



# WHY PREDICTIVE MODELLING IS THE NEW STANDARD FOR HEALTHCARE CLAIMS AND CLINICAL INTEGRITY

## Abstract

Healthcare claims have always been connected to two dimensions. The first is that the clinical record should support the code submitted. And the second is that the code should satisfy the payer's policy. That framing is still valid, but the operating environment around it has changed. Payers have industrialised claims review using AI-driven utilisation management at a speed and scale that provider-side processes, still largely manual, cannot match. The rise in denial rates is a symptom of that asymmetry. Predictive modelling addresses it by shifting intervention upstream from claims rework towards pre-submission accuracy, from post-payment fraud audit to pre-payment risk scoring, and from retrospective documentation review to real-time clinical decision support.



Denial rates in healthcare claims have climbed steadily for years. According to Experian, about 30% of providers reported denial rates above 10% in 2022, 38% in 2024, and 41% in 2025. Meanwhile, the average amount of a denied inpatient claim rose 12% and outpatient denials rose 14% between 2024 and 2025. Add to that the scale of fraud, waste, and abuse embedded across the claims lifecycle — estimated at \$100 billion or more annually in the US alone — and the

picture that emerges is not one of isolated inefficiencies but a systemic integrity problem.

Predictive modelling is not new to healthcare. However, its scope of deployment and its practical reach have burgeoned and matured. The global predictive analytics market in healthcare is projected to hit \$30.7 billion by 2028, at a CAGR of over 22%. But the rather large figure masks a significant underutilisation

gap. About 95% of US hospitals now have access to advanced analytics, yet only 15% report any meaningful use of them for financial forecasting and strategic planning. This period presents a valuable opportunity to delve into what predictive modelling actually delivers across the claims cycle, where clinical integrity analytics adds the most value, and why fraud analytics has developed into an operational imperative.

## Why reactive tools must be put out to pasture

The rising denial rates since 2022 reflect increasingly complex payer requirements, more aggressive prior authorisation enforcement, and a widening gap between payer-side automation and provider-side capabilities. Medicare Advantage-related denied claim amounts rose 22.4% in a single year. Telehealth-related denials jumped 84%.

Working a denial consumes staff time, delays revenue, and often damages

patient experience when unexpected balances arrive or treatment is postponed. More than 40% of providers report that claim denials are now a substantial part of their daily operations. About half of the providers are still reviewing claims manually, which leads to an undesirable surge in cost as complexity increases.

Incomplete documentation, coding errors, eligibility mismatches, and prior authorisation failures account for the

majority of preventable denials. These are, unequivocally, data problems. Predictive analytics targets them at the origin, before a claim is submitted. AI-driven platforms trained on historical claims data can identify which claims carry the highest denial risk and why, allowing teams to intervene while there is still time. Of healthcare providers already using AI for this purpose, 69% report reduced denials or improved resubmission success rates.

## Predictive modelling for healthcare claims

Predictive modelling in healthcare claims processing combines historical claims data, clinical information, and statistical algorithms to identify patterns that precede adverse outcomes like denials, fraud flags, readmissions, or cost overruns. The process integrates data across multiple sources: EHRs, insurance claims, lab results, pharmacy records, and administrative systems, then applies machine learning to find patterns invisible to manual review.

Three core model types underpin most healthcare analytics use cases.

- Classification models use supervised machine learning to categorise claims

into outcome groups, such as likely approved, likely denied, and high fraud risk. Trained on sufficiently large historical datasets, they achieve accuracy rates that rule-based systems cannot match, particularly when payer requirements shift frequently.

- Clustering models group claims or patient records by shared attributes, identifying cohorts of patients with similar risk profiles or billing patterns that deviate from expected norms. They are particularly valuable in fraud analytics, where unusual billing patterns need to be detected without pre-

labelled training examples.

- Time-series models analyse patterns across defined time intervals to forecast future demand, patient volume, or denial surges linked to payer policy changes.

The limiting factor for all three is data quality. Models built on incomplete, inconsistent, or siloed data do not produce reliable predictions. Poor interoperability between EHRs, claims systems, and billing platforms remains another technical constraint, and no amount of algorithmic sophistication compensates for missing or inaccurate input data.



## The rise of clinical integrity analytics

Clinical integrity analytics extends predictive modelling beyond the claims submission event to the clinical decisions that drive it. The claim submitted for a high-cost procedure is the downstream output of a chain of clinical decisions, including the diagnosis, the treatment pathway, and the documentation. If any of those decisions are incomplete, inconsistently coded, or unsupported by the clinical record, the claim is vulnerable to denial or audit.

### At the clinical documentation stage

Predictive models in healthcare decision support can flag unsupported codes, identify missing authorisation

requirements, and detect documentation gaps while a clinician is still engaged with the patient's record. These models are planted at the clinical documentation stage, i.e., before billing. Generative AI, increasingly integrated into this workflow, can interpret complex payer medical policies in real time and provide guidance at the point of care.

### The population health dimension

Predictive models identify high-risk patient populations, like those with chronic disease burdens, high readmission probability, or gaps in medication adherence. This enables coordinated care interventions that reduce avoidable utilisation. It is itself a claims

integrity outcome with fewer emergency admissions, fewer unplanned procedures, and fewer claims that lack the supporting documentation.

### The dual benefit of hospital readmission models

Predictive scoring applied at discharge identifies patients most likely to return within 30 days, enabling targeted follow-up that reduces readmissions. It also reduces the administrative cost of working those post-readmission claims. Many of these arrive already under scrutiny by payers who apply their own predictive tools to identify potentially avoidable stays.

## Fraud analytics and modelling for prevention

Healthcare fraud, waste, and abuse represent one of the most financially detrimental integrity problems. The 2025 National Health Care Fraud Takedown charged 324 defendants in schemes involving intended losses above \$14.6 billion. Many of these frauds were identified through claims analysis. Estimates from HHS programme data suggest that \$30 to \$110 billion of Medicare expenditure annually represents some form of fraud, waste, or abuse.

Models trained on claims patterns can preemptively flag suspicious billing behaviour like upcoding, unbundling, duplicate submissions, and provider billing anomalies.

Clustering algorithms identify networks of providers, patients, and suppliers operating in coordinated fraud schemes. Anomaly detection models flag real-time deviations from a provider's historical billing norms. Predictive risk scores applied at the claim level prioritise human review

toward the cases where it has the highest return, rather than distributing audit effort randomly across a claims volume no team can review comprehensively.

The healthcare fraud analytics market is projected to grow steadily. That growth is driven by the documented scale of what unchecked fraud costs, and by the increasing inadequacy of rule-based detection systems as fraud tactics evolve to evade them.





## The challenges

Predictive modelling in healthcare has genuine limitations that optimistic market projections may understate.

### Algorithmic bias

Models trained on historical claims data inherit the inequities embedded in that data. If certain populations have historically received less comprehensive care, models trained on their clinical records produce predictions calibrated to that lower baseline. They may erroneously identify them as lower risk because their documented care history is sparser. The result can be systematic underservice of already-underserved groups. Detecting and correcting bias requires ongoing performance audits across demographic subgroups and continuous retraining on diverse datasets.

### Data interoperability

Healthcare data is often located across disconnected EHR systems, claims platforms, lab systems, and pharmacy databases that were not designed to communicate with each other. Predictive

models require clean, integrated, and complete data. Building that foundation is as important as the model itself and in many organisations, considerably more difficult.

### Explainability

A model that identifies a claim as high denial risk but cannot explain why in terms a billing specialist can act on merely adds complexity to it. Healthcare payers using AI to automate claim reviews face the same challenge from the other direction. Providers contesting AI-generated denials need access to the reasoning. Explainability is the cornerstone of AI ethics and accountability. It is a prerequisite for trust in the model's outputs, both internally and in patient contexts.

### Clinical adoption

A predictive model must be able to change its behavior to deliver accurate outputs. When analytical insight does not translate into clinical action, it leads to the most important challenge in healthcare predictive analytics. It is also the most

commonly underestimated concern.

Clinicians are already resource-constrained. A model that predicts risk on a separate dashboard, requires login to a separate platform, or interrupts established workflows will risk being ignored as a result of rational prioritisation. Successful implementation requires embedding predictions directly into EHR interfaces, presenting outputs in a language clinicians recognise, and designing alerts that guide action rather than simply raise an alarm.

Equally important is the question of clinical authority over model outputs. Predictive models should augment clinical judgment, not override it. For example, when a model flags elevated sepsis risk, the clinician decides how to respond. This principle is important for clinical governance as well as patient trust. There has been consistent concern about modes of communicating model outputs, the uncertainty around predictions, and whether patients understood the role of algorithmic recommendations.

## End note

Predictive modelling in healthcare claims and clinical integrity is gaining impetus due to

- Climbing denial rates
- Skyrocketing fraud losses
- Lack of trust in clinical documentation quality
- Technological gap between analytical capability and deployment in health systems

These causes represent a significant,

measurable competitive and financial disadvantage, which can be addressed with [comprehensive and AI-led healthcare claims processing services](#).

The organisations closing that gap are not waiting for a single integrated solution to emerge. They are building modelling capability into existing claims workflows, investing in data infrastructure that makes models reliable, and treating clinical integrity as an end-to-end operational discipline.

Modern enterprises are deploying predictive modelling across the claims lifecycle, from pre-submission clinical integrity analytics and denial prevention through fraud analytics, clinical decision support, and population health management. Thus, it becomes exceedingly important to build the analytical infrastructure that turns healthcare data into defensible, actionable outcomes.

## References

1. <https://www.experian.com/blogs/healthcare/state-of-claims-2025/>
2. <https://www.fiercehealthcare.com/finance/payer-audits-denial-amounts-rise-again-2025-vendor-data-show>
3. <https://scholar.smu.edu/datasciencereview/vol6/iss2/18/>
4. <https://kms-technology.com/blog/predictive-analytics-in-healthcare/>
5. <https://online.champlain.edu/blog/predictive-analytics-in-healthcare>
6. <https://www.medicaleconomics.com/view/2025-state-of-claims-does-payer-provider-collaboration-need-to-improve->
7. <https://scoop.market.us/healthcare-fraud-analytics-market-news/>
8. [https://drj.com/journal\\_main/healthcare-waste-ai-risk/](https://drj.com/journal_main/healthcare-waste-ai-risk/)
9. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11751774/>

For more information, contact [infosysbpm@infosys.com](mailto:infosysbpm@infosys.com)



© 2026 Infosys Limited, Bengaluru, India. All Rights Reserved. Infosys believes the information in this document is accurate as of its publication date; such information is subject to change without notice. Infosys acknowledges the proprietary rights of other companies to the trademarks, product names and such other intellectual property rights mentioned in this document. Except as expressly permitted, neither this documentation nor any part of it may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, printing, photocopying, recording or otherwise, without the prior permission of Infosys Limited and/ or any named intellectual property rights holders under this document.