

Report No. 009

Koru Research Institute

Six Fronts, One War: How AI Turned Healthcare Billing Into an Arms Race

Marcus J. Frazier

Founder & Managing Partner, Koru Health

A Koru Capital Management Company

New York, NY

mfrazier@korucm.com | korucm.com

October 2026

A Note From the Author

There is a dangerous point in intelligence work when a room becomes too certain of itself. The briefing is polished, the numbers agree, and the logic closes neatly around the conclusion. That is often when the smallest inconsistency becomes the most important fact in the room. National security taught me that truth rarely announces itself by volume. More often, it reveals itself through resistance: the detail that will not conform, the assumption no one can defend, the absence that changes the meaning of everything around it. I came to live by a simple maxim: when the picture is too perfect, study the frame. The gravest failures are not always born from bad information. Sometimes they come from good information arranged around the wrong question. That lesson stayed with me. So when I encounter a number now, I do not ask whether it is accurate. I ask what reality had to be ignored for that accuracy to feel complete.

I had that same feeling again reading a Blue Cross Blue Shield Association analysis about hospital billing. On its own, it read like an accounting dispute: hospitals coding patients as more complex, an insurer group crying foul, a hospital association pushing back that patients are simply sicker than they used to be. I have read a hundred disputes shaped exactly like that one. What stopped me was not the dispute; it was what showed up once I asked the only question my training ever taught me to ask first, which is what else is happening in this system right now that nobody has connected to it.

The answer took months to assemble, and it is the subject of this report. Hospitals are using artificial intelligence to find every billable complication a human coder might have missed. Insurers are using artificial intelligence to find every deniable technicality a human adjuster might have missed. Some of those same insurers are using artificial intelligence to find every extra diagnosis they can attach to their own government payments, a fact currently under federal audit. The federal government has now built its own version of the same tool and pointed it at Traditional Medicare. And running beneath every one of these fronts sits a legal doctrine, decades old, that most participants in this fight do not know exists, and that determines how much any of it actually costs the side doing the harm.

None of the parties in this report are the villain the public conversation has already cast them as. Each one is responding, more or less rationally, to an incentive nobody designed to work together with anyone else's. That is what makes this a war rather than a scandal. A scandal has a guilty party. A war has six fronts, a great deal of collateral damage, and, so far, no one keeping score across all of them at once.

This report keeps that score.

A scandal has a guilty party. A war has six fronts, and until now, no one has been keeping score across all of them at once.

Marcus J. Frazier

Abstract

This report documents a healthcare billing conflict that most public coverage treats as a single dispute and that this investigation shows is a six front war, fought with the same class of artificial intelligence, deployed by different parties, pointed in different directions, against a population that can generally see only one front at a time. On the first front, hospitals deploy coding intensity artificial intelligence to identify billable clinical complexity a human coder might miss, a pattern documented directly in the Blue Cross Blue Shield Association's own published analyses and disputed by the American Hospital Association. On a second and third front, insurers, UnitedHealth through its nH Predict algorithm and Cigna through its Px Dx system, deploy denial artificial intelligence to identify deniable technicalities a human claims adjuster might miss, now the subject of active federal litigation in the District of Minnesota and the Eastern District of California respectively.

On a fourth front, some of those same insurers stand accused by the United States Department of Health and Human Services Office of Inspector General of using artificial intelligence assisted chart review to add diagnosis codes to their own Medicare Advantage risk adjustment submissions, inflating what the federal government pays them rather than what they pay a provider. On a fifth front, the federal government has adopted a structurally similar technology and payment model of its own: the Centers for Medicare and Medicaid Services' Wasteful and Inappropriate Service Reduction model, extending prior authorization review into Traditional Medicare for the first time. A sixth front is structural rather than algorithmic: a single corporate parent can now employ the physician documenting a clinical encounter, insure the patient the encounter concerns, and separately face

federal scrutiny for how its own subsidiaries code government payments, a conflict of interest with no dedicated oversight body.

Underneath all six fronts sits a legal architecture, established by the United States Supreme Court in *Pilot Life Insurance Co. v. Dedeaux* and *Aetna Health Inc. v. Davila*, that limits the financial consequences of a wrongful algorithmic denial for the majority of insured Americans to the value of the benefit itself. This report also introduces an original methodological contribution: an application of the selective labels problem, a documented concern in computer science literature on algorithmic decision making, to healthcare claims adjudication for the first time, showing how a denial model retrained on its own historical outcomes may be structurally unable to learn from its own errors.

This report extends Koru Research Institute's prior work on healthcare's infrastructure dependencies in Report No. 008, *Healthcare Runs on Compute Now: How AI, Data Centers, Cloud Infrastructure, and Energy Are Reshaping Healthcare Delivery*, and its analysis of artificial intelligence's geopolitical footprint in the Global Risk Brief, *The Bottleneck Beneath the AI Boom: How a Global Energy Average Hides Where Power, Water, and Chips Run Out*.

Executive Findings / Intelligence Assessment

1 RETAINED DENIAL MODELS ARE LEARNING FROM A BROKEN SAMPLE.

Fewer than one percent of denied claims are ever appealed, a pattern documented independently in two separate Kaiser Family Foundation analyses: one covering Medicare Advantage, cited directly in the Estate of Gene B. Lokken litigation, and one covering Affordable Care Act marketplace plans, cited directly in the Kisting-Leung v. Cigna Corp. litigation. A denial algorithm retrained on its own historical outcomes is therefore learning overwhelmingly from claims nobody challenged, not from claims independently proven correct. Computer science literature on algorithmic decision making describes this exact structural risk as the selective labels problem. Applying that concept to healthcare claims adjudication for the first time is this report's original contribution, developed in full in Part Four, and it implies these systems may grow more confident in their own errors with each retraining cycle rather than less.

2 THE SAME COMPANIES SIT ON BOTH SIDES OF THEIR OWN INCENTIVE.

UnitedHealth Group and Humana are simultaneously under federal audit for using artificial intelligence assisted chart review to add diagnosis codes that increase what the government pays them, and separately named in litigation alleging their use of artificial intelligence to deny claims they owe to patients and providers. One underlying capability, aimed in whichever direction the money happens to flow.

3 A DECADES OLD SUPREME COURT RULING NOW GOVERNS THE STAKES OF A TECHNOLOGY THAT DID NOT EXIST WHEN IT WAS DECIDED.

Aetna Health Inc. v. Davila and Pilot Life Insurance Co. v. Dedeaux hold that most employer sponsored health plans are shielded from state law bad faith and wrongful death claims, limiting a wrongfully denied beneficiary's recovery to the value of the benefit that should have been paid. Justice Ginsburg's own concurrence in Davila called the result a regulatory vacuum. That vacuum, built for a world of paper claims and human adjusters, now absorbs the consequences of algorithmic errors the Court had no occasion to imagine.

4 THE GOVERNMENT ADOPTED THE INCENTIVE STRUCTURE IT IS SIMULTANEOUSLY INVESTIGATING.

The Centers for Medicare and Medicaid Services pays the private technology companies operating its WISer model based on their success reducing services and lowering program spending, the same savings contingent structure already drawing scrutiny on the insurer denial side. Several of WISer's selected vendors bring direct experience managing prior authorization for the Medicare Advantage plans currently under federal investigation for the inverse behavior.

5 ONE COMPANY CAN OCCUPY THREE SIDES OF THE SAME TRANSACTION.

Optum, UnitedHealth Group's physician services division, is the largest employer or affiliate of physicians in the United States.

The same corporate parent can employ the physician documenting a clinical chart, insure the patient that chart describes, and separately face federal scrutiny for how its Medicare Advantage subsidiaries code their own government payments. No dedicated regulatory body currently examines this concentration as a single conflict of interest.

6 THE HUMAN COST IS DOCUMENTED, NOT HYPOTHETICAL, AND MUST BE STATED PRECISELY.

Gene Lokken, 91, and his family paid approximately 150,000 dollars out of pocket for continued rehabilitation after UnitedHealthcare ended coverage on the recommendation attributed to its nH Predict algorithm. He died five months later. His estate's complaint survived UnitedHealth's motion to dismiss on breach of contract and breach of the implied covenant of good faith and fair dealing, and the case remains in active discovery. This is an allegation a federal court has permitted to proceed, not a judicial finding of fact, and this report treats it with that precision throughout.

BOTTOM LINE

No single actor in this war is the villain the public conversation has assumed. The hospital association, the insurers, the government payer, and the courts are each responding in a manner rational to the incentive immediately in front of them, and none of those incentives were designed to reconcile with any of the others. The result is a system in which artificial intelligence is winning on every front simultaneously, not because any one party is uniquely malicious, but because no party, and no regulator, has yet built the mechanism to check whether any side's optimization target still points at the patient in the middle of it.

Introduction: The War Nobody Declared

Every war this report documents is being fought by someone who did not choose to enlist. The hospital finance director did not set out to fight an insurer's algorithm; she is responding to a coding intensity vendor's five to one return on investment promise because her margins already assume it. The claims adjuster did not choose to stop reading patient files; the batch review software made that choice for her years ago, and reversing it would mean staffing a department that no longer exists. The 91 year old patient recovering from a fall did not choose to become a data point in an appeal statistic. Nobody who actually touches this system chose the war they are now fighting inside it.

That is the central obstacle this report exists to name. A reader who follows healthcare news has almost certainly encountered one skirmish from this war already, most likely the Blue Cross Blue Shield Association's dispute with hospitals over coding intensity, since it broke into national coverage in September 2026. Almost no coverage of that dispute has connected it to the litigation against UnitedHealth's nH Predict algorithm, or to Cigna's Px Dx system, or to the federal audit findings against UnitedHealth and Humana's own Medicare Advantage coding, or to the government's own new prior authorization model, or to the decades old Supreme

Court doctrine that caps what any of this costs the party causing the harm. Each front has its own beat reporter, its own trade press, its own specialized legal analysis. None of them, as far as this institute has found, has been asked to look at all six fronts on the same map at once.

This report is that map, written using the same discipline this institute has applied to healthcare's compute dependencies in Report No. 008, Healthcare Runs on Compute Now: How AI, Data Centers, Cloud Infrastructure, and Energy Are Reshaping Healthcare Delivery, and to the geopolitics of artificial intelligence in the Global Risk Brief, The Bottleneck Beneath the AI Boom: How a Global Energy Average Hides Where Power, Water, and Chips Run Out. The discipline is the same in every case: establish what the primary documents actually say, separate an allegation from a finding, and follow the incentive past the headline to the mechanism underneath it. What that discipline surfaces here is not a story about which side of the billing fight deserves more sympathy; it is a story about six independently verified fronts, one recurring villain hiding in a Supreme Court decision from 2004, and a population of patients and honest providers caught in the crossfire of a war that none of the combatants agreed to declare against one another, and that all of them are currently losing ground in without realizing it.

Part One: The Measurement War

Before any single front in this war can be understood, one pattern needs to be named, because it recurs at every level of this story and it is the reason the war persists instead of resolving. Nobody involved, not the hospitals, not the insurers, not the federal government, not even the companies describing their own scale, can agree on what the underlying measurement actually is, and that ambiguity is not a minor technical footnote. It is the load bearing structural feature that keeps every front of this war unresolved.

Two BCBSA Reports, Not One

Start with the dispute that opened this investigation. The Blue Cross Blue Shield Association, working with its analytics partner Blue Health Intelligence, published two distinct documents in 2026 that press coverage has frequently conflated into a single number. The first, an issue brief titled Rising Coding Intensity and Its Impact on Health Care Affordability, published in March 2026 and authored by name, David Wennberg, Francois Fressin, Ryan Everist, Elizabeth Horng, Elise Nelson, and Sara Clemente, examined commercial inpatient claims from April 2022 through March 2025 across a subset of Blue Cross Blue Shield companies representing approximately 62 million members. It found a 9 percent per member increase in inpatient costs from 2023 to 2024, with roughly 20 percent of that increase attributable to rising coding intensity rather than a corresponding increase in documented treatment. Its most damning finding is a direct quote from an anonymous senior clinical executive at one BCBS plan, describing an actual chart audit at a hospital system with outlier coding growth: less than 20 percent of postpartum anemia diagnoses in that audit met established clinical criteria.

The second document, formally titled Hospital Coding Intensity Analysis: Major Bowel Procedures and authored by Chris Birkmeyer, David Wennberg, Keith Kamons, and Luke Chalker, David Wennberg's name appearing on both the March and September publications, analyzed claims data from the first quarter of 2023

through the fourth quarter of 2025 across BCBSA's member companies, which the association states collectively cover roughly one in three Americans. It is the source of the 942 million dollar figure that dominated subsequent press coverage, with roughly 70 percent, more than 650 million dollars, tied to more than 55,000 cases in which a secondary diagnosis moved a claim into a higher paying severity tier, and the overall share of inpatient stays billed as complex rising from approximately 37 percent to approximately 40 percent. Chalker, on the record, framed the finding bluntly: if patients are truly sicker, we would expect to see more treatment.

Some September 24 coverage extrapolated the Blue plan specific figure to 2.3 billion dollars nationally, a press extrapolation BCBSA itself did not make as its headline claim. A report built on this material has an obligation to state plainly which number belongs to which document, because collapsing them is precisely the kind of imprecision this report exists to criticize when other parties commit it. It is also worth noting, in the same spirit, that BCBSA's own analysis has drawn a methodological critique of its own. At least one physician reviewer, writing independently, has pointed out that BCBSA analyzed claims data without chart review, does not show any individually fabricated diagnosis, and does not establish that AI specifically, rather than more complete documentation generally, caused the increase. That is the identical class of objection the American Hospital Association has raised, and the identical measurement problem this report opened with in Part One.

	Actor	Mechanism	What the AI Optimizes For
Front 1	Hospitals	Coding intensity AI	Find more billable complexity
Front 2	UnitedHealth	Denial AI (nH Predict)	Find more deniable technicalities
Front 3	Cigna	Denial AI (Px Dx)	Find more deniable technicalities
Front 4	UnitedHealth / Humana	Upcoding AI	Find more billable diagnoses (from CMS)
Front 5	CMS (government)	WISeR prior authorization	Find more reducible services
Front 6	Optum / UnitedHealth	Vertical integration	Sit on all three sides of the transaction

Exhibit 1. The six independently documented fronts of the healthcare AI billing war, the actor deploying each mechanism, and what each system is optimized to find.

The Same Fight, One Level Up: How Big Is UnitedHealth's Own Conflict of Interest

This measurement dispute is not confined to clinical coding. It recurs at the level of corporate self description.

UnitedHealth Group's Optum division is widely reported, including in the company's own investor facing materials, to employ or affiliate with approximately 90,000 physicians, a figure regularly cited as evidence of Optum's position as the largest employer of physicians in the United States. In May 2024 congressional

testimony, then chief executive Andrew Witty offered a markedly different number to lawmakers: just under 10,000 physicians directly employed, with the remaining 80,000 characterized as merely affiliated. Same company, same year, two different numbers depending on whether the audience is an investor conference or a congressional hearing, the identical measurement ambiguity problem found in the BCBSA dispute, recurring one level higher, at the scale of describing the very conflict of interest examined in Finding 5 of this report.

Why This Section Matters to Every Front That Follows

Every front examined in Part Two rests on a contested measurement. Whether a coded diagnosis reflects genuine clinical severity or aggressive documentation. Whether a denied claim reflects genuine medical unnecessary or an algorithm's technicality. Whether a Medicare Advantage risk score reflects a genuinely sicker patient population or an upcoded one. In each case, the party with the financial incentive to resolve the ambiguity in its own favor is also, not coincidentally, the party controlling the measurement, a pattern that recurs six times because it is the single mechanism generating all six fronts. Understanding it here makes every subsequent finding in this report legible as a variation on one theme rather than six unrelated stories.

Part Two: The Six Fronts

FRONT 1: HOSPITAL CODING INTENSITY AI

The mechanism documented in Part One's BCBSA analysis has a clear technological driver. Ambient listening tools convert clinician patient conversations directly into structured notes, and coding artificial intelligence then scans that expanded documentation, alongside lab results and problem lists, to propose additional billable diagnoses or higher severity coding combinations. BCBSA's own March 2026 brief documents the clearest example in maternity care: hospitals with the largest increases in postpartum anemia coding, defined as growth of 30 percent or more in the diagnosis rate, saw that diagnosis rise from 4.0 percent of maternity admissions in the second quarter of 2022 to 12.3 percent by the first quarter of 2025. Transfusion rates, the standard treatment for the condition, rose only from 0.8 percent to 1.2 percent over the same period, an eight percentage point rise in diagnosis matched by less than half a percentage point rise in treatment.

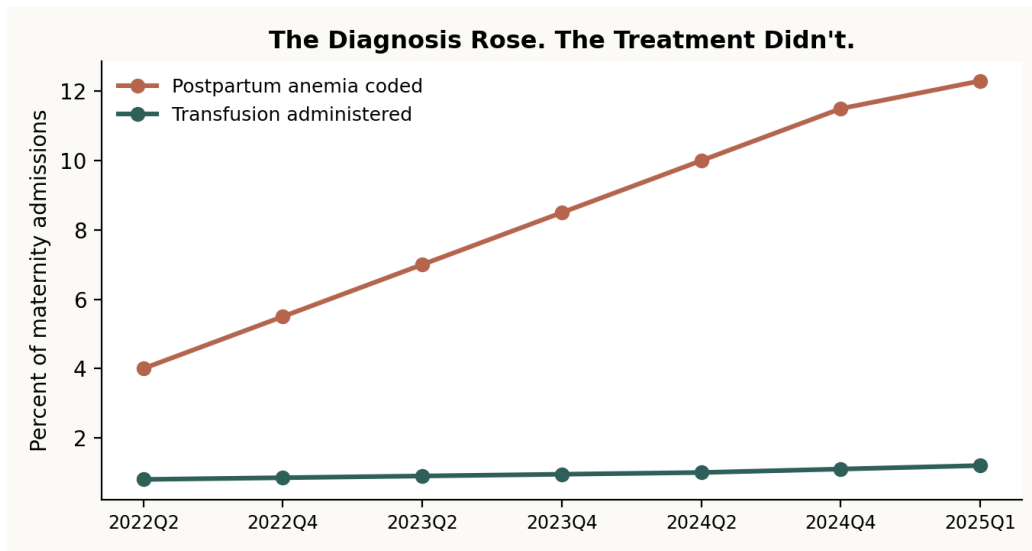


Exhibit 2. Postpartum anemia coding rose more than eight percentage points at high growth hospitals between 2022 and 2025. The treatment for the condition barely moved.

Coding vendors serving this market are frequently structured as contingency arrangements. Public case studies from vendors including SmarterDx describe returns on investment as high as five to one with no workflow changes required, and separate industry reporting describes contingency fees in the range of a fifth to a sixth of the additional revenue surfaced. The American Hospital Association's counter argument is not without merit and deserves to be stated fairly: an aging, increasingly complex patient population, combined with ambient documentation tools that genuinely capture clinical detail a rushed clinician previously failed to dictate, would produce a similar signature in the data. BCBSA's own analysis concedes it examined claims data rather than clinical charts, the one data source that would definitively settle which explanation is correct. Neither side has published a comprehensive, independent, patient level chart validation of the disputed cases. That validation gap, not a verdict on which side is right, is this report's finding on Front 1.

FRONT 2: INSURER DENIAL AI, UNITEDHEALTH'S NH PREDICT

Gene B. Lokken, 91, fractured his leg and ankle in a fall in May 2022. UnitedHealthcare covered nineteen days of post acute rehabilitation before the recommendation attributed to its nH Predict algorithm ended that coverage. His family paid approximately 150,000 dollars out of pocket for his continued care until his death in July 2023. His estate is the lead plaintiff in Estate of Gene B. Lokken, et al. v. UnitedHealth Group, Inc., et al., filed November 14, 2023, in the United States District Court for the District of Minnesota, Case No. 0:23-cv-03514. The complaint alleges nH Predict carries a 90 percent error rate on appealed determinations. UnitedHealth's own filings deny using nH Predict in the manner alleged. A February 2025 ruling dismissed some claims but allowed the breach of contract and breach of the implied covenant of good faith and fair dealing claims to proceed, because UnitedHealth's own plan documents promised that clinical staff and physicians, not an algorithm, would make coverage determinations. Discovery was ongoing as of March 2026.

The statistic that makes this front's information asymmetry concrete rather than anecdotal comes from the complaint's own citation to Kaiser Family Foundation research on Medicare Advantage: only about 0.2 percent

of denied members ever file an appeal. Separate reporting on this population has found that when naviHealth processed denials were appealed, a very high share, cited in multiple accounts as approximately 90 to 97 percent, were overturned. An almost certainly wrong decision, almost never checked, is the precise mechanism this report calls information asymmetry throughout, and it is documented here to a federal complaint and a cited KFF study, not inferred from a news summary.

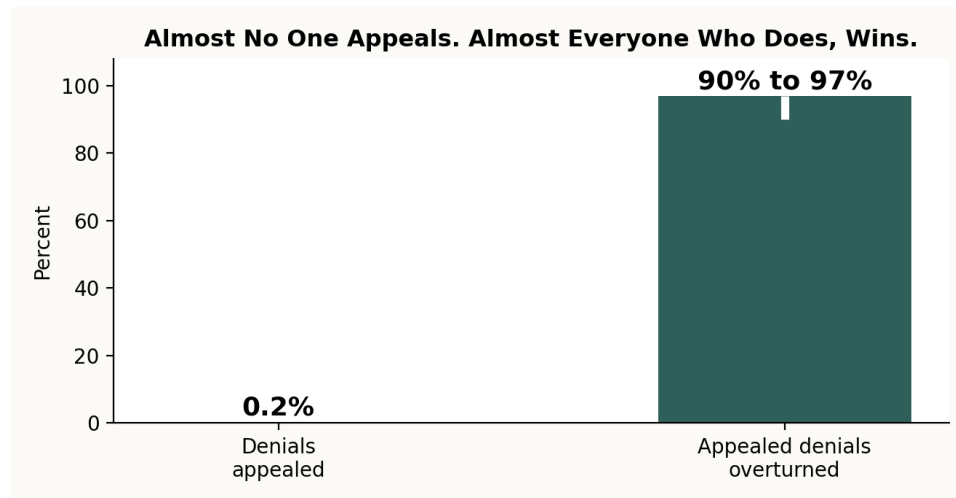


Exhibit 3. Only about 2 in 1,000 denied Medicare Advantage members ever appeal. When they do, the overwhelming majority win.

FRONT 3: INSURER DENIAL AI, CIGNA'S PDX

The same information asymmetry recurs independently in a second insurer and a second court. Kisting-Leung, et al. v. Cigna Corp., et al., filed July 24, 2023, in the United States District Court for the Eastern District of California, Case No. 2:23-cv-01477, alleges that Cigna's PxDx algorithm identifies discrepancies between a submitted diagnosis and a preset list of acceptable tests or procedures, then allows a physician to sign off on the resulting denials in batches of hundreds or thousands without opening an individual patient file. The complaint, citing the same ProPublica investigation that first surfaced the practice, states that Cigna's doctors denied more than 300,000 payment requests over a two month period in 2022, spending an average of 1.2 seconds reviewing each one. Named plaintiff Suzanne Kisting-Leung was denied coverage for a transvaginal ultrasound, ordered after a suspected ovarian cancer risk, twice, for bills of 198 dollars and 525 dollars respectively; Cigna's own coverage policy for the same procedure describes it as medically necessary for exactly the indication her doctor cited.

A March 31, 2025 ruling narrowed the case in a way that independently confirms this report's Front on the ERISA liability structure discussed in Part Three: three named plaintiffs lost standing because Cigna successfully argued the algorithm had not, in fact, been used on their specific claims, while the core breach of fiduciary duty theory, that Cigna's plan documents promised medical director review the algorithm displaced, survived. The complaint separately cites a Kaiser Family Foundation study of Affordable Care Act marketplace plans finding a comparably low appeal rate to the Medicare Advantage figure in Front 2, an independent confirmation, in a second insurance market entirely, that near zero appeal rates are not unique to one company or one line of business.

FRONT 4: INSURER UPCODING AI, THE SAME WEAPON, THE OPPOSITE DIRECTION

This is the front that makes the preceding two look incomplete on their own.

A United States Department of Health and Human Services Office of Inspector General audit, report number A-07-24-01214, issued September 10, 2026, examined Medicare Advantage Contract H5253, UnitedHealthcare of Wisconsin, for payment years 2020 and 2021. OIG did not sample diagnoses at random from the plan's full population; it drew a stratified random sample of 250 enrollee years specifically from eleven groups of diagnosis codes its methodology had independently identified as high risk for unsupported documentation, then asked UnitedHealthcare to produce the medical record behind each one. Of those 250, 183, 73.2 percent, lacked chart support.

The sample itself contained 722,280 dollars in overpayments; OIG then applied that error rate to the plan's broader population of similar high risk codes for the same two payment years and estimated at least 46,913,121 dollars in improper payments, against 16.57 billion dollars CMS paid this single subsidiary over the audit window. UnitedHealth's rebuttal, preserved in the audit and repeated to the press, does not dispute that this was the sample's finding. It argues instead that a sample deliberately targeted at the categories most likely to contain errors cannot fairly represent the plan's overall accuracy, and that 250 enrollee years is too small a foundation for an extrapolation of this size, the identical measurement dispute structure documented in Part One, a targeted sample's honest results against a defendant's objection that the target itself was chosen to find what it found, now recurring inside a federal audit rather than a trade association's white paper.

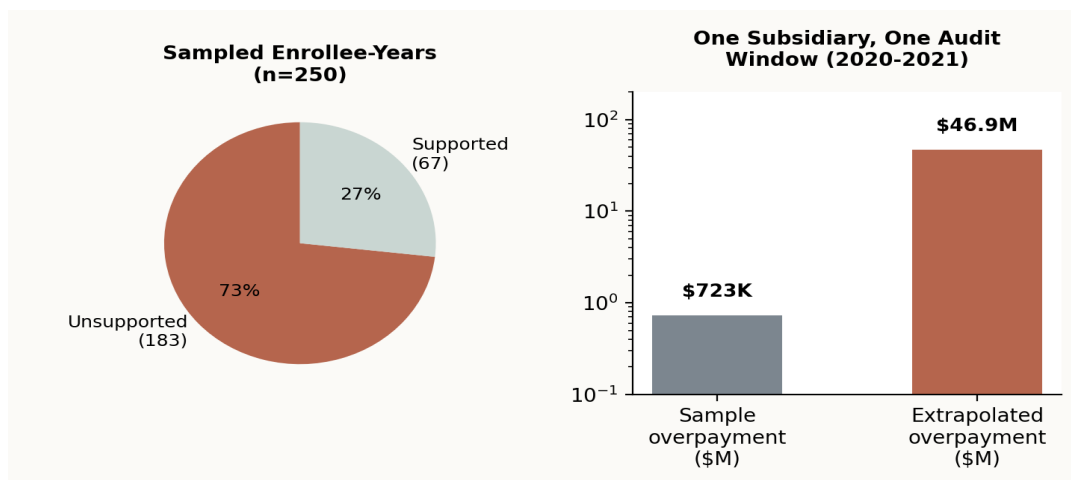


Exhibit 4. In a federal sample of 250 UnitedHealthcare Wisconsin Medicare Advantage enrollee years, nearly three in four diagnosis codes lacked chart support.

This is not an isolated finding against one company. OIG conducted a parallel audit of HumanaChoice using the same eleven high risk diagnosis categories and the same stratified sampling design over the same payment years, and reporting on both audits together found the same pattern, the large majority of each plan's sampled high risk diagnoses unsupported by the underlying medical record. The Medicare Payment Advisory Commission independently estimates that Medicare Advantage coding intensity, net of the statutory adjustment already

applied, will inflate 2026 payments to MA plans by 4 percent nationally, a congressional advisory finding cited directly in a March 2026 letter from Senator Jeff Merkley to CMS. Department of Justice settlements in 2026 alone include 556 million dollars from Kaiser Permanente in January, 117.7 million dollars from Aetna in March, and 541.5 million dollars from The Villages Health System in August, nearly 1.1 billion dollars in two settlements alone against the same general allegation: artificial intelligence assisted chart review, generating additional diagnosis codes for risk adjustment payment, unsupported by the underlying medical record.

FRONT 5: GOVERNMENT AI, THE INCENTIVE STRUCTURE THE GOVERNMENT IS SIMULTANEOUSLY INVESTIGATING

The Centers for Medicare and Medicaid Services Innovation Center's Wasteful and Inappropriate Service Reduction model, WISeR, launched January 1, 2026, with prior authorization requests accepted beginning January 5 for thirteen service categories in New Jersey, Ohio, Oklahoma, Texas, Arizona, and Washington.

CMS selected six technology companies as model participants, described in the program's own materials as firms with demonstrated success managing prior authorization for other payers, including Medicare Advantage plans, the same plans examined in Front 4.

CMS's own fact sheet states the model does not apply to Medicare Advantage beneficiaries and will have no impact on them; WISeR is Traditional Medicare only, extending this dynamic for the first time into the one part of the program historically understood as a refuge from prior authorization friction.

The payment structure closes the loop on this report's central thesis. CMS states its model participants are paid based on their ability to reduce unnecessary or non covered services and lower program spending, a savings contingent architecture structurally identical to the contingency fee model documented on the opposite side of this war in Front 1, and to the incentive already under OIG scrutiny in Front 4. The federal government did not design a new incentive to correct the pattern documented elsewhere in this report. It adopted the same one, aimed at itself, and staffed by contractors experienced in aiming it at others.

FRONT 6: VERTICAL INTEGRATION, ONE COMPANY, THREE SIDES OF THE SAME TRANSACTION

Optum, UnitedHealth Group's physician services and technology division, sits at the center of the measurement dispute documented in Part One and at the center of a structural conflict of interest no single regulator is positioned to examine in full. The same corporate parent can, in principle, employ or affiliate with the physician generating the clinical documentation on Front 1's coding intensity question, insure the patient described in that documentation and run the nH Predict style denial review on Front 2, and separately answer to federal auditors on Front 4 for how its Medicare Advantage subsidiaries code the diagnoses generating its own government payment. No dedicated body currently examines vertical integration of this kind as a unified risk; antitrust regulators examine market concentration, HHS OIG examines individual payment integrity, and state insurance regulators examine individual insurer conduct, but no single authority is charged with asking whether one company touching every side of a transaction chain is itself the risk.

Part Three: The Villain Hiding in Plain Sight

Every front documented in Part Two produces the same question in the mind of an informed reader: if the harm is this well documented, why does it persist. The answer is not regulatory inattention or corporate impunity in the simple sense. The answer is a specific, decades old body of federal law that most participants in this war, including many of the physicians, patients, and hospital executives living inside it, do not know exists.

The Employee Retirement Income Security Act of 1974, ERISA, governs the majority of Americans with employer sponsored health coverage. In *Pilot Life Insurance Co. v. Dedeaux*, decided in 1987, the United States Supreme Court held that ERISA preempts state law causes of action, including bad faith claims, against employer sponsored benefit plans. In *Aetna Health Inc. v. Davila*, decided in 2004, the Court extended this holding specifically to claims arising from wrongful coverage denials, ruling that when a plan administrator makes a coverage determination, even one alleged to have caused serious harm, the beneficiary's remedy is generally limited to the value of the benefit that should have been provided under the plan, not tort damages, not punitive damages, and not, in the ordinary case, wrongful death damages.

Justice Ginsburg, joined by Justice Breyer, wrote separately in *Davila* to make a point the majority opinion left unstated. Her concurrence describes the resulting legal landscape in terms this report adopts directly: a regulatory vacuum exists, in which virtually all state law remedies are preempted, but very few federal substitutes are provided. A sitting Justice of the United States Supreme Court, in 2004, named the exact structural gap this report is describing in 2026, more than two decades before an algorithm existed that could deny a claim in 1.2 seconds without a physician opening the file.

The consequence for every front documented in Part Two is direct. When a denial contributes to a patient's death, as alleged in the Lokken litigation, and the coverage arose through an employer sponsored plan, the family's remedy is typically capped at the value of the denied benefit itself. This is not a hypothetical reading of the doctrine; one federal filing states plainly that a plan denied a decedent benefits under an ERISA plan, causing death, and the claim was still held preempted. The financial deterrent that would ordinarily discourage a company from deploying an aggressive, error prone denial algorithm is, for the majority of the insured population, capped by federal law decided before the technology existed. This is why the Kisting-Leung litigation's surviving claims are narrowly contract flavored, breach of fiduciary duty under ERISA's civil enforcement provision, rather than the broader tort theories state law would otherwise supply. The doctrine is not a background detail; it is the reason every front in this war can operate at the scale it does without the legal system pricing in the full cost of the harm.

A number without a check on the party controlling it is not information. It is an unaudited weapon, and this report has now found six of them pointed at once.

Part Four: The Selective Labels Problem

Every front documented in this report so far rests on evidence already public: court filings, a federal audit, a trade association's own numbers, a Supreme Court opinion. This section is different: it is Koru Research Institute's own original analytical contribution, and it should be read as inference rather than established fact, clearly labeled as such, because that distinction is the entire discipline this report has applied to every other party's claims and must apply with equal rigor to its own.

Computer science research on algorithmic decision making has identified a specific, well documented failure mode called the selective labels problem, formalized in a widely cited 2017 paper by Lakkaraju, Kleinberg, Leskovec, Ludwig, and Mullainathan, originally developed to examine bail and pretrial risk assessment algorithms. The problem arises whenever a model's training outcomes are only observed for a non random subset of cases, specifically, the subset where a particular decision was already made. In the bail context: a judge's decision to detain a defendant means the outcome, whether that defendant would have committed a new offense if released, is never observed. The model can only learn from the cases where its own predecessor's decision allowed the outcome to be seen at all.

Apply this directly to the statistic anchoring Fronts 2 and 3 of this report. Only about 0.2 percent of denied claims are ever appealed, across two independently documented insurance markets, Medicare Advantage and the Affordable Care Act marketplace. If a denial algorithm is ever retrained on its own historical decisions and their outcomes, the training signal it receives is drawn almost entirely from the 99.8 percent of denials nobody challenged, not from an independently adjudicated determination that those denials were correct. A denial that was, in fact, wrong, but never appealed, looks statistically identical, from inside the model's own training data, to a denial that was genuinely correct. The model has no way to tell an error it got away with from a decision it got right, because the only cases that would teach it the difference are the cases systematically underrepresented in what it is trained on.

This report is not aware of any public disclosure, by UnitedHealth, Cigna, or any other insurer, of whether nH Predict, Px Dx, or any comparable system is in fact retrained using this kind of outcome data, and this section should not be read as an allegation that it is. What this report can state with confidence is narrower and, this institute believes, more useful: if any such system is retrained this way, the selective labels problem is not a remote theoretical risk; it is close to the specific, empirically grounded scenario the original 2017 research was written to warn against, applied here, for what this institute believes is the first time, to healthcare claims adjudication rather than criminal justice. A model built to get better at denying claims may, under these conditions, actually be getting better at denying claims nobody happens to challenge, which is a meaningfully different and considerably more dangerous thing for a health system to have built at scale without anyone checking. The corrective is not complicated to describe, even if it would be uncomfortable for an insurer to volunteer: validate a denial model's performance against outcomes from a randomly audited sample of upheld denials, not merely against the small, self selected fraction that happened to be appealed. No insurer examined in this report has disclosed doing so.

Part Five: What Happens When the Weapons Get Smarter

Every mechanism documented in Part Two depends on a large language model or a comparable machine learning system, and every one of those systems will inherit whatever capability gains the broader artificial intelligence industry achieves over the next several years.

This section examines that trajectory honestly, which means resisting the temptation to state a specific date with more confidence than the underlying forecasting research actually supports.

Daniel Kokotajlo's widely discussed AI 2027 scenario, published in April 2025, projected rapid automation of AI research and development compounding through 2027, a forecast Kokotajlo himself has since revised toward a longer timeline in a July 2026 follow up analysis. OpenAI has stated an internal target of March 2028 for full automation of its own AI research function. Anthropic's own research institute, discussing the same question, states plainly that current systems are not yet at this threshold while describing 2027 as a plausible point at which AI systems could reliably handle tasks spanning several weeks of equivalent human effort. Serious researchers, including Arvind Narayanan at Princeton, directly dispute the aggressive end of this range, arguing that AI capability diffuses through the economy over the course of decades rather than fiscal quarters.

The defensible version of the concern this section raises is a risk multiplier, not a prediction. Three of the six fronts in this war, nH Predict, PxDx, and the AI assisted chart review tools implicated in Front 4's upcoding findings, are systems that will likely inherit whatever frontier model capability gains arrive over the next several years, before the oversight mechanisms surrounding them, an OIG audit cadence measured in years, a 0.2 percent appeal rate that has shown no sign of improving, and ERISA litigation timelines measured in the Lokken case at three years and counting, have any realistic opportunity to adapt at a comparable pace. A capability gap that is already difficult to correct today is not a gap likely to remain static while only one side of it keeps improving.

The Playbooks Change on Both Sides, Not Just One

A smarter coding intensity model does not simply speed up what hospitals already do on Front 1. It finds combinations of clinical detail a human coder would never think to cross reference, a lab value on one page, a medication order on another, a nursing note three days later, and proposes a diagnosis code justified by all three at once. That is a different order of capability than an ambient listening tool catching a diagnosis a rushed clinician forgot to dictate, and it moves the entire coding intensity debate onto ground the BCBSA methodology in Part One was never built to audit, since a chart review looking for a single missed diagnosis is not designed to catch a diagnosis assembled from a pattern spread across an entire admission.

The insurer side of this war changes in the opposite direction and arguably faster. A denial model that improves at the same pace does not merely deny more claims. It becomes better at identifying the specific technicality, buried in a specific line of a specific prior authorization form, that makes a given denial defensible on paper even when the underlying care was appropriate. Cigna's PxDx already demonstrates what a comparatively simple system can do at scale, 300,000 denials in two months, 1.2 seconds each. A version of that system built on a materially more capable model does not need more time per claim to find more grounds for denial. It needs

less, which is the opposite of the direction every other safeguard in this system is built to expect.

Does a Smarter Algorithm Mean More Legal Exposure for Insurers

This is a genuine open question, and it is no longer a purely speculative one. Legal scholarship has already begun to move in this exact direction. University of Minnesota law professor Daniel Schwarcz, in a widely discussed article on AI claims handling, argues that insurance bad faith law has traditionally focused on the result, whether a claim was wrongly denied, and that AI claims handling introduces a distinct problem: the process itself can be unfair independent of whether any single denial turns out to be correct, a theory legal commentators have described as procedural bad faith. Separately, insurance litigation is already producing what practitioners have termed institutional bad faith claims, shifting the legal question from an individual adjuster's subjective intent to the insurer's own design choices in building and deploying an automated system, with at least one state court allowing exactly this theory to proceed against an insurer's algorithmic claims tools.

The selective labels problem documented in Part Four fits directly inside this emerging framework. It implies that a denial model retrained on its own outcomes has no reliable way to tell an error it got away with from a decision it got right, because only about 0.2 percent of denials are ever appealed. A more capable model does not correct this blind spot on its own; a more capable model becomes more confident inside it, since the underlying training data does not improve merely because the model examining it does. Under a procedural bad faith or institutional bad faith theory, that is no longer only a data science problem. It is a design choice a court can examine directly.

If an insurer's own data science team understands this dynamic, and continues retraining a denial model on its own historical outcomes regardless, the argument that the resulting errors were an unforeseeable byproduct of a well intentioned system becomes considerably harder to sustain in litigation. Courts examining the breach of fiduciary duty theory that survived in both the Lokken and Kisting-Leung litigation have generally asked whether a plan administrator's process was reasonable, not merely whether its outcome was correct. A known, published, named statistical failure mode, left uncorrected in a system an insurer continues to scale, is the kind of fact pattern that turns a process argument from a defense into a liability.

None of this changes the ERISA liability shield documented in Part Three. The cap on damages established in Davila and Pilot Life does not move simply because a denial model gets smarter. What may move is the standard for what counts as good faith administration under the narrower fiduciary duty and breach of contract theories that have already survived dismissal in this report's two central cases, and that distinction, not the size of any eventual judgment, is where the real legal exposure in this scenario is most likely to concentrate.

What Happens to Actuarial Science Itself

A separate and less discussed question concerns the actuaries and underwriters on the insurer side of this war, not merely the denial algorithms deployed against individual claims.

Actuarial risk models have traditionally been built on a relatively fixed set of variables, diagnosis codes, procedure codes, demographic factors, historical claims experience, refined slowly as new data accumulates over years. A materially more capable AI system applied to this same underwriting function could plausibly

identify combinations of variables, buried across thousands of pages of policy documents, medical records, and claims history, that a human actuary would never manually cross reference, the same capability shift documented above for coding and denial, applied one level further upstream, to how risk itself gets priced before a single claim is ever filed.

This report has found no evidence that this is happening today, and it should not be read as an allegation that it is. It is, however, a direct and logical extension of every mechanism already documented in this report, and it raises a question worth stating plainly for the reader qualified to investigate it further: if the same class of AI capability that is already finding new grounds for denial in individual claims is eventually applied to the actuarial models that set premiums and coverage terms in the first place, the six fronts documented in this report would no longer describe the whole war. They would describe only the fronts visible after the underwriting decision has already been made, with an entirely separate and largely invisible front sitting upstream of all of them.

Part Six: The Koru Six Front Index

Report No. 008 introduced the Koru Clinical Sovereignty Index to give health systems a disciplined way to score their own infrastructure dependency.

The Bottleneck Beneath the AI Boom introduced the Koru National Compute Sovereignty Index to apply the same discipline at the level of nations competing for AI infrastructure advantage.

This report introduces a third instrument in the same family, scoring each of the six fronts documented above across five dimensions: how strong the financial incentive driving the behavior is, how large the gap is between the harm caused and the legal accountability available to address it, how mature the applicable regulatory regime is, whether verified human harm has been documented rather than merely alleged, and how independently the front has been documented outside this report's own research.

Dimension	What It Measures	Score Range
Financial Incentive Strength	How directly the actor profits from the AI behavior in question, from contingency fees to risk adjustment payment.	1 to 5
Legal Accountability Gap	The distance between documented harm and the real financial consequence available under current law, informed directly by the ERISA analysis in Part Three.	1 to 5
Regulatory Maturity	Whether a dedicated regulatory body currently examines this specific front as its own risk category.	1 to 5

Verified Human Harm	Whether harm to a specific, named patient has been documented in a primary source, distinct from aggregate financial harm.	1 to 5
Independent Documentation	How many independent primary sources, beyond this report, have documented the front.	1 to 5

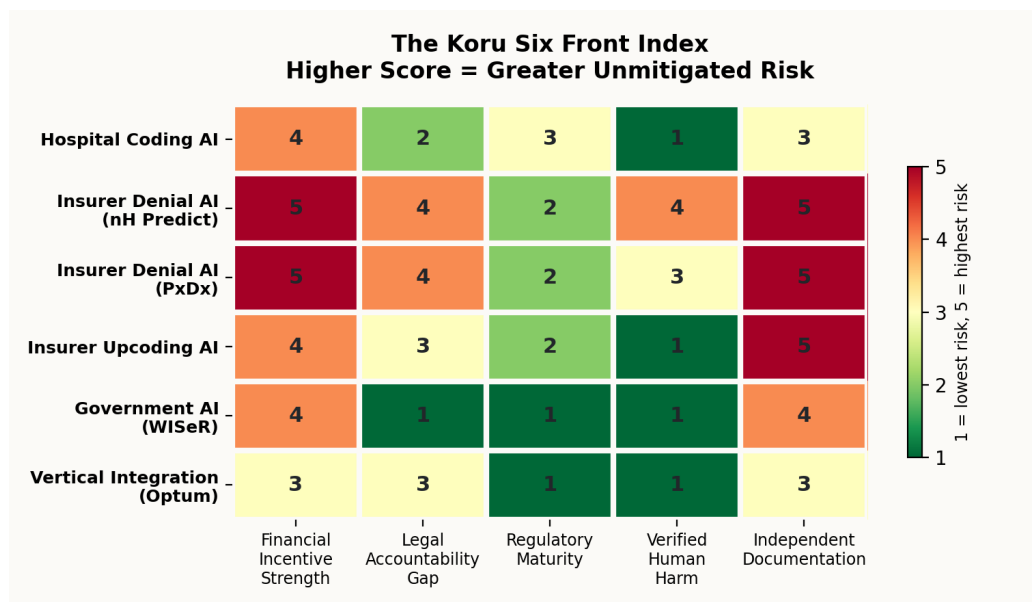


Exhibit 5. The Koru Six Front Index. No front scores low risk across all five dimensions; every front carries at least one serious, unmitigated exposure.

Applied to the evidence in this report, the pattern that emerges is itself a finding. The two insurer denial fronts, nH Predict and Px Dx, score highest on financial incentive, verified human harm, and independent documentation, reflecting their extensive litigation record, but score comparatively lower on regulatory maturity, since state insurance regulators and federal ERISA enforcement were never designed with algorithmic batch denial in mind. The government's own WISeR model scores among the lowest on legal accountability gap and regulatory maturity specifically, since it is new enough that no comparable oversight or litigation record yet exists, a genuine blind spot rather than a resolved risk. No front scores favorably across all five dimensions, which is the report's central structural finding restated in numeric form: this is not six problems awaiting six separate fixes; it is one incentive architecture, replicated with minor variation, everywhere it has been permitted to operate unaudited.

Where This Leaves the Reader

Six villains, named plainly. The measurement ambiguity that lets every party in this war define its own ground truth. The insurer denial algorithms that treat a 0.2 percent appeal rate as license rather than liability. The upcoding algorithms that turn the same capability toward the government the moment the payment direction reverses. The federal government's own WISeR model, built on the same incentive it is simultaneously auditing elsewhere. The vertical integration inside a single company that no regulator is positioned to see whole. And beneath all five, the ERISA liability shield, decided in 1987 and again in 2004, determining all along what any of this actually costs the party causing the harm.

The reader who finishes this report, whether a hospital executive, a health plan actuary, a congressional staffer, or a patient who has personally received a denial letter that felt as though nobody had actually read their file, should leave with one specific capability they did not have before: the ability to ask, of any single front in this war, what the other five fronts would predict about it. A denial that seems anomalous on Front 2 looks considerably less anomalous once Front 3 confirms the same mechanism at a second company, and considerably more urgent once Part Three explains why the legal system may never fully price in its cost. That is the actual argument of this report: not that any one front is the crisis, but that the crisis is the war none of the six fronts, examined alone, was ever going to reveal.

References

Aetna Health Inc. v. Davila, 542 U.S. 200 (2004).

American Hospital Association. Fact Sheet: AI and Coding Intensity. AHA, July 31, 2026.

Birkmeyer, C., Wennberg, D., Kamons, K., and Chalker, L. Hospital Coding Intensity Analysis: Major Bowel Procedures. Blue Cross Blue Shield Association white paper and press release, September 24, 2026.

Blue Cross Blue Shield Association and Blue Health Intelligence (Wennberg, D., Fressin, F., Everist, R., Horng, E., Nelson, E., and Clemente, S.). Rising Coding Intensity and Its Impact on Health Care Affordability. Issue Brief, March 2026.

Estate of Gene B. Lokken, et al. v. UnitedHealth Group, Inc., et al., No. 0:23-cv-03514 (D. Minn., filed Nov. 14, 2023).

Frazier, Marcus J. Healthcare Runs on Compute Now: How AI, Data Centers, Cloud Infrastructure, and Energy Are Reshaping Healthcare Delivery. Koru Research Institute, Report No. 008, 2026.

Frazier, Marcus J. The Bottleneck Beneath the AI Boom: How a Global Energy Average Hides Where Power, Water, and Chips Run Out. Koru Research Institute, Global Risk Brief, 2026.

Healthcare Dive. Federal Watchdog Accuses Humana, UnitedHealthcare Medicare Advantage Plans of Upcoding. September 2026.

Kaiser Family Foundation. Analysis of Medicare Advantage Denial and Appeal Rates. KFF, cited in Estate of Gene B. Lokken v. UnitedHealth Group litigation, 2023.

Kaiser Family Foundation. Claims Denials and Appeals in ACA Marketplace Plans. KFF, 2023.

Kisting-Leung, et al. v. Cigna Corp., et al., No. 2:23-cv-01477 (E.D. Cal., filed Jul. 24, 2023).

Kokotajlo, D., et al. AI 2027. Scenario report, April 2025; and AI 2040: Plan A, follow up analysis, July 2026.

Lakkaraju, H., Kleinberg, J., Leskovec, J., Ludwig, J., and Mullainathan, S. The Selective Labels Problem: Evaluating Algorithmic Predictions in the Presence of Unobservables. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017.

Medicare Payment Advisory Commission. Chapter 11, The Medicare Advantage Program: Status Report. MedPAC, March 2025 and January 2026 updates.

Merkley, J. Letter to the Centers for Medicare and Medicaid Services on Medicare Advantage Coding Intensity. U.S. Senate, March 30, 2026.

Pilot Life Insurance Co. v. Dedeaux, 481 U.S. 41 (1987).

Rucker, P., et al. How Cigna Saves Millions by Having Its Doctors Reject Claims Without Reading Them. ProPublica, March 25, 2023.

Schwarcz, D. Distributing Risk in an Age of AI: Procedural Bad Faith and AI Claims Handling. Law review article, 2026.

U.S. Centers for Medicare and Medicaid Services. Wasteful and Inappropriate Service Reduction (WISeR) Model, Fact Sheet, Frequently Asked Questions, and Provider and Supplier Operational Guide. CMS Innovation Center, 2026.

U.S. Department of Health and Human Services, Office of Inspector General. Medicare Advantage Compliance Audit of Specific Diagnosis Codes That UnitedHealthcare of Wisconsin, Inc. (Contract H5253) Submitted to CMS. Report No. A-07-24-01214, issued September 10, 2026.

U.S. Department of Justice settlements: Kaiser Permanente (\$556 million, January 2026); Aetna (\$117.7 million, March 2026); The Villages Health System (\$541.5 million, August 2026).

Washington Examiner. Minnesota Based UnitedHealthcare Overbilled Medicare by \$47 Million: HHS. September 2026.

About the Author

Marcus J. Frazier is Founder and Managing Partner of Koru Health, a Koru Capital Management Company, and leads the Koru Research Institute. A former intelligence officer trained in pattern of life analysis, he moved from national security into Wall Street risk and operations before founding Koru. His research examines healthcare operations, artificial intelligence infrastructure, institutional risk, workforce systems, and valuation risk in healthcare M&A and private equity, including Report No. 008, Healthcare Runs on Compute Now: How AI, Data Centers, Cloud Infrastructure, and Energy Are Reshaping Healthcare Delivery, Report No. 007, The Veto Outside the Model, and the Global Risk Brief, The Bottleneck Beneath the AI Boom: How a Global Energy Average Hides Where Power, Water, and Chips Run Out. He holds an MBA from Cornell and is a Georgetown alumnus.

Koru Research Institute is the research and publishing arm of Koru Health, based in New York, NY.

Connect with the author: [linkedin.com/in/markf91](https://www.linkedin.com/in/markf91) | [ORCID 0009-0005-5094-9040](https://orcid.org/0009-0005-5094-9040) | [SSRN Author ID 12444988](https://ssrn.com/author/12444988) | korucm.com