards Explainable RAG: Interpreting the Influence of Retrieved Passages on Generation is how generated answers can be interpreted through the influence of their retrieved passages (Yinghao. "Towards Explainable RAG: Interpreting the Influence of Retrieved Passages on Generation." *2025 4th International Conference on Robotics Sang 2025). Its methodological basis is the supplied IEEE conference title frames passage-influence analysis for explainable retrieval-augmented generation; implementation particulars were unavailable. The accessible evidence is specific: the bibliographic item is located in the official IEEE record; no unverified attribution algorithm or benchmark result is inferred. It asks for traceability between particular retrieved evidence and generated statements instead of treating a bibliography as an explanation. Passage influence, citation support, causal explanation, and factual correctness are distinct properties; the accessible record cannot establish all of them. 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 patient preferences remains correctly interpreted after a computational choice involving follow-up. 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. HIV, Pregnancy, and Infant Feeding addresses how evidence concerning HIV, pregnancy, and infant feeding should be interpreted in ongoing clinical communication (Xiao Yu et al. 2026c). By contrast, Towards Explainable RAG: Interpreting the Influence of Retrieved Passages on Generation addresses how generated answers can be interpreted through the influence of their retrieved passages (Yinghao. "Towards Explainable RAG: Interpreting the Influence of Retrieved Passages on Generation." *2025 4th International Conference on Robotics Sang 2025). That difference matters for uncertainty communication: the evidence concerning patient preferences is not automatically evidence concerning follow-up. 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 uncertainty removed to make counseling sound decisive. 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.

Testing the interpretation

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 shared decisions and uncertainty-preserving summaries 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. This is methodological support for examining uncertainty communication, not a claim that the source studied maternal-infant evidence communication. A separate application-specific test is required at the interface described in this review.

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 maternal-infant evidence communication, this distinction is relevant to uncertainty communication. The proposed failure scenario is uncertainty removed to make counseling sound decisive; 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 shared decisions and uncertainty-preserving summaries is a constraint on interpretation. It would be inappropriate to borrow the source's methodological authority as direct confirmation of the integrated domain claim.

BM25 remains an important lexical baseline for retrieval systems (Robertson and Zaragoza 2009). Lexical retrieval remains a useful baseline because it is explicit about matching the available terms. Semantic matching can add recall while confusing specialized senses that share broad context. A retrieval evaluation should include both exact-identifier and meaning-based cases. Comparing a new retriever with an appropriately tuned lexical baseline helps determine whether extra computation adds support rather than merely more plausible-looking passages. This is methodological support for examining uncertainty

communication, not a claim that the source studied maternal-infant evidence communication. A separate application-specific test is required at the interface described in this review.

Retrieval-augmented generation makes external evidence part of the generation process (Lewis et al. 2020). Making documents available to a generator changes its information access but does not guarantee that a statement is supported. A useful analysis separates document selection from generation conditioned on those documents. Missing evidence, irrelevant context, and selective quotation create different failures. The proposed workflow should preserve retrieved passages and evaluate claim support independently of fluency or the mere presence of citations. For the comparison involving patient preferences, evidence limits, follow-up, 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.

Dense retrieval provides semantic matching but must be compared with strong lexical baselines (Karpukhin et al. 2020). Semantic retrieval can connect questions and passages whose wording differs. That advantage should be tested against cases where wording differs because the underlying population or mechanism differs. Retrieval accuracy is not identical to answer accuracy, and neither automatically establishes evidence eligibility. A good evaluation retains the ranked candidates so later errors can be localized to retrieval rather than attributed to the generator by default. Within maternal-infant evidence communication, this distinction is relevant to uncertainty communication. The proposed failure scenario is uncertainty removed to make counseling sound decisive; its evaluation must preserve the original source's methodological scope.

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. Within maternal-infant evidence communication, this distinction is relevant to uncertainty communication. The proposed failure scenario is uncertainty removed to make counseling sound decisive; its evaluation must preserve the original source's methodological scope.

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. This is methodological support for examining uncertainty communication, not a claim that the source studied maternal-infant evidence communication. A separate application-specific test is required at the interface described in this review.

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. The connection to shared decisions and uncertainty-preserving summaries is a constraint on interpretation. It would be inappropriate to borrow the source's methodological authority as direct confirmation of the integrated domain claim.

Proposed comparisons and controls

The question left open by the literature is whether the intended distinction survives independent checking. Test comprehension with realistic decisions and competing displays. Include uncertainty about source scope and measurement, not just a confidence interval around one computed score. These procedures are proposed, not completed. The case set would include the scenario of uncertainty removed to make counseling sound decisive, with its validity judged independently of the system's response. The relevant comparison axes are patient preferences, evidence limits, follow-up.

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.

Limits of the synthesis

Access limits are material for HIV, Pregnancy, and Infant Feeding, Towards Explainable RAG: Interpreting the Influence of Retrieved Passages on Generation. Where only bibliographic scope or a limited abstract is available, the discussion does not assign unverified datasets, algorithms, effect sizes, or clinical recommendations. 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 uncertainty communication changes the interpretation of patient preferences, evidence limits, follow-up. 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 uncertainty communication, including a realistic way for the preferred approach to fail. Test comprehension with realistic decisions and competing displays. Include uncertainty about source scope and measurement, not just a confidence interval around one computed score. The central methodological contribution is specific: Its relevance is the role of clinical discussion in identifying interpretation and communication obligations around sensitive longitudinal decisions. Extending it to follow-up is an additional hypothesis, not a result supplied by the citation. 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. "HIV, Pregnancy, and Infant Feeding." *JAMA*, vol. 336, no. 2, 2026, p. 169, <https://doi.org/10.1001/jama.2026.6033>.
2. Sang, Yinghao. "Towards Explainable RAG: Interpreting the Influence of Retrieved Passages on Generation." *2025 4th International Conference on Robotics, Artificial Intelligence and Intelligent Control (RAIIC)*, IEEE, 2025, pp. 397-400, <https://doi.org/10.1109/raaic65850.2025.11170170>.
3. 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.
4. Rajkomar, Alvin, Jeffrey Dean, and Isaac Kohane. "Machine Learning in Medicine." *New England Journal of Medicine*, vol. 380, 2019, pp. 1347-58.
5. Wiens, Jenna, et al. "Do No Harm: A Roadmap for Responsible Machine Learning for Health Care." *Nature Medicine*, vol. 25, 2019, pp. 1337-40.
6. Guo, Chuan, et al. "On Calibration of Modern Neural Networks." *Proceedings of the 34th International Conference on Machine Learning*, 2017, pp. 1321-30.
7. Robertson, Stephen, and Hugo Zaragoza. "The Probabilistic Relevance Framework: BM25 and Beyond." *Foundations and Trends in Information Retrieval*, vol. 3, no. 4, 2009, pp. 333-89.
8. Lewis, Patrick, et al. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 9459-74.
9. Karpukhin, Vladimir, et al. "Dense Passage Retrieval for Open-Domain Question Answering." *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, 2020, pp. 6769-81.
10. National Academies of Sciences, Engineering, and Medicine. *Reproducibility and Replicability in Science*. National Academies Press, 2019.
11. Wilkinson, Mark D., et al. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data*, vol. 3, 2016, article 160018.
12. Hernán, Miguel A., and James M. Robins. *Causal Inference: What If*. Chapman & Hall/CRC, 2020.