

Early Vision after Pituitary Surgery and Immune-State Narratives

Abstract

In immune-state and tumor evidence synthesis, a convincing output can still be invalid when early postoperative recovery treated as long-term function. The article develops a research agenda through critical comparison of published evidence. The primary literature contributes a bounded result: the review integrates cellular metabolism, nutrient/oxygen conditions, mitochondrial stress, and stromal metabolites as candidate mechanisms of persistence. Early and late outcomes can disagree because recovery is a trajectory. A short-term endpoint is useful for some decisions but should not be relabeled as permanent success or failure. Report the exact endpoint day and compare later trajectories. Treat early classification as a time-bounded prediction and quantify the uncertainty of longer-term extrapolation. The proposed comparison concerns day-seven vision, tumor size, immune concepts; it would test whether the specific failure scenario is detected without erasing legitimate uncertainty. The original source's limits remain material: A unifying set point is a hypothesis, not an already validated clinical measurement or an established therapeutic protocol.

The decision and its evidence contract

A narrow reading of early vision after pituitary surgery and immune-state narratives reveals a broader problem of evidence design. The question is what an informed user can conclude from day-seven vision when immune concepts remains uncertain. Consider the possibility of early postoperative recovery treated as long-term function. 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 tissue contexts, clinical endpoints, and mechanistic claims, 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.

Early and late outcomes can disagree because recovery is a trajectory. A short-term endpoint is useful for some decisions but should not be relabeled as permanent success or failure. Timing also affects missingness and who remains under observation. In this article, the relevant contrast is day-seven vision, tumor size, immune concepts. The proposed counterexample of early postoperative recovery treated as long-term function puts that contrast under pressure. Report the exact endpoint day and compare later trajectories. Treat early classification as a time-bounded prediction and quantify the uncertainty of longer-term extrapolation. A favorable finding would show that the intended interpretation of immune concepts remains defensible when day-seven vision is perturbed under the specified conditions. An unfavorable finding should narrow the claim rather than be discarded as an inconvenient case.

The transition at issue is specific to immune-state and tumor evidence synthesis. It concerns day-seven vision, tumor size, immune concepts, and the scenario of early postoperative recovery treated as long-term function shows why a locally correct result can still become a misleading integrated claim. 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.

Timing also affects missingness and who remains under observation. 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 set-point framing asks whether durable remission requires changing a coupled metabolic state rather than only lowering an inflammatory marker. 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.

Methodological comparisons

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 tissue contexts, clinical endpoints, and mechanistic claims. The result and its evidence qualification should remain recoverable even when the interface condenses them into a short report.

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. An implementation should expose this limitation to reviewers of tissue contexts, clinical endpoints, and mechanistic claims. The result and its evidence qualification should remain recoverable even when the interface condenses them into a short report.

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. The connection to early vision after pituitary surgery and immune-state narratives is a constraint on interpretation. It would be inappropriate to borrow the source's methodological authority as direct confirmation of the integrated domain 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. Within immune-state and tumor evidence synthesis, this distinction is relevant to endpoint timing. The proposed failure scenario is early postoperative recovery treated as long-term function; its evaluation must preserve the original source's methodological scope.

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. The connection to early vision after pituitary surgery and immune-state narratives is a constraint on interpretation. It would be inappropriate to borrow the source's methodological authority as direct confirmation of the integrated domain claim.

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. Within immune-state and tumor evidence synthesis, this distinction is relevant to endpoint timing. The proposed failure scenario is early postoperative recovery treated as long-term function; its evaluation must preserve the original source's methodological scope.

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. This is methodological support for examining endpoint timing, not a claim that the source studied immune-state and tumor evidence synthesis. A separate application-specific test is required at the interface described in this review.

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 early vision after pituitary surgery and immune-state narratives 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. Within immune-state and tumor evidence synthesis, this distinction is relevant to endpoint timing. The proposed failure scenario is early postoperative recovery treated as long-term function; 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 endpoint timing, not a claim that the source studied immune-state and tumor evidence synthesis. A separate application-specific test is required at the interface described in this review.

What the focal studies establish

A useful boundary case is supplied by *Resetting Immunometabolic Set Points in Autoimmune Disease*, concerned with why autoimmune activity can persist or recur when conventional inflammation suppression does not durably change underlying cellular and tissue states (Xiao Yu et al. 2026e). Its methodological basis is a conceptual review proposing immunometabolic set points across immune and tissue compartments. The accessible evidence is specific: the review integrates cellular metabolism, nutrient/oxygen conditions, mitochondrial stress, and stromal metabolites as candidate mechanisms of persistence. The set-point framing asks whether durable remission requires changing a coupled metabolic state rather than only lowering an inflammatory marker. A unifying set point is a hypothesis, not an already validated clinical measurement or an established therapeutic protocol. The contribution becomes relevant to endpoint timing when the workflow moves from day-seven vision to immune concepts. That transition is the review's proposed analytical focus, not a new result attributed to the source.

Machine Learning-Based Prediction of Early Visual Field Nonrecovery after Endoscopic Surgery for Nonfunctioning Pituitary Adenomas with Clinically Accessible Biomarkers: A Dual-Center Study asks whether accessible preoperative biomarkers can predict early visual-field nonrecovery after surgery for nonfunctioning pituitary adenomas (Sunjiang Liu et al. 2026). Its methodological basis is a retrospective dual-center design with LASSO selection, multivariable logistic modeling, and a nomogram; training comprised 268 patients and validation 100. The accessible evidence is specific: the study includes 368 patients, defines early improvement at postoperative day seven, and reports age, vertical tumor diameter, and fasting glucose as predictors; reported C-indices are approximately 0.747 and 0.753. It develops an accessible early-risk tool with geographically separate validation instead of relying solely on model fit in development data. Day-seven nonrecovery is not permanent visual loss; retrospective selection, limited validation events, and unmeasured care differences constrain interpretation. The contribution becomes relevant to endpoint timing when the workflow moves from day-seven vision to immune concepts. That transition is the review's proposed analytical focus, not a new result attributed to the source.

(Qian Guo et al. 2026) examines how spatial glial-immune interactions organize permissive and suppressive microdomains within brain tumors in *Immune Niches in Brain Tumors: Glial-Immune Cell Interactions and Spatial Microdomains*. Its methodological basis is a review of astrocytes, microglia, oligodendrocyte-lineage cells, T cells, and myeloid cells across spatial tumor niches. The accessible evidence is specific: the abstract integrates cytokine, metabolic, hypoxic, and immunoediting mechanisms associated with immune suppression and exhaustion. It shifts attention from tumor-wide averages to interactions whose meaning depends on location and cell context. The review does not establish efficacy of a new intervention; co-location, state annotation, and functional interaction require different evidence. For endpoint timing, the important connection is the distinction between day-seven vision and immune concepts. The source informs that distinction, but a stronger conclusion would require evidence at the point where one is used to stand for the other.

The relevant question in *Splenic Extramedullary Granulopoiesis and Neutrophil States in Cancer Immunity: Systemic Immunosuppression and Therapeutic Resistance* is how splenic extramedullary granulopoiesis and neutrophil heterogeneity contribute to systemic cancer immunosuppression and resistance (Yidan Zhang et al. 2026). Its methodological basis is a review organized around stress-associated splenic

blood-cell production and reprogrammed neutrophil states. The accessible evidence is specific: the official Cancer Bioscience recent-articles listing confirms the review's scope and emphasizes sustained stress and systemic reprogramming; no new patient cohort is inferred. The framing connects extra-tumoral immune production with local cancer resistance questions. The accessible summary cannot establish specific quantitative contributions, universal cell-state functions, or clinical effects of proposed targeting strategies. The contribution becomes relevant to endpoint timing when the workflow moves from day-seven vision to immune concepts. That transition is the review's proposed analytical focus, not a new result attributed to the source.

(Bibo Gao et al. 2026) examines which retrospective clinical features predict extended hospital stay after endoscopic endonasal surgery for nonfunctioning pituitary adenomas in *The Prediction of Extended Hospital Length of Stay in Patients after Endoscopic Endonasal Transsphenoidal Surgery for the Resection of Non-Functioning Pituitary Adenomas: A Dual-Center Retrospective Analysis*. Its methodological basis is a dual-center retrospective prediction analysis with a modeled extended-stay outcome and separate validation. The accessible evidence is specific: the accessible record describes a nomogram and a 100-patient validation sample; specific predictor effects and thresholds are not assumed. It treats hospitalization burden as an explicitly predicted outcome that can inform care-process evaluation. Length of stay includes organizational and discharge factors; it is not an interchangeable measure of tumor biology, surgical quality, or long-term recovery. The contribution becomes relevant to endpoint timing when the workflow moves from day-seven vision to immune concepts. That transition is the review's proposed analytical focus, not a new result attributed to the source.

The sources should be compared through the roles they play, not merged into a single performance claim. *Resetting Immunometabolic Set Points in Autoimmune Disease* addresses why autoimmune activity can persist or recur when conventional inflammation suppression does not durably change underlying cellular and tissue states (Xiao Yu et al. 2026e). By contrast, *The Prediction of Extended Hospital Length of Stay in Patients after Endoscopic Endonasal Transsphenoidal Surgery for the Resection of Non-Functioning Pituitary Adenomas: A Dual-Center Retrospective Analysis* addresses which retrospective clinical features predict extended hospital stay after endoscopic endonasal surgery for nonfunctioning pituitary adenomas (Bibo Gao et al. 2026). That difference matters for endpoint timing: the evidence concerning day-seven vision is not automatically evidence concerning immune concepts. 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 early postoperative recovery treated as long-term function. 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.

A discriminating evaluation agenda

The proposed test of endpoint timing is concrete: Report the exact endpoint day and compare later trajectories. Treat early classification as a time-bounded prediction and quantify the uncertainty of longer-term extrapolation. These procedures are proposed, not completed. The case set would include the scenario of early postoperative recovery treated as long-term function, with its validity judged independently of the system's response. The relevant comparison axes are day-seven vision, tumor size, immune concepts.

A timing-focused protocol would preserve issue, acquisition, exposure, and outcome clocks separately. Reviewers would identify the information available at each decision and compare conclusions before and

after later information is introduced. The key contrast is whether a claim changes because the underlying state changes or because the analysis has borrowed a future observation. Retrospective reconstruction and prospective use must receive different labels.

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.

Scope and unresolved questions

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 endpoint timing changes the interpretation of day-seven vision, tumor size, immune concepts. 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 counterexample of early postoperative recovery treated as long-term function shows why local success and integrated validity must be evaluated separately. Report the exact endpoint day and compare later trajectories. Treat early classification as a time-bounded prediction and quantify the uncertainty of longer-term extrapolation. The central methodological contribution is specific: The set-point framing asks whether durable remission requires changing a coupled metabolic state rather than only lowering an inflammatory marker. Extending it to immune concepts is an additional hypothesis, not a result supplied by the citation. A stronger conclusion will require evidence at the exact interface described in the failure scenario. Repeating a source, improving fluency, or increasing execution success cannot substitute for that test.

References

1. Yu, Xiao, et al. "Resetting Immunometabolic Set Points in Autoimmune Disease." *Journal of Translational Autoimmunity*, vol. 13, 2026, <https://doi.org/10.1016/j.jtauto.2026.100391>.
2. Liu, Sunjiang, et al. "Machine Learning-Based Prediction of Early Visual Field Nonrecovery after Endoscopic Surgery for Nonfunctioning Pituitary Adenomas with Clinically Accessible Biomarkers: A Dual-Center Study." *International Journal of Clinical Practice*, vol. 2026, no. 1, 2026, <https://doi.org/10.1155/ijcp/2504255>.
3. Guo, Qian, et al. "Immune Niches in Brain Tumors: Glial-Immune Cell Interactions and Spatial Microdomains." *Acta Neuropathologica Communications*, vol. 14, 2026, <https://doi.org/10.1186/s40478-026-02373-0>.
4. Zhang, Yidan, et al. "Splenic Extramedullary Granulopoiesis and Neutrophil States in Cancer Immunity: Systemic Immunosuppression and Therapeutic Resistance." *Cancer Bioscience*, 2026, <https://doi.org/10.1016/j.cbs.2026.100009>.
5. Gao, Bibo, et al. "The Prediction of Extended Hospital Length of Stay in Patients after Endoscopic Endonasal Transsphenoidal Surgery for the Resection of Non-Functioning Pituitary Adenomas: A Dual-Center Retrospective Analysis." *Cancers*, vol. 18, no. 10, 2026, <https://doi.org/10.3390/cancers18101582>.
6. 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.
7. Rajkomar, Alvin, Jeffrey Dean, and Isaac Kohane. "Machine Learning in Medicine." *New England Journal of Medicine*, vol. 380, 2019, pp. 1347-58.
8. Wiens, Jenna, et al. "Do No Harm: A Roadmap for Responsible Machine Learning for Health Care." *Nature Medicine*, vol. 25, 2019, pp. 1337-40.
9. Guo, Chuan, et al. "On Calibration of Modern Neural Networks." *Proceedings of the 34th International Conference on Machine Learning*, 2017, pp. 1321-30.
10. 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.
11. Geifman, Yonatan, and Ran El-Yaniv. "Selective Classification for Deep Neural Networks." *Advances in Neural Information Processing Systems*, vol. 30, 2017.
12. Lundberg, Scott M., and Su-In Lee. "A Unified Approach to Interpreting Model Predictions." *Advances in Neural Information Processing Systems*, vol. 30, 2017.
13. World Health Organization. *Ethics and Governance of Artificial Intelligence for Health*. World Health Organization, 2021.
14. National Academies of Sciences, Engineering, and Medicine. *Reproducibility and Replicability in Science*. National Academies Press, 2019.
15. Wilkinson, Mark D., et al. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data*, vol. 3, 2016, article 160018.