

A Context Engine for Medicare Advantage Risk Adjustment and Documentation Integrity

How Enterprise AI Could Strengthen Evidence, Policy, and Review Decisions Before They Reach an Auditor

A practical premise: Risk adjustment is most defensible when an organization can connect every proposed condition to contemporaneous clinical evidence, applicable MA policy context, and a qualified reviewer decision before any downstream use — the core discipline of documentation integrity.



Cole Talton

Director of AI Healthcare Markets

A white paper by Cole Talton, Director of AI Healthcare Markets, Cyberhill Partners. For Medicare Advantage plan, risk-bearing provider, clinical documentation integrity, risk adjustment, and compliance leaders. July 2026.

EXECUTIVE SUMMARY

What it is: A context engine built on a knowledge graph that connects source evidence, policy context, and reviewer decisions into one inspectable view for Medicare Advantage risk adjustment and documentation integrity.

What it does: Assembles source-linked evidence packets before a qualified reviewer decides, so every submitted diagnosis can be traced back to its clinical support, governing policy version, and documented rationale.

Why now: CMS expanded RADV audits from ~60 contracts to all ~550 eligible contracts annually. HHS OIG flagged \$462M in potential overpayments from a single diagnosis category in 2026. The question is no longer whether your organization will be audited, but whether you can reconstruct the reasoning behind your decisions when it is.

WHO THIS IS FOR

VP of Risk Adjustment, Chief Compliance Officer, Chief Medical Officer, VP of Clinical Documentation Integrity, Medicare Advantage plan and risk-bearing provider leadership.

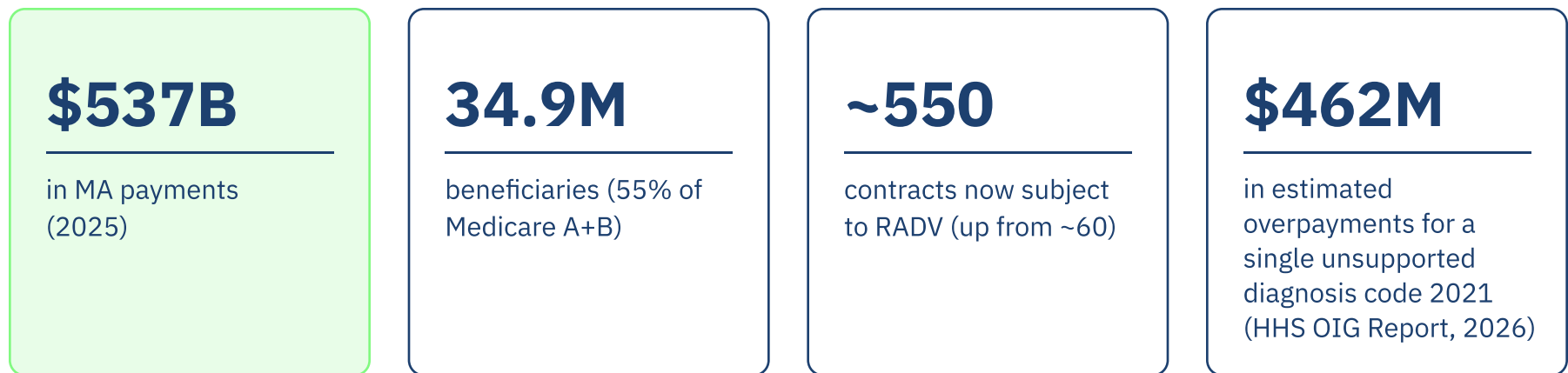
If time is short: the appendix contains a one-page, ten-question self-assessment. It takes about two minutes and is designed to be read on its own.

1. The current state: MA context is scattered across systems and time

Medicare Advantage risk adjustment depends on diagnoses supported in the medical record and submitted under payment-year requirements. MedPAC reported that MA enrolled about 34.9 million beneficiaries in 2025 — 55 percent of Medicare beneficiaries with both Part A and Part B coverage — and that Medicare paid MA plans an estimated \$537 billion, not including payments for drug coverage. Documentation integrity is therefore a program-integrity, compliance, clinical-governance, and financial-stewardship concern.

Scrutiny is expanding. CMS announced in May 2025 that newly initiated RADV audits would cover all eligible MA contracts for each payment year — approximately 550 contracts, up from roughly 60 annually. CMS states that unsupported diagnoses may result in overpayment findings. In 2026, HHS OIG found that none of 97 sampled acute-stroke diagnoses were supported by related records and estimated \$462 million in potential net overpayments for 2021 — a finding that reflects a targeted, high-risk coding pattern and should not be generalized to every diagnosis category or organization.

The Scale of MA Risk Adjustment



The problem is architectural, not a shortage of data or effort. Evidence lives across clinical notes, encounters, problem lists, claims, and policy rules, usually assembled late as disconnected fragments. History can raise a review question but does not prove current support: chronic conditions persist, while support must meet the applicable encounter and documentation requirements. The final decision is typically stored apart from the evidence that justified it.

Policy context compounds this. CMS finalized an updated risk model in the CY 2024 Rate Announcement — the “2024 CMS-HCC model,” commonly called v28 — and phased it in through 2026. The same historical diagnosis can carry a different review implication depending on the payment year in force.

The extrapolation posture is also unsettled: the Northern District of Texas vacated the 2023 RADV final rule in *Humana Inc. v. Becerra* on procedural grounds in September 2025, CMS appealed, and as of publication the Fifth Circuit appeal remains in briefing with no decision issued. RADV audits have continued regardless. Confirm the current legal position with qualified counsel rather than relying on this summary.

The rest of this paper describes an alternative approach (Section 2), what it costs to find out whether it fits your organization (Section 3), and what to do next (Section 4).

2. The context engine: connecting evidence before the decision

The alternative to the fragmented, retrospective review just described is a context engine: a governed layer connecting the evidence behind a review decision into one inspectable view. Architecturally, it is a semantic layer built on a knowledge graph — representing the person, encounter, source note, condition candidate, risk-model association, policy, and reviewer decision as connected entities rather than isolated rows. An ontology specifies the permitted entity types and relationships, so a candidate ties to a note, author, service date, evidence cue, policy version, and reviewer action. The engine does not decide a condition is valid; it makes the relationship inspectable.

The unit of work is an evidence packet: a source-linked, time-aware view showing provenance, chronology, evidence cues, governing policy context, missing-data flags, and the reviewer’s written rationale. Each packet shows what is present, absent, and uncertain, so a reviewer can assess the question without assuming the answer.

The workflow is deliberately constrained, shown in Figure 1 below. Automation retrieves, organizes, and prioritizes within defined boundaries; it does not create an open-ended diagnosis or silently convert narrative text into a submission decision. MEAT-oriented cues are treated as review prompts, not proof.

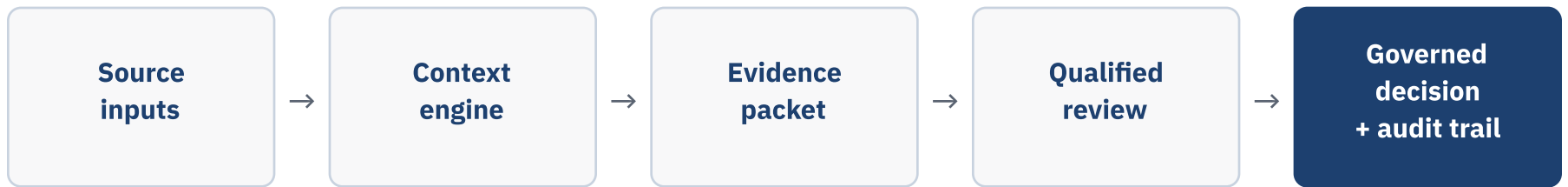


Figure 1. The constrained workflow: automation assembles and routes; a qualified reviewer decides.

A worked example

Traditional candidate: prior history lists chronic systolic heart failure; today’s note states “Continue medications”; a recapture queue labels an HCC opportunity. The excerpt doesn’t establish whether the condition is current or what policy applies.

In a governed, context-centered review: the reviewer sees the encounter, source note, author, service date, longitudinal history, medication context, and applicable payment-year policy — and a note stating, “Chronic systolic heart failure, reports feeling about the same. Daily weights reviewed, appears stable. Continue current regimen; recheck in three months.” The note is unremarkable, not exemplary — but it is current, attributable, and dated. Same condition, same code. The difference is whether MA context is organized before the decision, and whether rationale, source, policy version, and reviewer action are preserved for a later auditor.

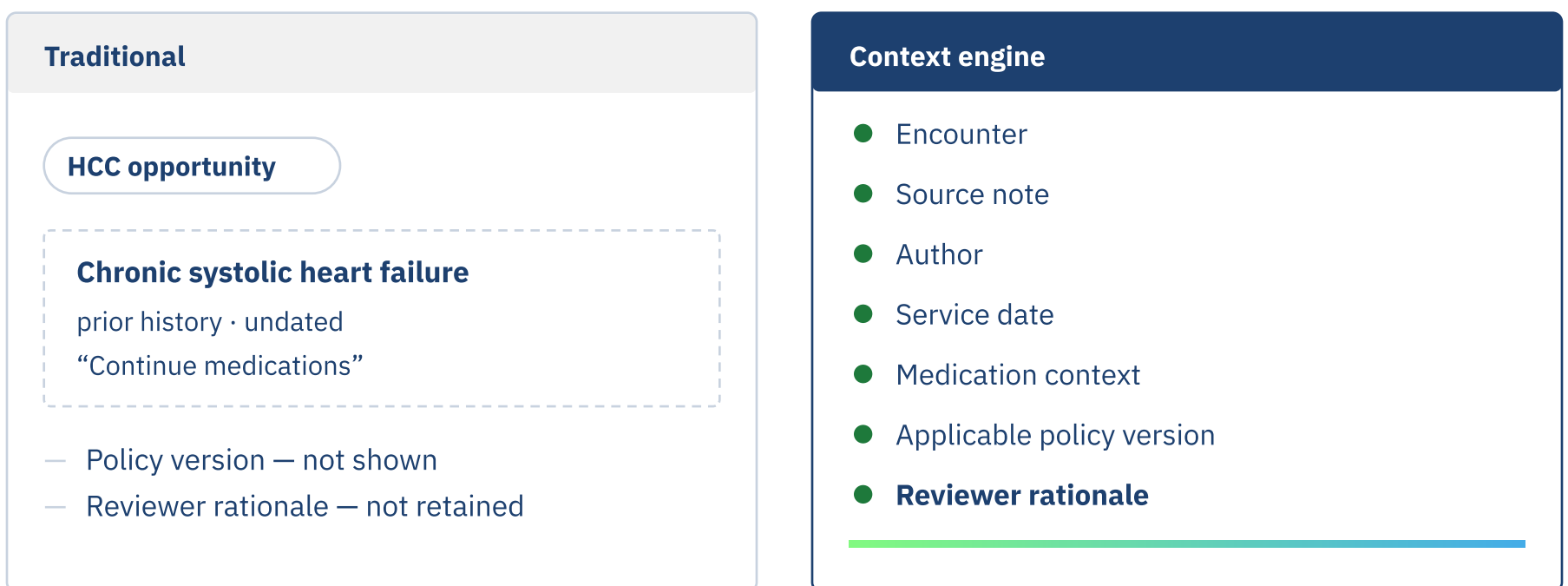


Figure 2. The same condition as a queue label and as an evidence packet.

Dimension	Traditional retrospective review	Governed context engine review
Unit of work	Code candidate or chart list.	Condition question with source-linked evidence and policy context.
Failure mode	History or detached problem list drives work.	Evidence, policy, and missing-data gaps are visible.
RADV defensibility	Logic often reconstructed after the fact.	Provenance, policy, context, and rationale retained.
Role of AI	Opaque code-like prediction.	Constrained context assembly, retrieval, and prioritization; a human decides.

Table 1. Traditional retrospective review versus a governed context engine review.

3. Implementation, measurement, and honest limits

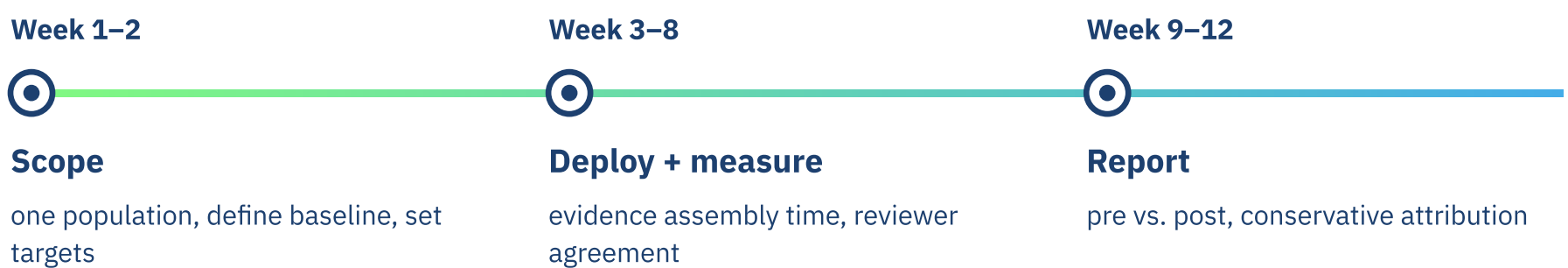
A governed context engine workflow carries no universal RAF-uplift claim. Value should be quantified in a bounded pilot, not asserted in advance.

What a 90-day pilot looks like

Begin with one population or documentation question. Define the baseline before deployment and set targets in advance — for example, evidence-assembly time per case, or reviewer agreement.

Report pre- versus post-deployment change with conservative attribution. Potential, accepted, submitted, and realized outcomes are four different numbers; conflating them is how pilots produce figures that don't survive audit.

Success is a documented outcome, not a favorable one. A pilot showing no improvement, honestly measured, is still a useful result.



Governance must define ownership for policy changes, reviewer qualifications, and release approval. Each decision should preserve source, timestamp, missing-data disclosures, policy version, reviewer, rationale, and downstream use. Measurement is a safety control: track evidence sufficiency, false positives, suppression patterns, and reviewer disagreement — including where the engine performs poorly.

Where this can fall short

No context engine can create support data absent from the record; missing sources must stay visible, not be assumed away. Payment-year mappings require expert review and change control. A weakly prioritized queue merely moves chart chase to a new screen, and a workflow tuned for apparent RAF opportunity over defensible context creates compliance risk. Guidance, models, and payer policies change, requiring active maintenance.

Addressing common objections

“We already have a risk-adjustment vendor and a chart-review process.” Existing tools may remain essential. The question is whether they consistently connect source evidence, chronology, policy, and reviewer rationale.

“How would this integrate with our EHR and our current risk-adjustment vendor?” A context engine sits across existing systems rather than replacing them. Integration scope drives pilot timeline and depends on which systems are involved — an engineering assessment, not a claim this paper can make, and the pilot's first deliverable rather than a precondition for it.

“What does this cost, and how long before it does anything?” Both are scoped in a short discovery conversation, not guessed at here: a bounded, time-boxed assessment with defined scope, data requirements, success measures, and an explicit go/no-go decision point — not an open-ended implementation. That scoping conversation is the actual first step, not a delay before one.

“Is this just another RAF-lift engine?” It should not be. Suppression and deferral must be as visible as approval, and financial models must separate potential, submitted, and realized outcomes. The value is context and defensibility before financial exposure.

“Can AI determine whether a diagnosis is supported?” No. AI can retrieve evidence, highlight cues, summarize chronology, connect policy context, and flag gaps. Qualified clinical, coding, or compliance reviewers make the consequential decision.

“How does this protect us in a RADV audit?” No technology guarantees an audit outcome or replaces compliant documentation. The contribution is readiness: source-linked evidence, visible gaps, versioned policy, and a recorded decision trail showing how the MA context supported the reviewer’s decision.

4. Conclusion and next steps

Risk adjustment is judged not only by whether a diagnosis was submitted, but by whether the organization can show the record supported it. A context engine connects source records, condition candidates, policy, and audit history — and the same foundation can extend to quality and payment-integrity questions.

What Cerebro does today, stated plainly: as currently implemented, it assembles source-linked evidence packets for condition review, applies CMS-published risk-adjustment reference data, routes each candidate to a qualified human reviewer, requires a written rationale for every decision including suppression, and records that decision in an append-only audit trail. These capabilities have been demonstrated in a controlled environment and are not a representation about performance in any production deployment. Cerebro does not determine whether a diagnosis is supported.

What a deployment could achieve is separate and organization-specific: integration, ontology design, policy configuration, reviewer training, and security controls are all required. The approach is designed to reduce avoidable context loss and improve audit traceability — it does not remove audit exposure or guarantee a financial result.

Cyberhill Partners builds enterprise AI on ontologies, semantic layers, and governed context. The right starting point is narrow: identify the single population or documentation question where context loss is most costly today, and scope it as a bounded pilot with a defined baseline and pre-set targets — the structure set out in “What a 90-day pilot looks like” in Section 3. Contact Cyberhill Partners to scope that pilot.

Appendix: Documentation Context Readiness

Ten yes/no questions about whether the reasoning behind your risk-adjustment decisions can be reconstructed later. Nothing is submitted to anyone. The useful signal is in the questions you cannot answer with a clear yes. This page is designed to be used on its own.

	Question	Yes	No
1	For any diagnosis you submitted, can you produce the source note, its author, and the service date in under five minutes?	☐	☐
2	Can you tell whether support for that diagnosis was current for the review period, rather than carried forward from prior history?	☐	☐
3	For a decision made three payment years ago, can you determine which risk model and policy configuration were in force on that service date?	☐	☐
4	When a reviewer declines a candidate, is that decision recorded with its stated reason — or does the candidate simply disappear from the queue?	☐	☐
5	Does your review process distinguish evidence that is genuinely absent from evidence that was never expected?	☐	☐
6	Is a written reviewer rationale required before a decision is saved, rather than optional?	☐	☐
7	Would any alteration of a past review decision or its audit record be detectable and attributable after the fact?	☐	☐
8	For any submitted diagnosis, can you name the individual who approved it and confirm their role and qualification?	☐	☐
9	If your current risk-adjustment platform ended tomorrow, could you still answer questions 1 through 8 for the prior three payment years?	☐	☐
10	Do you track reviewer disagreement, suppression rates, and false positives — not only candidates captured?	☐	☐

Reading your answers

8–10

YES

Your context is already well governed. The remaining question is whether that holds under a records request for a payment year that predates your current systems.

5–7

YES

Common. Evidence is usually reachable but assembled manually, and the reasoning behind past decisions is reconstructed rather than retained.

0–4

YES

Context is being lost between systems. The gap is architectural rather than a matter of coding effort, and more review capacity will not close it.

Where this leads. Every question above describes context that exists somewhere in your organization but is not connected: the source, the chronology, the policy version in force, the reviewer’s reasoning, and what was declined. A context engine connects those into one governed view before a decision is made and preserves it afterward. If three or more answers above were uncomfortable, that is a useful place to begin — start with the single population or documentation question where the loss is most costly today.

This self-assessment is not a compliance, legal, coding, or audit assessment and does not evaluate compliance with any CMS or other requirement. The notice below applies.

Important notice and disclaimer

This white paper is provided by Cyberhill Partners for general informational and educational purposes only. It describes a conceptual approach and design goals; it does not describe measured performance, warrant any product capability, or constitute legal, regulatory, clinical, coding, compliance, actuarial, or financial advice, and it does not create any advisory or client relationship. Descriptions of laws, regulations, litigation, audit methods, and agency guidance summarize publicly reported developments believed accurate as of the date of publication, are not legal analysis, and are subject to change; the legal status of RADV extrapolation in particular remains unsettled. Statistics and findings are attributed to their cited public sources and have not been independently audited or endorsed by Cyberhill Partners. Any comparisons refer to workflow characteristics generally and are not assertions about any specific product or vendor. This approach is intended to support accurate, well-documented risk adjustment and must not be used to justify submitting diagnoses that the medical record does not support. Any implementation involving protected health information requires appropriate legal agreements, minimum-necessary access, and security controls; the governance practices described are design principles, not representations about any specific deployment. Organizations should consult qualified legal counsel and their own clinical, coding, and compliance advisors before acting on any information in this paper. Nothing herein guarantees any audit outcome, compliance result, or financial return.

References

1. Medicare Payment Advisory Commission. “The Medicare Advantage Program: Status Report.” Chapter 12 in Report to the Congress: Medicare Payment Policy. March 2026.
https://www.medpac.gov/wp-content/uploads/2026/03/Mar26_Ch12_MedPAC_Report_To_Congress_SEC.pdf (accessed July 28, 2026).
2. Centers for Medicare & Medicaid Services. “CMS Rolls Out Aggressive Strategy to Enhance and Accelerate Medicare Advantage Audits.” Press release, May 21, 2025.
<https://www.cms.gov/newsroom/press-releases/cms-rolls-out-aggressive-strategy-enhance-accelerate-medicare-advantage-audits> (accessed July 28, 2026).
3. Centers for Medicare & Medicaid Services. “Medicare Advantage Risk Adjustment Data Validation Program.” Page last modified March 4, 2026.
<https://www.cms.gov/data-research/monitoring-programs/medicare-risk-adjustment-data-validation-program> (accessed July 28, 2026).
4. HHS Office of Inspector General. “CMS Potentially Overpaid Medicare Advantage Organizations \$462 Million Based on Certain Unsupported Acute Stroke Diagnosis Codes.” Report A-02-23-01020, issued May 28, 2026; posted June 1, 2026.
<https://oig.hhs.gov/reports/all/2026/cms-potentially-overpaid-medicare-advantage-organizations-462-million-based-on-certain-unsupported-acute-stroke-diagnosis-codes/> (accessed July 28, 2026).
5. Centers for Medicare & Medicaid Services. “Announcement of Calendar Year (CY) 2024 Medicare Advantage Capitation Rates and Part C and Part D Payment Policies” (CY 2024 Rate Announcement). March 31, 2023.
<https://www.cms.gov/files/document/2024-announcement-pdf.pdf> (accessed July 28, 2026). Summarized in CMS, “Fact Sheet: 2024 Medicare Advantage and Part D Rate Announcement,” March 31, 2023,
<https://www.cms.gov/newsroom/fact-sheets/fact-sheet-2024-medicare-advantage-and-part-d-rate-announcement>. Phase-in completion for CY 2026 described in CMS, “2026 Medicare Advantage and Part D Advance Notice Fact Sheet,”
<https://www.cms.gov/newsroom/fact-sheets/2026-medicare-advantage-part-d-advance-notice-fact-sheet> (all accessed July 28, 2026).
6. Humana Inc. and Humana Benefit Plan of Texas, Inc. v. Becerra, No. 4:23-cv-00909-O (N.D. Tex. Sept. 25, 2025) (ECF No. 76) (O’Connor, C.J.), vacating and remanding the 2023 RADV final rule, 88 Fed. Reg. 6,643 (Feb. 1, 2023); notice of appeal filed Nov. 21, 2025, docketed Dec. 30, 2025 as No. 25-11293 (5th Cir.); appellants’ opening brief filed Mar. 21, 2026. No Fifth Circuit decision had issued as of the publication date.