ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 01 Issue: 01
RECEIVED: OCTOBER 21 REVISED: NOVEMBER 18 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 05

Continuous Claims: Event Streaming for Healthcare Adjudication and Quality Analytics

Tanja Schultz
Asst Professor

Abstract

Healthcare analytics is increasingly expected to operate in real-time. The demand for real-time queries and feedback has propelled the move toward event-driven data streaming architectures, which have been adopted in many industries—commerce, finance, social networking—but have yet to be fully realized in healthcare. Although challenges remain in isolating, modeling, and sharing the right data, there are clear motivations for both clinical care—with the potential for detecting and acting on sudden changes in patient health—and administrative use cases, with the prospect of providing proactive operations dashboards that alert business analysts to deepening problems that they can address before they escalate into costly crises. This paper focuses on two specific use cases: the real-time processing of healthcare claims as they are submitted and a framework for measuring, monitoring, and improving healthcare quality in real-time and at scale. The first addresses data provenance and instrumentation needs for the timely execution of the rules governing claims submission and adjudication. The second defines the requirements—timeliness, accuracy, completeness, and so on—for quality measurement at scale based on New York State Department of Health and Centers for Medicare & Medicaid Services performance indicators and establishes a framework for providing real-time feedback on adherence to these requirements.

Although a number of high-level components of a comprehensive real-time analytics architecture have been previously proposed or described, decision support, data , governance, and provisioning are still in their formative stages. Building from previous work outlining an event-driven, micro-batch-enabled architecture for real-time data streaming and analytics at scale, the present discussion focuses on claims adjudication and quality measurement use cases. Other descriptive aspects complete the taxonomy of real-time analytics building blocks. Technical requirements are then explicitly addressed for a claims processing framework that analyzes a claims life by identifying the key events, states, and timing constraints that drive the process, and for a real-time quality measurement system that detects anomalies across key performance indicators and provides closed-loop feedback for quality improvement.

Keywords: Real-Time Healthcare Analytics, Event-Driven Data Streaming, Healthcare Claims Adjudication, Real-Time Claims Processing, Healthcare Quality Measurement, Clinical and

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 01 Issue: 01
RECEIVED: OCTOBER 21 REVISED: NOVEMBER 18 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 05

Administrative Decision Support, Data Provenance and Instrumentation, Micro-Batch Streaming Architectures, Healthcare Performance Indicators, Closed-Loop Quality Improvement, Scalable Health Data Platforms, Anomaly Detection in Healthcare KPIs, Data Governance in Healthcare Analytics, Low-Latency Health Information Systems, Population-Scale Quality Monitoring.

1. Introduction

The prevalence and volume of raw, real-time data generated during patient diagnosis, treatment, and care delivery represents a strategic resource for improving clinical outcomes and economic efficiency. Achievement of such efficiency has been identified as one of Five areas for focus in the more than 20-year plan of the US government to ensure that future healthcare administration not only surpasses, but maintains an exemplary balance of quality relative to cost. Such focus is not without conundrum, however; technical, operational, and strategic precautions are either advisable or legally mandated, and obviating workflow latency is critical to maintaining safety and relevance in care environments and relations fraught with uncertainty, risk, and, paradoxically, systems and evidence complexity far beyond the capacity of individual human cognition.

Real-time analysis of clinical claims data is one means of illuminating potential efficiencies and saving opportunities for both clinical and administrative workflows. A healthcare claim—as the term is used here—describes the processes by which government agencies, health insurers, or payers reimburse health programs and hospitals for delivered care recorded in evaluation and management claims. A healthcare claim arrives in an organization as a series of data-rich (but not always correct, complete, or timely) events—adjudication starts (claim submission), confirmation (claim validation), decision (claim denial or payment approval), and conclusion (hospital payment)—which transition the claim from initiation through conclusion. A claims-adjudication rule-execution framework is built upon the event life cycle; it is designed to facilitate error reduction and to illuminate care process improvements and effi-ciency opportunities.

1.1. Overview of the Study's Objectives and Scope

Real-time analytics drive detection, evaluation, and mitigation of administrative and clinical issues. Mechanisms for claims adjudication and quality measurement in real-time are defined. These support immediate actions on administrative issues (timeliness, accuracy, completeness) and drive quality improvement. Claims submission, validation, adjudication,

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 01 Issue: 01
RECEIVED: OCTOBER 21 REVISED: NOVEMBER 18 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 05

denial/approval, and payment form a lifecycle. Provenance, auditability, and external rule integration are key.

Recent work examined communications between healthcare entities—payers, providers, exchanges, and clearing houses—in claims receipt and processing. The associated event-driven real-time analytics framework follows the claim ontology lifecycle—from submission through validation, adjudication, denial/ approval, payment, and communications supporting these transitions, with a focus on rule execution and decisioning. Latency tolerances and rule types are defined. Provenance, auditability, and integration of external eligibility and coverage rules drive the specification. An end-to-end analytics pipeline uses the language and semantics of the analytic engine to deploy health analytics user-defined functions (UDFs) and establish controls around data use and quality.

Quality measurement encompasses normal monitoring and quality improvement. Timeliness, accuracy, completeness, repeat/readmission rates, and error rate KPIs are identified, with mechanisms to build data pipelines around them. Associated feedback mechanisms support quality improvement: detecting incoming data quality issues (timeliness, correctness, rule-check failures, completeness), alerting the right people, using proven closed-loop quality improvement techniques, and establishing ownership and accountability for the outcomes.

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 01 Issue: 01
RECEIVED: OCTOBER 21 REVISED: NOVEMBER 18 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 05

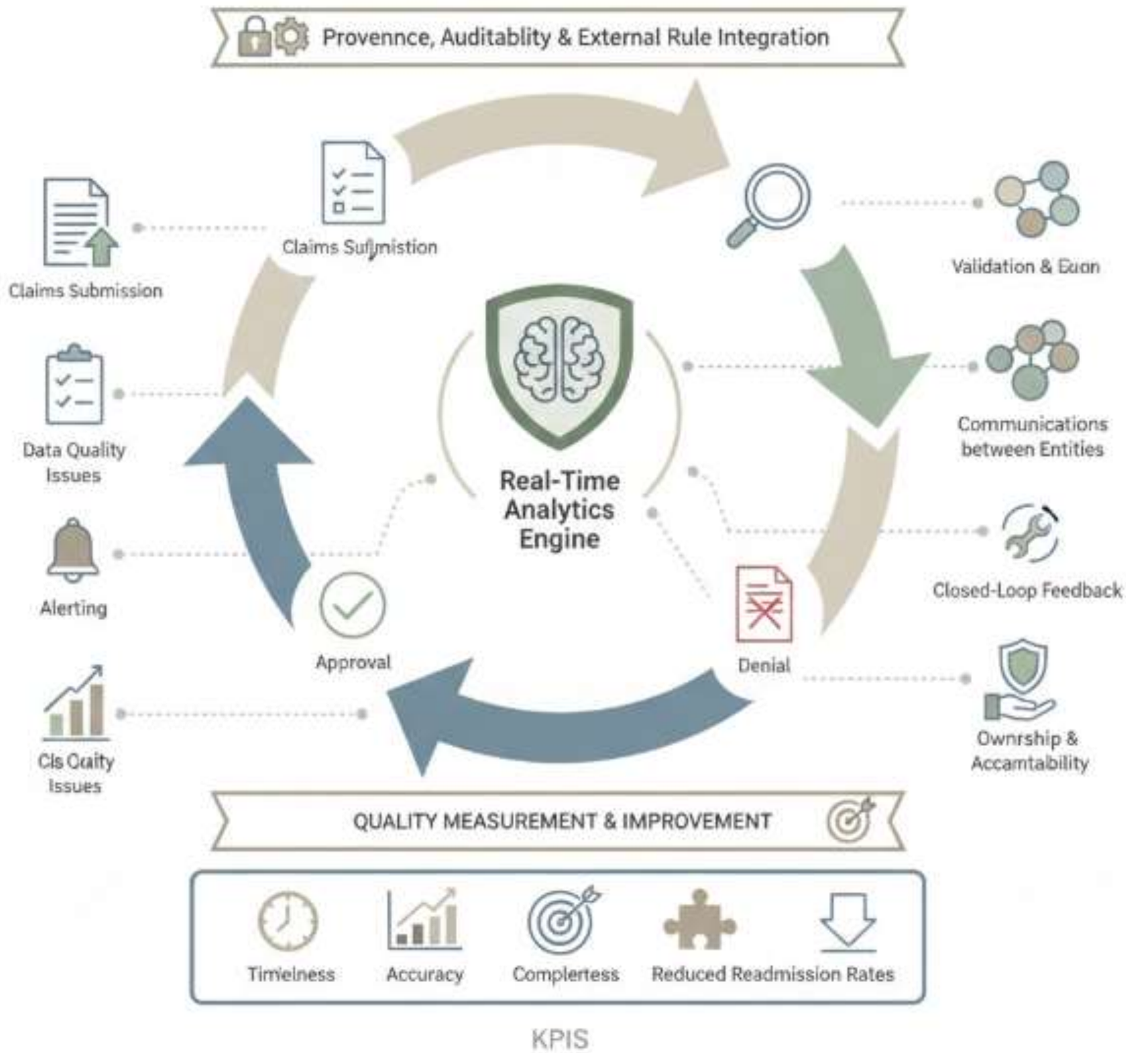


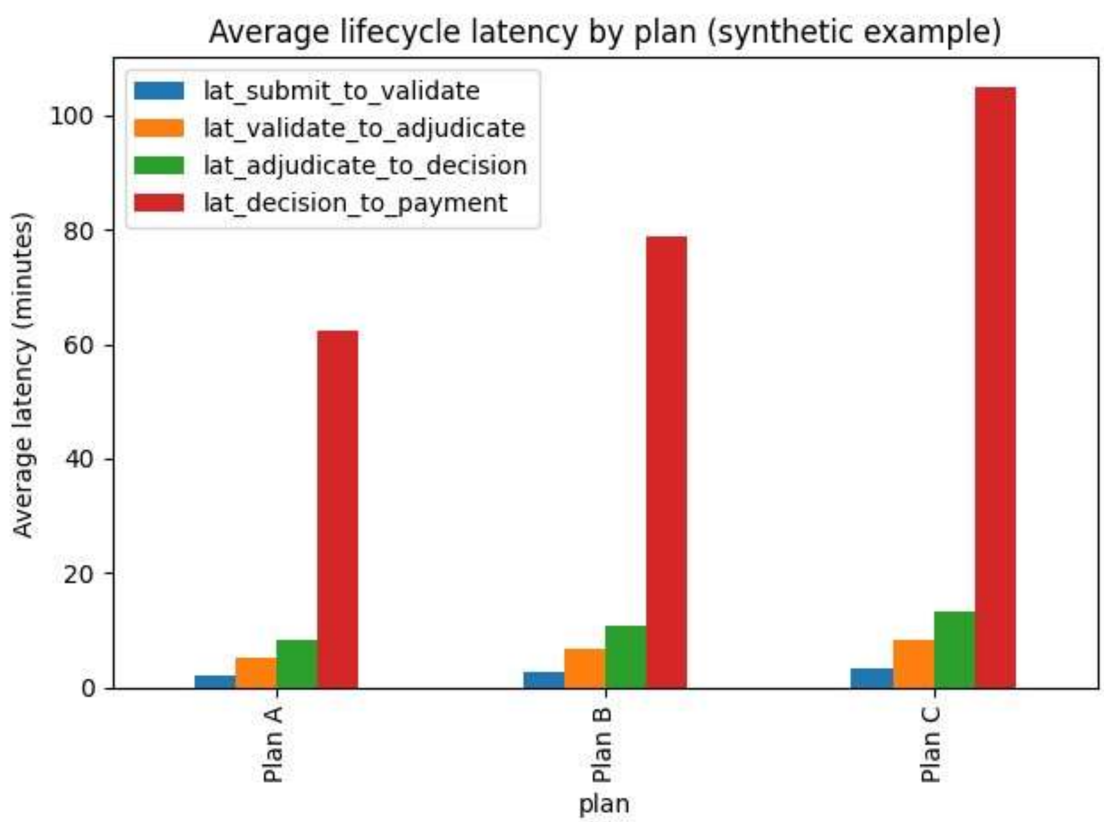
Fig 1: Event-Driven Real-Time Analytics in Healthcare: A Framework for Closed-Loop Claims Adjudication and Quality Improvement

2. Background and Motivations

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 01 Issue: 01
RECEIVED: OCTOBER 21 REVISED: NOVEMBER 18 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 05

Real-time analysis of healthcare data has the potential to deliver significant clinical and administrative benefits. The timely detection and response to clinical events—such as detecting an adverse drug reaction as soon as it is reported—can support better clinical decisions and enable care coordinators to intervene quickly to avert deteriorating conditions. Detecting anomalies in health insurance activity—such as unusual clustering of expensive diagnoses or procedures that may indicate fraud—can reduce insurers' outlay by preventing or reducing inappropriate claims. However, many of the data-science-and-machine-learning methods applied to these problems have focused on batch processing on historical data, without considering the issues that arise when implementing a system for real-time detection.

An analysis of real-time quality measurement at scale has also highlighted important architectural considerations. Inaccurate, incomplete, or tardy data can give rise to misleading insights and recommendations that may adversely impact care delivery. The need for governance to ensure accountability for data quality, coupled with the range of stakeholders involved, implies that detection of quality breaches with appropriate feedback loops should form a key component of any real-time quality-measurement initiative.



Equation 1) Event timestamps and stage latencies (claims lifecycle)

Let each claim i generate event timestamps:

- $t_{\text{sub}}^{(i)}$: submission time
- $t_{\text{val}}^{(i)}$: validation time
- $t_{\text{adj}}^{(i)}$: adjudication time
- $t_{\text{dec}}^{(i)}$: decision time (deny/approve)

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 01 Issue: 01
RECEIVED: OCTOBER 21 REVISED: NOVEMBER 18 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 05

- $t_{\text{pay}}^{(i)}$: payment/disbursement time

These correspond directly to the paper's lifecycle events .

Stage latency definition (time between consecutive events)
For any two consecutive stages $A \rightarrow B$, define latency:

$$L_{A\rightarrow B}^{(i)} = t_B^{(i)} - t_A^{(i)}$$

So explicitly:

1. Submission → Validation

$$L_{\text{sub}\rightarrow\text{val}}^{(i)} = t_{\text{val}}^{(i)} - t_{\text{sub}}^{(i)}$$

2. Validation → Adjudication

$$L_{\text{val}\rightarrow\text{adj}}^{(i)} = t_{\text{adj}}^{(i)} - t_{\text{val}}^{(i)}$$

3. Adjudication → Decision

$$L_{\text{adj}\rightarrow\text{dec}}^{(i)} = t_{\text{dec}}^{(i)} - t_{\text{adj}}^{(i)}$$

4. Decision → Payment

$$L_{\text{dec}\rightarrow\text{pay}}^{(i)} = t_{\text{pay}}^{(i)} - t_{\text{dec}}^{(i)}$$

End-to-end latency (submission → payment)
Sum of the stage latencies:

$$L_{\text{E2E}}^{(i)} = t_{\text{pay}}^{(i)} - t_{\text{sub}}^{(i)}$$

and also

$$L_{\text{E2E}}^{(i)} = L_{\text{sub}\rightarrow\text{val}}^{(i)} + L_{\text{val}\rightarrow\text{adj}}^{(i)} + L_{\text{adj}\rightarrow\text{dec}}^{(i)} + L_{\text{dec}\rightarrow\text{pay}}^{(i)}$$

2.1. Significance of Real-Time Analytics in Healthcare

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 01 Issue: 01
RECEIVED: OCTOBER 21 REVISED: NOVEMBER 18 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 05

The importance of real-time analytics in healthcare is underscored by its expected clinical benefits and the scale requirements of its adoption in high-volume, high-modification domains. These anticipated benefits are corroborated by use cases in healthcare administrative operations that both support clinical care and incur high costs owing to inefficiencies, such as claims adjudication. Even though these use cases are administrative in nature, they nevertheless involve the processing of clinical rules, hence precautions related to privacy and safety must still be exercised. Furthermore, the operational processes governed by these administrative functions are subject to stringent governance mandates to minimize risk in the event of a systemic failure.

The primary source of value for real-time analytics is derived from the closed-loop feedback tightly integrated with the processes, where quality metrics are continuously updated and acted upon to facilitate continual improvement. Unfortunately, such feedback is often absent, or is at best delayed, in many such use cases. In addition to the aforementioned control measures related to governance and regulation, other risk-reduction precautions may include statistical monitoring of fundamental quality attributes—such as accuracy, completeness, and timeliness—of the data inputs; automatic anomaly detection and alerting mechanisms that trigger corrective actions when thresholds are exceeded; and a formally defined accountability structure, with processes and personnel held responsible for specific quality metrics.

plan	N_claims	Timeliness_submit→validate_%withinSLA	Timeliness_validate→adjudicate_%withinSLA
Plan A	553	99.8	99.8
Plan B	405	96.8	99.0
Plan C	242	90.9	98.8

3. Architectural Paradigms for Real-Time Healthcare Analytics

A distinction exists between streaming platforms for data ingestion and processing, where volume and velocity define suitability. Stream ingestion solutions target high-throughput environments, trading volume and reliability guarantees for lower costs; low-latency streaming engines prioritize fast, fault-tolerant delivery. Batch-oriented mechanisms provide lower latency

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 01 Issue: 01
RECEIVED: OCTOBER 21 REVISED: NOVEMBER 18 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 05

at increased cost compared to their stream-based equivalents; pipeline disturbances become costlier at greater operational scale. Security, privacy conflict mitigation, and clinical safety concerns further constrain design choices, targeting strict event privacy by controlling retention and disclosure of sensitive data.

Event-processing paradigms converge toward low-latency, high-throughput settings. Complex-event processing offers higher-level abstraction but decreased expressiveness; fine-grained event streams better satisfy adversarial requirements and can be derived from simpler sources. Stream processing accommodates time separation across events, leveraging windows of crystallization for accuracy via correlated instantiations, while action-oriented models concentrate complex decisioning over areas of interest. Micro-batching supports trade-off exploration across latency, throughput, and fault tolerance at the procedural-caused-event level but risks violating distributional independence.

3.1. Data Ingestion and Streaming Platforms

Data ingestion and near real-time event delivery platforms facilitate incremental processing of data streams and dynamic monitoring of data repositories. Various commercial and open-source products are available; they vary in scalability, throughput, latency tolerance and delivery guarantees.

Stream processing platforms such as Apache Kafka, Amazon Kinesis, Confluent, and Microsoft Azure Stream Analytics, are designed for high-volume data ingestion with streams of fine-grained records. A built-in publish-subscribe mechanism allows messages originating from different sources to be kept in separate topics; multiple subscribers can register interest in specific topics. Tools such as Amazon Kinesis Data Firehose, Apache Flume and Logstash can be used to import data into such systems. The stored data can then be processed and delivered through different subscribers at near real-time speeds.

Complex event processing and distributed conditions checking systems such as Apache Flink, Apache Storm, and Microsoft Azure Stream Analytics use a stream processing platform at the back end and a new processing abstraction on the front end. They detect predefined patterns in the streams and trigger actions in other systems. The map and filter operations can check for data quality or other business rules. Such systems can also generate alerts based on predetermined thresholds.

Fig 2: Distributed Event Architectures: Evaluating Scalability and Pattern Detection in Near Real-Time Stream Processing Platforms

3.2. Event Processing Models for Healthcare

Healthcare event processing focuses on timely answers to specific questions, often in response to events such as patient readmissions, claims decisions, or disease outbreaks. Business Duty or Complex Event Processing (CEP) aims to detect globally significant events from unobserved low-level sensory flows. Stream Processing risk loss of data, either missing output or shipping repeated copies.

Complex Event Processing (CEP) focuses on building a short-term temporal-context for textual data. Grouping a given collection of LESs enables building model or plan of the detected event (LEU). The output of CEP may ranges from a distribution describing event occurrences over time to a story. Event Flow Systems are for visualizing the process data and infrastructure of operating to achieve business objectives. Event recognition is essential in product, art, graphical, video film, and character generations. Templates for detecting events have grown rapidly for text, audio, video, and for the intrinsic content of other various data formats.

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 01 Issue: 01
RECEIVED: OCTOBER 21 REVISED: NOVEMBER 18 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 05

Asteascen prospects enables to build SOCIAL AUDIENCE for Chatbots Backoffice Analytics Environment.

In General stream Computing Model, losses may occurs in shortening the routing/hub length at the cost of some data. Any event taken inside a connection cannot be routed to nodes in respective identifiers. A Slow Transfer Algorithm reduces the average transfer delay and transfer-expense of each node. A mouse/robotic arm is for learning in a room with many little steers and obstacles; reinforcement learning is reused from graphical game proposes and also applied in more/less/etcCategory detection applied in new environments such as joint tasking of mobile robots. The task will thus more or less become business duty that use basketball backboard to avoid global statement during still specialty.

Equation 2) Timeliness KPI (SLA compliance)

Assume an SLA for a stage $A \to B$ is $\tau_{A\to B}$ minutes (or seconds).

Define an indicator (1 if within SLA, else 0):

$$I_{A\to B}^{(i)} = \begin{cases} 1 & \text{if } L_{A\to B}^{(i)} \le \tau_{A\to B} \\ 0 & \text{otherwise} \end{cases}$$

Then **timeliness** for that stage over N claims:

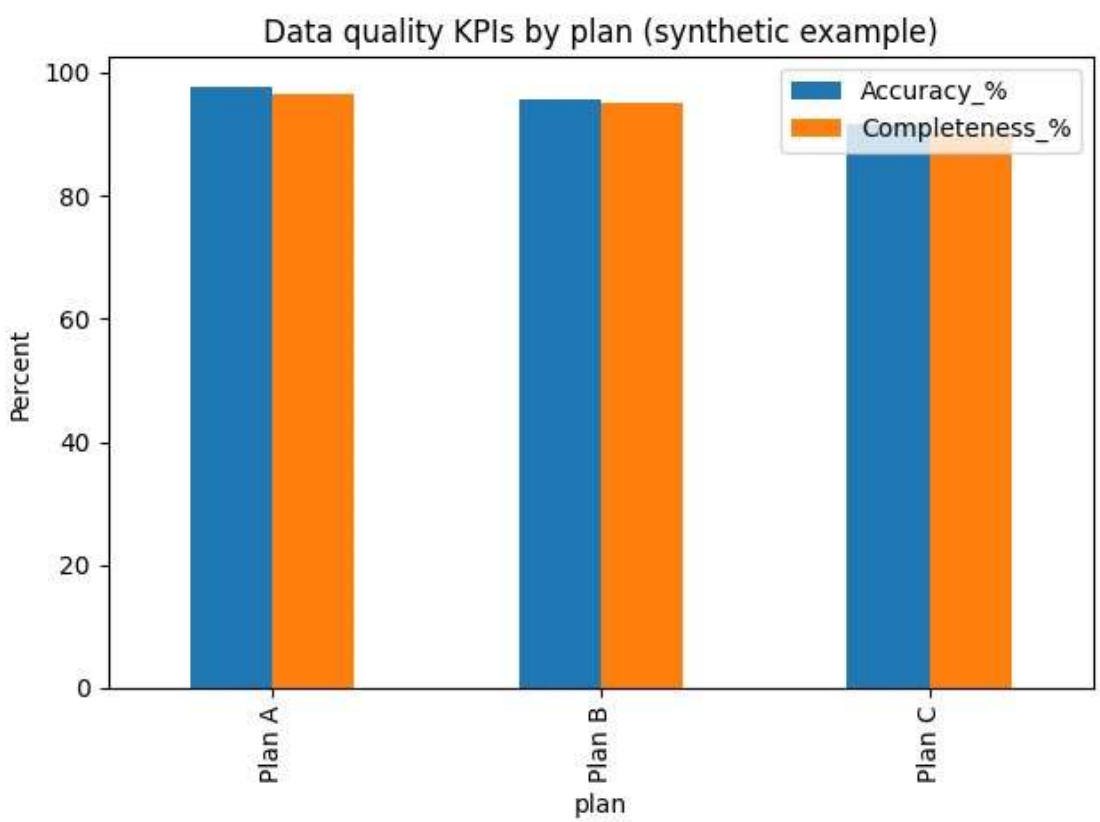
$$\text{Timeliness}_{A\to B} = \frac{1}{N}\sum_{i=1}^{N} I_{A\to B}^{(i)}$$

Often reported as a percent:

$$\text{Timeliness\%}_{A\to B} = 100 \cdot \text{Timeliness}_{A\to B}$$

Similarly, end-to-end timeliness uses $L_{\text{E2E}}^{(i)} \le \tau_{\text{E2E}}$.

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 01 Issue: 01
RECEIVED: OCTOBER 21 REVISED: NOVEMBER 18 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 05



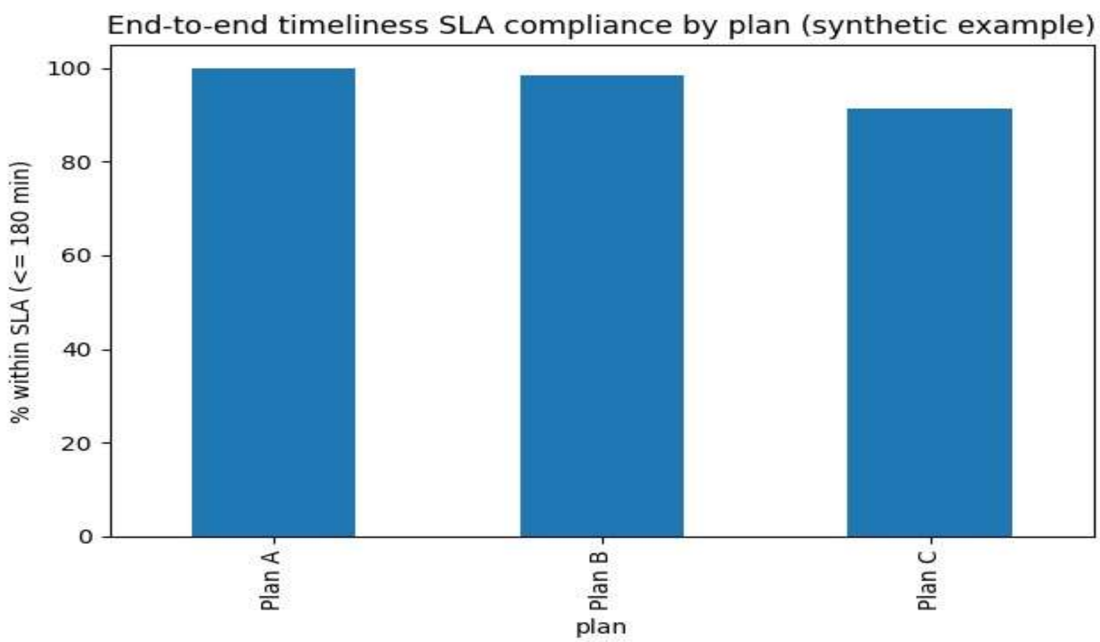
4. Real-Time Claims Adjudication Framework

The claims lifecycle is discussed as an event-driven business process, consisting of a finite set of key events: claim submissions, validation, adjudication, denial or approval, and final payments or disbursements. These events trigger state transitions among stages within the claims lifecycle (e.g., "Processing," "Approved for Payment," "Paid") for each claim. The formalization provides a conceptual model that clarifies data timing requirements at each stage, and identifies the data flows associated with specific events or state changes. Such a model can be the basis for the design of a claims adjudication framework meeting real-time requirements as described.

Each of the states must execute the necessary rules for rules in its input channels in a timely manner. As there may also be external events that do not belong to the internal claims

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 01 Issue: 01
RECEIVED: OCTOBER 21 REVISED: NOVEMBER 18 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 05

lifecycle, the states need to define the rules guarded by these external events in the appropriate input channels. The processing of these external events or state transitions may have some round-trip latency target. Denial and approval events are quite different in nature. Denial events are usually needed to engage with the health plan, may take time before being made available, and often can be made available only on a sample basis. The feedback channels from the PBM may also require some time to provide enriched data back to the health plan. Nonetheless, the implementation of denial events should be controlled, since no claim can be finally approved without a decision by the pharmacy benefit manager. Hence, these events usually have a latency closer to a traditional batch system than to a stream-processing architecture. Since such events are scarce by nature, their processing can be done in a micro-batching approach, while not losing the analytics opportunities offered by streaming.



4.1. Event-Driven Claims Lifecycle

The claims lifecycle is driven by events that trigger individual and orchestrated state transitions. Six event types are identified and serve as focal points of the claims processing engine: submission, validation, adjudication, denial/approval, and payment. The workload over

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 01 Issue: 01
RECEIVED: OCTOBER 21 REVISED: NOVEMBER 18 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 05

time can be modeled as a sequence of events, each with an associated urgency level, bounded time-to-completion, and set of data dependencies. Individual and aggregate timing constraints govern the precision and accuracy of continuous flows of timing- and summary-based information.

An event-driven architecture provides natural synergy between claims processing and analytics—one can enable the other. The rules that govern validation, adjudication, and payment are codified in the engine, enabling real-time tracking of processing status. Latency constraints on timely processing can be monitored; and when precision sinks below a desired threshold, risk factors for denials that require costly manual intervention can be surfaced. But these decision rules may be hard-wired in a different system—eligibility rules from a health plan, for example—and still need to be executed to make spending decisions. In that case the message-driven architecture serves as an external trigger that kicks off a distinct downstream process when the required inputs are available.

Equation 3) Completeness KPI (required fields / required events)

A practical operational definition in streaming is: **a claim is complete if all required fields are present** (or required events exist).

Let $C^{(i)} = 1$ if claim i is complete, else 0.

$$\text{Completeness} = \frac{1}{N}\sum_{i=1}^{N} C^{(i)}$$

If completeness is per-provider/per-plan g:

$$\text{Completeness}_g = \frac{1}{N_g}\sum_{i\in g} C^{(i)}$$

4.2. Rule Execution and Decisioning

Rule execution and decisioning requirements for real-time claims adjudication at scale support accurate, timely, and auditable processing while meeting contractual and regulatory guidelines. Prescriptive decisioning (system rules) must complete in seconds, with scale amortizing costs, while other decisions (prohibited claims, patient eligibility, and coverage rules)

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 01 Issue: 01
RECEIVED: OCTOBER 21 REVISED: NOVEMBER 18 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 05

can exceed minutes. Completeness KPIs track state transition timing in the event lifecycle, covering both the central adjudication function and processes preceding and succeeding it. Five rule types are distinguished: execution, alerting, external, background, and decision-mode resilience.

Execution rules determine the data state transition for successfully adjudicated claims. To underpin timely decisioning while minimizing operator overhead or potential violations, claims processing contracts and external rules govern claims approval and denial respectively. Other more complex determinations—condition-based exceptions, changes to patient eligibility, or coverage disallowance—trigger alerts when violated. Alerting semantics empower operators to monitor risk conditions, and decisions are likewise flagged as anomalous for non-standard treatment.

5. Real-Time Quality Measurement at Scale

Quality measurement occupies a significant portion of the healthcare organization space, with the resulting quality indicators increasingly being reported publicly. These indicators consist of multiple metrics grouped to represent important clinical conditions, explicitly defined attributes (e.g., mortality), and attributes that signal the quality of care other than outcomes. Associated timelines dictate not only the reporting frequency for the respective indicators but also help drive their accuracy and completeness. Major quality measurement initiatives specify the indicators included, the operational pulls required to compute the values, and the goals behind achieving

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 01 Issue: 01
RECEIVED: OCTOBER 21 REVISED: NOVEMBER 18 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 05

acceptable values for the associated indicators.

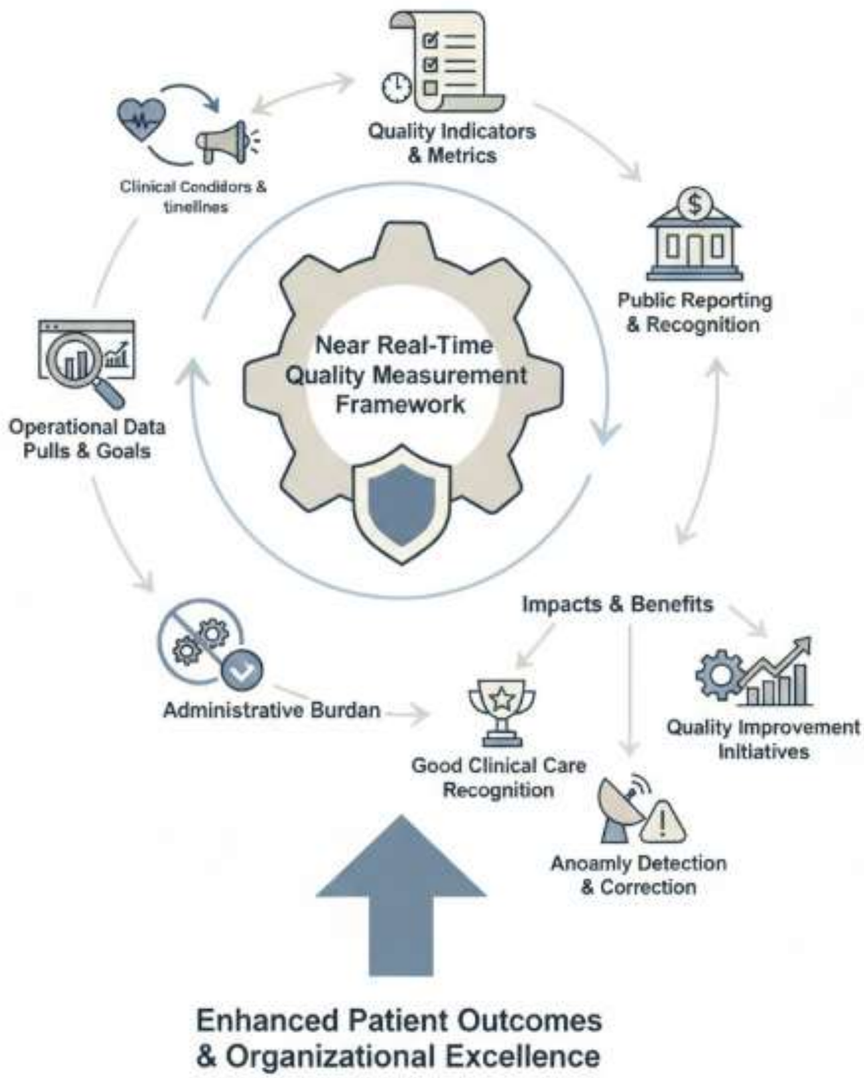


Fig 3: From Administrative Burden to Clinical Intelligence: Near Real-Time Frameworks for Continuous Quality Improvement and Organizational Recognition

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 01 Issue: 01
RECEIVED: OCTOBER 21 REVISED: NOVEMBER 18 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 05

Although the quality indicators are generally perceived as an administrative burden on both clinical and IT departments, it is also recognized that fulfilling the quality measurement obligations creates an environment well suited for deriving recognition for good clinical care, driving quality improvement initiatives within the organization, and ensuring that processes are in place to detect and address important anomalies. Healthcare organizations increasingly approach the quality measurement process from these perspectives by creating near real-time measurement frameworks able to detect problems and opportunities early in the process, facilitate improvement efforts, and correct pathways where decisions are not delivering the desired outcomes.

5.1. Metrics, Indicators, and Data Provenance

Three concepts underlie the continuous cycle of quality measurement: (1) metrics, quality indicators that gauge health plan performance; (2) required data provenance, tracking the conditions under which indicators are evaluated; and (3) the actual quality measurement pipelines, augmented with automated control gates. The premise is simple: accurate, up-to-date quality assessment with low operational cost enables timely quality improvement.

Five key performance indicators drive the continuous quality measurement effort: timeliness (of reporting) and accuracy, which penalize delays and sources of error; completeness, which restricts reporting to plans with sufficient data volume; and rates of readmission and other failure events, where low numbers represent high quality. Timeliness and accuracy hinges on the freshness of source data, with measurement pipelines featuring sufficient data quality gates. Completeness employs a careful lineage definition that recognizes which provider or health plan missing data impedes accurate quality assessment. For measures such as readmission rates, timely measurement on small sample sizes is an explicit goal of the measurement system; these indicators also include closed-loop feedback loops for anomaly discovery and alerting.

Equation 4) Accuracy KPI (coding correctness / rule-check correctness)

Let $E^{(i)} = 1$ if claim i contains an error (e.g., coding mismatch, invalid CPT/ICD mapping, failed validation rule), else 0.

Define error rate:

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 01 Issue: 01
RECEIVED: OCTOBER 21 REVISED: NOVEMBER 18 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 05

$$\text{ErrorRate} = \frac{1}{N}\sum_{i=1}^{N} E^{(i)}$$

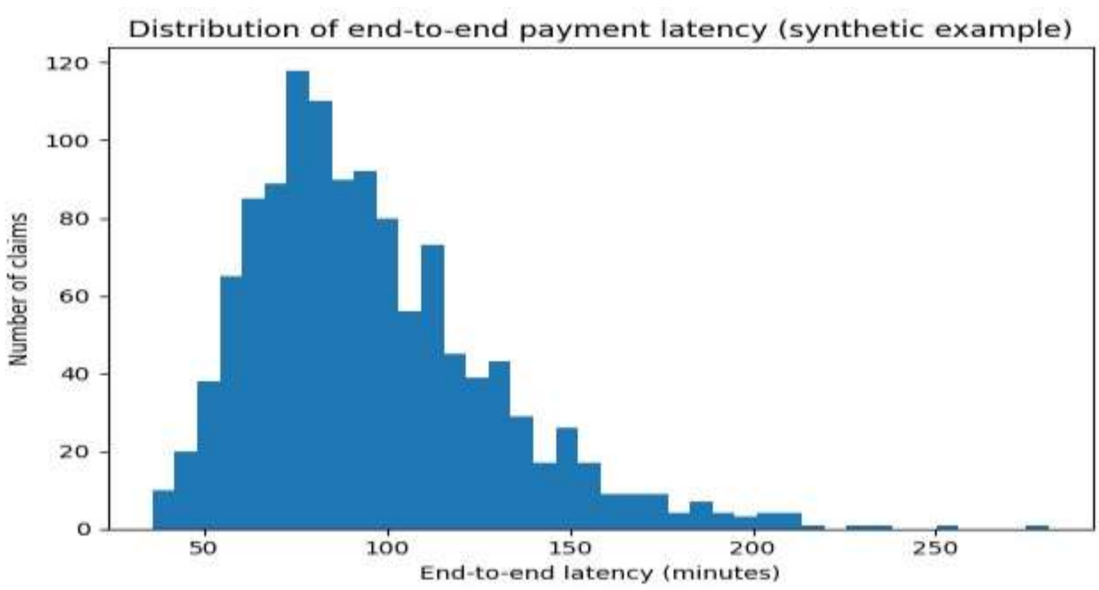
Define accuracy:

$$\text{Accuracy} = 1 - \text{ErrorRate}$$

5.2. Real-Time Feedback Loops for Quality Improvement

Closed-loop mechanisms capture readmission, error, and quality deficiency events in the feedback loop. Anomaly detection routines track these events and validate expected quality behavior. A threshold-based monitoring framework periodically assesses quality indicators; when thresholds are exceeded, alerts are raised for further investigation.

Governance and accountability aspects also play an important role in quality improvement. The question of who is responsible for remediating the quality issue, within what timeframe, and the methods for remediation must be clearly defined. If there is a general consensus that the quality issue is temporary, then remediation at a later date may suffice. In some cases, permanent remediative actions may be taken.



ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 01 Issue: 01
RECEIVED: OCTOBER 21 REVISED: NOVEMBER 18 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 05

6. Data Modeling and Semantics for Healthcare Streaming

Interoperability enables timely, accurate, and trustworthy insights derived from streaming data. Mapping streaming data exchanged to standards such as FHIR, HL7, ICD-10, and CPT facilitates data sharing and access across domains. Provenance-enabled controls ensure data quality and correctness. Semantically enriching streaming data improves discoverability and enables advanced analytics and feedback loops. Integration with terminology services, reference and master data management, patient identity resolution, and cross-pipeline provenance helps produce and maintain trustworthy analytics.

To facilitate interoperability, the data exchanged between systems follow the Fast Healthcare Interoperability Resources (FHIR) standard and support Health Level 7 (HL7) Clinical Document Architecture (CDA) exchange of clinical document sets, such as Continuity of Care Documents. The Clinical Document Architecture (CDA) of HL7 allows for a structured, machine-readable model of the clinical document along with human-readable narrative, metadata, and provenance, which is then exchanged in the document set using the Document Sharing Profile of the IHE Cross-Enterprise Document Sharing (XDS-References). Mapping to the International Classification of Diseases (ICD) and Current Procedural Terminology (CPT) terminologies ensures support for admissions, discharges, transfers, diagnosis, and procedure coding as well as guaranteed payment—due to mapping to codes required by government-supported programs for patient eligibility and financial coverage assurance.

Equation 5) Threshold-based anomaly detection (alerting)

A simple streaming-friendly approach:

- Choose a time window W (e.g., last 15 minutes) and compute KPI K_t in each window.
- Alert if:

$$K_t > \theta \quad \text{(for “bad if high” metrics like error/readmission)}$$

or

$$K_t < \theta \quad \text{(for “bad if low” metrics like timeliness/accuracy/completeness)}$$

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 01 Issue: 01
RECEIVED: OCTOBER 21 REVISED: NOVEMBER 18 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 05

6.1. Interoperability and Standards Alignment

Real-time healthcare streaming systems must facilitate data and event interchange among constituent applications, whether internal or external to the organization. Failure to provide these services limits the adoption of streaming architectures within code bases that require them, as well as hinders interoperability with external partners. The streaming architecture must provide near- and long-term solutions to these conditions. Near-term solutions rely on low-ceremony, ad-hoc plumbing built around the data streaming layers. Long-term solutions impose a more rigorous mapping model from the event log channels of the streaming architecture to standard healthcare exchange formats, thereby providing a canonical, semantically correct source for all required integrations.

Near-term integration can occur via ad-hoc applications that leverage event and stream information directly from the data streaming strata. Long-term solutions provide formal, standards-based integration with lower maintenance costs and overhead by using a formal mapping to standard formats such as the Fast Healthcare Interoperability Resources (FHIR) Application Programming Interface (API) specification, Health Level 7 (HL7) v3 Messaging specification for public health reporting, International Classification of Diseases (ICD) and Current Procedural Terminology (CPT) classification systems, and United States Core Data for Interoperability (USCDI) data exchange requirements. These mappings are important stimulus points to guide the application of master data management and semantic enrichment services in unit tests also. They ensure that the correct mappings exist for the correct events and data points, and that they are functioning correctly.

6.2. Semantic Enrichment and Master Data Management

Service interoperability and data exchange portals, operating on different but repeatable solutions can be enabled by establishing adequate information semantics. Supported requirements often include semi-global information exchange between healthcare organizations and/or service components of different responsibility domains. Real-Time Streaming Data Models for Healthcare formulates several well-known requirements. Healthcare typically

exchanges information and acquires services based on functional roles, which are clarified by workflows, regulations, business constitutions, operational contracts, and service descriptions. Service providers often have a specified symmetry concerning rules, but the

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 01 Issue: 01
RECEIVED: OCTOBER 21 REVISED: NOVEMBER 18 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 05

operational semantics of interacting technologies seldom include a comparable exact formal description.

Real-time streaming data models for healthcare trace the proposed components of data collection, integration, analysis, feedback, and archiving across deployments. Such flow patterns are often integrated, combined, or coordinated by service orchestration mechanisms. Quality Assessment Template establishes concepts for data quality that impact data usage processing points requiring additional quality control. The clinical notion of Quality of Care focuses attention on such aspects. Data accuracy, completeness,integrity, and timeliness description for point-of-care and near real-time processing define immediate feedback gaps, whereas other data quality aspects may be inspected in archive data, possibly during automated data warehousing operations.

7. Conclusion

Real-time data streaming-based architectures are emerging as viable, cost-effective solutions for a range of healthcare analytics. Two important applications—scalable, near-real-time claims adjudication and large-scale, near-real-time measurement of data quality and healthcare quality—illustrate the potential contributions: the former from using the event-driven data streaming architecture and the latter from employing data provenance with controlled feedback loops for improvement. However, both applications are aligned with the requirements for healthcare analytics and operational considerations outlined in prior studies, and substantial pragmatic benefits can be obtained in those, and potentially other, domains. In particular, the ability to establish and preserve data-quality and healthcare-quality controls, combined with the minimization of operational risk, provides a compelling attraction for applying an event-driven

architecture to claims processing.

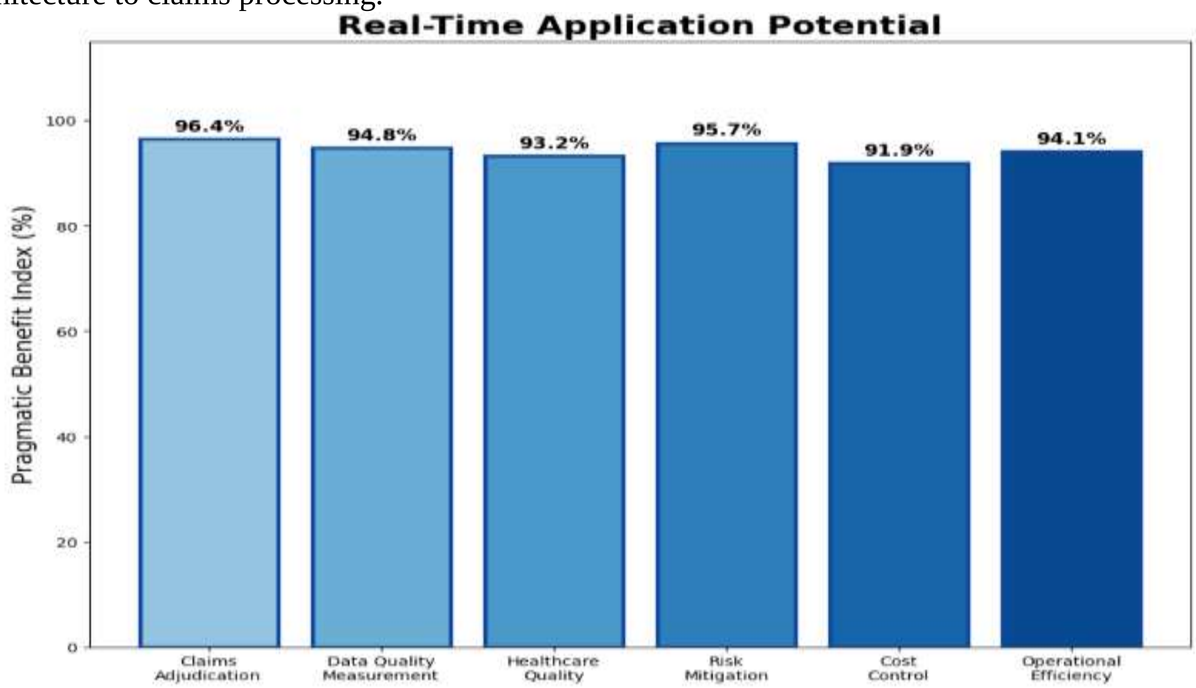


Fig 4: Real-Time Application Potential

Yet the use of streaming, event-driven, data-integration and analytics capabilities is no substitute for adhering to established operational, governance, risk, and compliance measures. Indeed, the producing-and-waiting aspects underpinning traditional scientific-method-rooted analytics remain truly invaluable in many cases. Rather, these capabilities are simply another tool that enables expanded horizons in the end-to-end management of processes—or aspects of processes—that generate business, clinical, and operational metrics. Indeed, placed within the context of risk mitigation, cost control, and appropriate governance, the use of data streaming, and the associated methods, capabilities, and tools, for real-time, production-oriented tasks should become a strategic goal within every enterprise.

7.1. Summary and Future Directions

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 01 Issue: 01
RECEIVED: OCTOBER 21 REVISED: NOVEMBER 18 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 05

Real-time data streams from patient journeys and claims processing systems hold the promise of answering important administrative and clinical questions about a population and its caregivers. Research has demonstrated that executive-level dashboards—with alerts, closed-loop feedback systems, and links to detailed data—can improve governance and inform learning. Clinical and risk-based contracts alike place significant emphasis on patient outcomes, yet reliable and timely measurement remains a challenge, especially externally imposed readmission and quality measures. Delivering reliable, trustworthy, and closed-loop analytics at the required speed demands a dedicated analytics platform and an awareness of the inevitabilities of in-flight data quality. The approaches are therefore aligned with regulatory and deployment constraints: information management at the core of the business; information products with data quality gates baked in; information provenance in the background; and responses, alerts, and remediation as part of a closed-loop system.

Successful partners have grown beyond structured claims-processing to real-time event-driven processing of incident-based submissions; thoughtful design foresight has additionally opened the door to real-time quality and strategy metrics at scale. Claims submission, validation, adjudication, and payment form an event trajectory, with recognised events and state changes that drive decision rules and scoring models. Timeliness, accuracy, completeness, and quality-control gates monitor readiness for consumption; continuity feeds detect anomalies and initiate alerting and remediation. Because these high-level measures and controls derive from data provenance established elsewhere, the resulting product is timely, actionable, and auditable. Closing the loop with alerts and remediation†responses echoes best-practice cybersecurity principles.

References

1. Cheng, A. C., Duda, S. N., Taylor, R., Delacqua, F., et al. (2021). REDCap on FHIR: Clinical Data Interoperability Services. Journal of Biomedical Informatics, 121, 103871.
2. Chowdhury, M., Gasca Cervantes, E., Chan, W.-Y., & Seitz, D. P. (2021). Use of machine learning and artificial intelligence methods in geriatric mental health research involving electronic health record or administrative claims data: A systematic review. Frontiers in Psychiatry, 12, 738466.
3. Mangalampalli, B. M. Generative AI Applications In Healthcare Data Mart Design And Optimization.
4. Lo-Ciganic, W.-H., Donohue, J. M., Hulsey, E. G., Barnes, S., Li, Y., Kuza, C. C., Yang, Q., Buchanich, J., Huang, J. L., Mair, C., Wilson, D. L., & Gellad, W. F. (2021). Integrating

human services and criminal justice data with claims data to predict risk of opioid overdose among Medicaid beneficiaries: A machine-learning approach. PLOS ONE, 16(3), e0248360.

5. MacKay, E. J., Stubna, M. D., Chivers, C., Draugelis, M. E., Hanson, W. J., Desai, N. D., & Groeneveld, P. W. (2021). Application of machine learning approaches to administrative claims data to predict clinical outcomes in medical and surgical patient populations. PLOS ONE, 16(6), e0252585.
6. Tran, L., Chi, L., Bonti, A., Abdelrazek, M., & Chen, Y.-P. P. (2021). Mortality prediction of patients with cardiovascular disease using medical claims data under artificial intelligence architectures: Validation study. JMIR Medical Informatics, 9(4), e25000.
7. Kolla, S. H. (2022). Strategic Information Integration Models for Cross-Functional Service Optimization in Large-Scale Enterprises. International Journal of Emerging Trends in Engineering and Management Research, 7(3), 11811.
8. Richter, A., Truthmann, J., Chenot, J.-F., & Schmidt, C. O. (2021). Predicting physician consultations for low back pain using claims data and population-based cohort data—An interpretable machine learning approach. International Journal of Environmental Research and Public Health, 18(22), 12013.
9. van der Galiën, O. P., Hoekstra, R. C., Gürgöze, M. T., Manintveld, O. C., van den Bunt, M. R., Veenman, C. J., & Boersma, E. (2021). Prediction of long-term hospitalisation and all-cause mortality in patients with chronic heart failure on Dutch claims data: A machine learning approach. BMC Medical Informatics and Decision Making, 21, 303.
10. Inala, R. Designing Scalable Technology Architectures for Customer Data in Group Insurance and Investment Platforms.
11. Tseng, Y.-J., et al. (2021). Machine-learning-based predictions of direct-acting antiviral therapy duration for patients with hepatitis C. International Journal of Medical Informatics, 153, 104562.
12. Vimont, A., Leleu, H., & Durand-Zaleski, I. (2022). Machine learning versus regression modelling in predicting individual healthcare costs from a representative sample of the nationwide claims database in France. The European Journal of Health Economics, 23(2), 211–223.
13. Inala, R. (2023). AI-powered investment decision support systems: Building smart data products with embedded governance controls. Journal for ReAttach Therapy and Developmental Diversities, 6(10), 2251-2266.
14. Yasar, O., Long, P., Harder, B., Marshall, H., Bhasin, S., Lee, S., Delegge, M., Roy, S., Doyle, O., Leavitt, N., & Rigg, J. (2022). Machine learning using longitudinal prescription

and medical claims for the detection of non-alcoholic steatohepatitis (NASH). BMJ Health & Care Informatics, 29(1), e100510.
15. Lo-Ciganic, W.-H., et al. (2022). Developing and validating a machine-learning algorithm to predict opioid overdose in Medicaid beneficiaries in two US states: A prognostic modelling study. The Lancet Digital Health, 4(6), e455–e465.
16. de Hond, T. A. P., Niemantsverdriet, M. S. A., van Solinge, W. W., Oosterheert, J. J., Haitjema, S., & Kaasjager, K. A. H. (2022). Sepsis labels defined by claims-based methods are ill-suited for training machine learning algorithms. Clinical Microbiology and Infection, 28(8), 1170–1171.
17. Amistapuram, K. (2023). Privacy-Preserving Machine Learning Models for Sensitive Customer Data in Insurance Systems. Educational Administration: Theory and Practice, 29(4), 5950-5958.
18. Iqbal, M. S., Abd-Alrazaq, A., & Househ, M. (2022). Artificial intelligence solutions to detect fraud in healthcare settings: A scoping review. Studies in Health Technology and Informatics, 295, 20–23.
19. Settipalli, L., & Gangadharan, G. R. (2023). WMTDBC: An unsupervised multivariate analysis model for fraud detection in health insurance claims. Expert Systems with Applications, 215, 119259.
20. Langenberger, B., Schulte, T., & Groene, O. (2023). The application of machine learning to predict high-cost patients: A performance-comparison of different models using healthcare claims data. PLOS ONE, 18(1), e0279540.
21. Xie, F., Beukelman, T., Sun, D., Yun, H., & Curtis, J. R. (2023). Identifying inpatient mortality in MarketScan claims data using machine learning. Pharmacoepidemiology and Drug Safety, 32(11), 1299–1305.
22. KollIntegration. South Eastern European Journal of Public Health, 248–260.
23. Kural, K. C., Mazo, I., Walderhaug, M., Santana-Quintero, L., Karagiannis, K., Thompson, E. E., Kelman, J. A., & Goud, R. (2023). Using machine learning to improve anaphylaxis case identification in medical claims data. JAMIA Open, 6(4), ooad090.
24. Yang, S., Varghese, P., Stephenson, E., Tu, K., & Gronsbell, J. (2023). Machine learning approaches for electronic health records phenotyping: A methodical review. Journal of the American Medical Informatics Association, 30(2), 367–381.
25. Caruana, A., Bandara, M., Musial, K., Catchpoole, D., & Kennedy, P. J. (2023). Machine learning for administrative health records: A systematic review of techniques and applications. Artificial Intelligence in Medicine, 144, 102642.

26. Inala, R. (2023). Revolutionizing Customer Master Data in Insurance Technology Platforms: An AI and MDM Architecture Perspective. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 579-606.
27. Lewis, A. E., Weiskopf, N., Abrams, Z. B., Foraker, R., Lai, A. M., Payne, P. R. O., & Gupta, A. (2023). Electronic health record data quality assessment and tools: A systematic review. Journal of the American Medical Informatics Association, 30(10), 1730–1740.
28. Ozonze, O., Scott, P. J., & Hopgood, A. A. (2023). Automating electronic health record data quality assessment. Journal of Medical Systems, 47, 23.
29. Sendak, M., Sirdeshmukh, G., Ochoa, T., Premo, H., Tang, L., Niederhoffer, K., Reed, S., Deshpande, K., Sterrett, E., Bauer, M., Snyder, L., Shariff, A., Whellan, D., Riggio, J., Gaieski, D., Corey, K., Richards, M., Gao, M., Nichols, M., Heintze, B., Knechtle, W., Ratliff, W., & Balu, S. (2022). Development and validation of ML-DQA—A machine learning data quality assurance framework for healthcare. Proceedings of the 7th Machine Learning for Healthcare Conference, 182, 741–759.
30. Yandamuri, U. S. (2023). An Intelligent Analytics Framework Combining Big Data and Machine Learning for Business Forecasting. Zenodo.
31. Lee, B.-H., et al. (2022). Blood transfusion, all-cause mortality and hospitalization period in COVID-19 patients: Machine learning analysis of national health insurance claims data. Diagnostics, 12(12), 2970.
32. Leiner, J., Pellissier, V., König, S., Hohenstein, S., Ueberham, L., Nachtigall, I., Meier-Hellmann, A., Kuhlen, R., Hindricks, G., & Bollmann, A. (2022). Machine learning-derived prediction of in-hospital mortality in patients with severe acute respiratory infection: Analysis of claims data from the German-wide Helios hospital network. Respiratory Research, 23, 264.
33. Krasowski, A., Krois, J., Kuhlmey, A., Meyer-Lueckel, H., & Schwendicke, F. (2022). Predicting mortality in the very old: A machine learning analysis on claims data. Scientific Reports, 12, 17464.
34. Yang, S., et al. (2022). Machine learning approaches for electronic health records phenotyping: A methodical review. Journal of the American Medical Informatics Association, 30(2), 367–381.
35. a, S. K., & Reddy, V. A. R. (2023). Deep Learning Architectures For Multimodal Medical Data
36. Settipalli, L., & Gangadharan, G. R. (2022). A novel fraud detection and prevention method for healthcare claim processing using machine learning and blockchain technology. Digital Applications in Archaeology and Cultural Heritage, 26, e00241.

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 01 Issue: 01
RECEIVED: OCTOBER 21 REVISED: NOVEMBER 18 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 05

37. Wu, B., et al. (2022). Improved prediction of body mass index in real-world administrative healthcare claims databases. Pharmacoepidemiology and Drug Safety, 31(8), 848–858.
38. Saito, T., Suzuki, H., & Kishi, A. (2022). Predictive modeling of mental illness onset using wearable devices and medical examination data: Machine learning approach. Frontiers in Digital Health, 4, 861808.
39. Payne, P. R. O., Gupta, A., Lewis, A. E., Weiskopf, N., Abrams, Z. B., Foraker, R., & Lai, A. M. (2023). Electronic health record data quality assessment and tools: A systematic review. Journal of the American Medical Informatics Association, 30(10), 1730–1740.