

A Prior Authorization SLA Must Follow the Whole Case

A Practical Measurement Method for Healthcare Outsourcing

Phil Hatch | Akholi

October 23, 2026

Healthcare outsourcing managers should measure the complete prior-authorization case through submission, acknowledgment, decision, and downstream action. Agent task counts can conceal an unfinished handoff while a patient continues to wait.

Define Completion at the Case Level

The healthcare outsourcing manager should define a prior-authorization service around the case state the buyer needs to achieve. Extracting information, completing a form, or drafting a request is an activity. Completion requires evidence that the intended submission reached the correct destination and that the contracted follow-up and recording obligations were performed.

Stanford's October 2026 HealthAdminBench report distinguishes subtask scores from complete administrative tasks. It reports 82.8% subtask performance but 26.7% whole-task completion for one tested computer-use configuration. These are synthetic benchmark results, not clinical deployment rates. They show why a task count cannot substitute for execution evidence.¹

The service-level agreement should preserve clinical decision ownership and patient escalation. An administrative agent should not infer authority to make a clinical determination

¹ Nigam Shah and Jonathan Chen, "Measuring What Medical AI Can Actually Do," Stanford University, October 19, 2026, "PhysicianBench and MedAgentBench" and "HealthAdminBench," <https://hai.stanford.edu/news/measuring-what-medical-ai-can-actually-do>.

or close an unresolved case. Define the permitted assistance, the qualified decision-maker, and the route for urgent or ambiguous cases.

Create a Shared Case State Dictionary

The buyer and provider should agree the states that matter: received, information incomplete, ready to submit, submitted, acknowledged, further information requested, decision received, appealed where in scope, and recorded or otherwise completed. Each transition needs an observable event and an owner. The exact dictionary should match the contracted workflow.

A case can be waiting on the buyer, provider, payer, clinician, or another party. Record that responsibility without losing total elapsed time. A pause in one contractual clock should not erase the patient's wait or a statutory deadline. Managers need both responsibility attribution and the full age of the case.

The study's reports of completion flags inconsistent with actual work provide a transferable reporting concern. They arise in financial-services outsourcing, not healthcare. The appropriate application is a reconciliation test between reported case states and independent execution evidence, rather than a claim about hospital failure prevalence.²

Bind Each State to Evidence

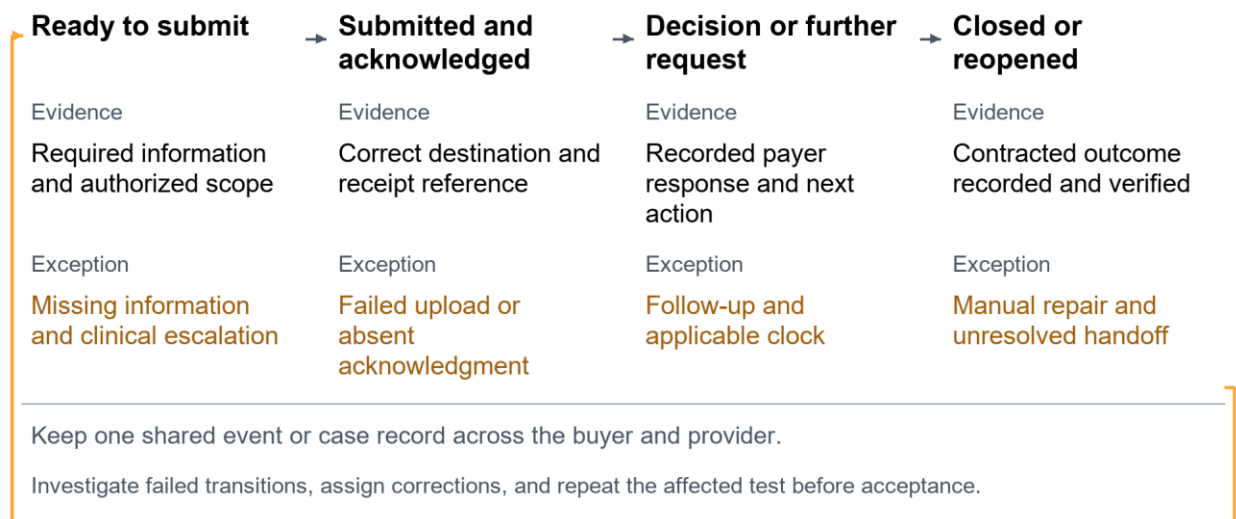
Figure 1 proposes a case-state record. Submission should require a destination, timestamp, case identifier, and available acknowledgment. A missing file or failed upload should leave the case incomplete. The service should detect duplicate attempts and preserve their relationship to the original case.

For a hypothetical portal submission, an agent may report success after completing the visible form while failing to attach a required document. The manager should inspect the receiving system's state or another reliable acknowledgment. A screenshot of the draft page cannot prove submission or completeness.

Define sampling across completed, delayed, corrected, rejected, and reopened cases. The assurance method should reflect exposure and volume; a convenient sample of successful cases will miss the handoffs most likely to matter. Retain the evidence needed for investigation without unnecessarily duplicating patient information.

² Phil Hatch, *Agentic AI Outsourcing at Tier 1 Financial Institutions: Full Study Report*, version 1.8.2 (Akholi, forthcoming manuscript dated November 1, 2026), Part 3, AI Technology, sec. 4.1, "Accuracy of the Tools' Own Success and Failure Reports," indicator C172, <https://doi.org/10.2139/ssrn.7164559>.

Figure 1. Follow the case rather than the agent activity



Author's proposed method; adapt the sequence and decision rights to the service.

Source: Phil Hatch, proposed framework developed for this paper.

Separate Contract Clocks from Legal Deadlines

The US CMS-0057-F fact sheet states decision timeframes for specified impacted payers, generally beginning in 2026: 72 hours for expedited requests and seven calendar days for standard requests, excluding the stated QHP issuers from that timeframe provision. The rule does not establish a universal deadline for all payers, drugs, or jurisdictions.³

Managers should have counsel and operational specialists identify the applicable clock, trigger, exclusions, escalation, and evidence for each service. The contract should state whether it measures the provider's preparation, submission, follow-up, or end-to-end coordination. It should not imply that the outsourcing provider controls a payer's decision.

For UK, EU, Swiss, and other services, replace US payer terminology and timing with the local authorization or funding process. The common measurement method remains useful, but local clinical, privacy, safety, and patient-rights obligations determine the permissible workflow.

³ Centers for Medicare & Medicaid Services, *CMS Interoperability and Prior Authorization Final Rule CMS-0057-F* (Centers for Medicare & Medicaid Services, January 17, 2024), pp. 1–3, operational provisions and decision timeframes, <https://www.cms.gov/files/document/fact-sheet-cms-interoperability-and-prior-authorization-final-rule-cms-0057-f.pdf>.

Report Patient Relevant Delay and Provider Performance

The dashboard should distinguish total case age, time in each state, missing evidence, provider-controlled delay, and cases needing clinical escalation. Present the oldest and most consequential exceptions, not only average processing time. An average can improve while a small unresolved queue continues to create material patient or service risk.

A case marked waiting on the payer should still have an assigned follow-up owner and the next required action. Otherwise, responsibility coding can become a way to remove difficult cases from the provider's measured queue. The manager should inspect aged waiting cases, verify the original submission and acknowledgment, and check that the correct escalation occurred. The provider should retain the total timeline even when the contract excludes some time from a provider-controlled measure.

Manual completion belongs in the record. The study distinguishes alteration of reporting from correction of outputs and processes. Applied here, a provider should preserve the original status, the human intervention, and the final result. The manager should not allow a manually completed case to remain described as unattended agent completion.⁴

Use incentives carefully. Paying for submissions can reward duplicate or incomplete requests; paying only for approvals can encourage inappropriate selection or obscure payer decisions outside the provider's control. The commercial measure should reward the contracted contribution and preserve required clinical judgment.

Accept the SLA Through a Reconciled Pilot

Before broader use, reconcile a bounded pilot from receipt through the agreed terminal state. Test missing documents, incorrect destinations, failed acknowledgments, urgent cases, and reopening. Assign corrective owners and repeat failed transitions before treating the service measure as reliable.

The SLA should tell the manager whether the case progressed as required and why it remains unresolved. A shared state dictionary, evidence at each transition, and separate legal and contractual clocks make that possible. Activity counts can remain diagnostic measures, but they should not define completion.

⁴ Hatch, *Agentic AI Outsourcing*, Part 5, "Five Manual Alteration Methods" and "Disclosure of Manual Alteration," table 5.8.

Evidence and Source Note

The study is a nonrepresentative cross-section of 850 vendor delivery-team respondents on 150 engagements serving Tier 1 financial institutions, collected during January–June 2026. Neither banks nor vendors corroborated the accounts. It establishes reported experience, not causation, legal breach, or market prevalence. Its application here is a mechanism to investigate, not an estimate of conditions in other industries. The author-supplied version 1.8.2 is dated November 1, 2026 and is cited as a forthcoming manuscript consulted October 23, 2026.⁵

This is an analytical decision paper based on the supplied study and selected primary public sources checked through October 23, 2026. The original contribution is a case-state SLA dictionary and reconciliation method separating activity, execution, responsibility, and elapsed time. The proposed method has not been independently validated. Legal requirements depend on the entity, activity, jurisdiction, and effective provisions; supervisory observations, research, and professional commentary have distinct authority.

Bibliography

Centers for Medicare & Medicaid Services. *CMS Interoperability and Prior Authorization Final Rule CMS-0057-F*.

Centers for Medicare & Medicaid Services, January 17, 2024. <https://www.cms.gov/files/document/fact-sheet-cms-interoperability-and-prior-authorization-final-rule-cms-0057-f.pdf>.

Hatch, Phil. *Agentic AI Outsourcing at Tier 1 Financial Institutions: Full Study Report*. Version 1.8.2. Akholi.

Forthcoming manuscript dated November 1, 2026. Author-supplied manuscript consulted October 23, 2026. <https://doi.org/10.2139/ssrn.7164559>.

Shah, Nigam, and Jonathan Chen. “Measuring What Medical AI Can Actually Do.” Stanford University. October 19, 2026. <https://hai.stanford.edu/news/measuring-what-medical-ai-can-actually-do>.

⁵ Hatch, *Agentic AI Outsourcing*, Introduction, “Executive Summary” and “Evidentiary Basis and Its Limits.”