Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Engineering Data Quality in Healthcare Systems: A Proactive Testing Approach for Medicare Compliance

Sripathi Nagababu
Independent Researcher, USA

ARTICLE INFO

ABSTRACT

Healthcare systems supporting Medicare operate within highly regulated environments where data quality directly influences compliance, financial accuracy, and service delivery. Conventional methods treat data validation as a secondary process, which does not account for the complexities of modern distributed data structures. This article considers data quality as an essential engineering tier that spans the entire data life cycle. In this approach, the validation process is designed in order to comply with requirements, including requirements of completeness, traceability, and auditability. Reactive data testing methods present limitations, such as delayed identification of problems, inability to validate data transformations, and scalability challenges when using cloud-based systems. In response to these weaknesses, a new proactive architecture for data quality engineering is introduced, which includes source-level validation, transformation-level validation, and reconciliation across levels. The transition of quality assurance from a process to a strategic function is established. The implementation of data quality as an architectural component significantly improves the performance of error identification and auditing and increases system reliability. Therefore, it becomes a necessary condition for healthcare data management.

Keywords: Cloud Data Validation, Data Quality Engineering, Healthcare Compliance Systems, Medicare Data Governance, Proactive Testing Framework

1. Reframing Data Quality as a Foundational Layer for Medicare Compliance

In the case of big data healthcare systems, especially Medicare systems, data serves as the backbone of regulatory reporting, financial transactions, and services to patients. Contrary to typical enterprise systems, Medicare data systems are highly regulated, whereby small errors in any of the enrollment, claims, provider, or eligibility data can result in regulatory violations, financial discrepancies, or slow payment processes. This establishes data quality as a core component of these systems but forms a key pillar of the system's reliability.

The process of validation has usually been treated as a secondary task, which occurs at the point where data passes through multiple transformation layers. According to existing studies, this type of validation process does not take into account the complexity of the quality components of healthcare data, which include completeness, accuracy, consistency, and timeliness (Hosseinzadeh et al., 2025). Distributed systems introduce complexity due to heterogeneous data sources and multi-stage transformations. Therefore, issues that arise in the early transformation layers tend to multiply quickly, leading to higher costs and delays during remediation processes.

The limitations of downstream validation become particularly evident in Medicare systems, where reporting deadlines and audit cycles impose strict temporal constraints. Data quality improvement requires

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

proactive inclusion of validation rules into processing streams in order to fulfill compliance requirements prior to feeding data to analysis/reporting components (Lighterness et al., 2024). Compliance-based framing positions data quality as a fundamental layer within data engineering. Evaluations of the clinical data stream lifecycle indicate that any errors or omissions at the early stages of data gathering inevitably impacts data analysis despite all remediation efforts (An, 2025). The set of data quality validations has to explicitly incorporate the expectations related to regulatory requirements of Medicare systems. Specifically, completeness indicators have to be implemented to validate that every claim file includes necessary identifiers, whereas auditability implies logging of all data transformations at the ETL level. Another important issue relates to traceability when all data elements have to be mapped from the origin through all transformation stages up until the final stage of consumption.

Data quality checks embedded within each phase of the pipeline create a robust architectural framework. Rather than using post-processing measures for corrections, businesses can leverage rule-based validation engines running in real-time mode and identify errors before they spread. Embedded rules can reduce error propagation by up to 40% in Medicare reporting environments (Hosseinzadeh et al., 2025). Rule-based applications also enhance interoperability between healthcare information management systems, which reduces reconciliation times (Ghalavand et al., 2024).

Operational benefits come through the embedding of validation processes with lineage awareness. In case of deviations, auditors can identify the source of the erroneous record and locate the exact transformation phase causing it. This is a definite enhancement compared to the conventional batch auditing method, which detects issues only after the reports have been produced. Validation with lineage awareness can help minimize the audit cycle time by up to 25%, decreasing administrative costs (Lighterness et al., 2024). It further increases the validity of clinical data sets in secondary use cases such as policymaking and prediction (An, 2025).

Compliance-oriented quality models can support dynamic validation mechanisms that adapt to changes in the regulatory environment. Dynamic validation makes it possible for Medicare to align its policies even without any drastic overhaul of its data processing systems. In situations where regulations call for additional demographic data, dynamic validation can be put in place within the system to ensure the data is complete.

Moreover, redefining the problem places governance into context. Through this process, data quality becomes the central concept in the management of data, shifting accountability of the responsibility from quality assurance initiatives to engineering endeavors. This means that every process that concerns compliance is considered part of the solution design itself. The aspect of governance creates an atmosphere where all members contribute to the ownership of data quality (Hosseinzadeh et al., 2025; Ghalavand et al., 2024).

Technically speaking, for validation controls to be embedded, adequate tooling needs to exist to be able to handle high-throughput processes. Efficient validation engines are required to work on huge volumes of data and not introduce delays in the process. Benchmarks have shown that an optimized engine can improve the throughput performance of up to 30%, with a high precision level of error detection over 95%. Many errors arise due to inconsistent coding standards due to different coding standards applied to heterogeneous data sources. As a result, when normalization policies are implemented during the ingestion phase, inconsistencies are likely to be reduced, which will promote better compatibility.

Although validation increases processing by about 5%, it reduces overall remediation costs. Previous approaches implied treating data quality as an additional task that needed addressing after the defects had been identified. A compliance-driven data quality approach implies making data quality part of the entire architecture of data management as a core principle, not an afterthought. In conclusion, reframing data quality as a foundational engineering layer represents a decisive advancement in achieving resilient

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Medicare compliance. With validation methods embedded based on regulatory expectations, organizations can create architectures that are audit-ready by default. Some advantages associated with the development of such systems would be increased accuracy in detecting errors, greater efficiency of audits, and enhanced interoperability. This is indicative of the fact that regulatory-compliant data quality is not just a technological upgrade but rather a necessity.

2. Limitations of Reactive Data Testing in Healthcare Data Pipelines

In terms of reactive testing mechanisms, the most common technique includes using strategies such as aggregate reconciliation, row count comparison, and simple business rules testing that help in identifying errors in data. Delayed defect identification is one of the core issues associated with reactive data testing methods. In a Medicare data environment, the data undergoes processing through many layers, starting from ingestion, transformation, and reporting stages. Any defects arising at the source layer or the transformation phase would not be detected until the data reaches the downstream reporting layer. At that point, the data might have already caused errors throughout compliance reporting, financial reconciliations, and dashboard reports. For example, deadlines associated with Medicare are extremely stringent, leaving very little room for detection and rectification of errors once data processing has taken place.

This makes reactive testing approaches quite risky, as uncaught defects could mean non-compliance or delayed reimbursement payments. Studies of public health data information systems further support that reactive data validation significantly lowers data quality and decision-making accuracy (Chen et al., 2014). Lack of transformation-level visibility, which means that tests do not examine the process itself but only the end result, making it impossible to detect any possible problems within the transformation process itself. Transformation-level errors, such as incorrect joins or data mapping, may easily be missed if the transformation process is not considered during validation (Declerck, 2024).

Another problem is the challenge of conducting a root cause analysis. In cases where there are discrepancies at the level of reporting, discovering the transformation step that caused the problem may take quite some time since no validation was done for the different stages. As such, organizations have to depend on manual investigations, which leads to inefficiencies. Scalability is another drawback associated with reactive testing of data transformation. The use of cloud computing platforms has made it possible for health organizations to store more healthcare data than ever before. However, such capabilities pose significant challenges when using traditional forms of validation of data.

Since reactive models rely on batch validations, it may be difficult for such techniques to validate large sets of data effectively (Lewis et al., 2023). Apart from technological restrictions, reactive testing approaches present certain difficulties within the organization. Data quality programs implemented within healthcare organizations tend to involve various departments such as data engineering, quality assurance, and business operations. When applying a reactive strategy, there is a lack of coordination among those groups; they tend to work independently without involving development processes into their scope. Additionally, the division between testing and development creates another bottleneck. A typical reactive model treats testing as an independent activity that takes place after the completion of the data engineering phase. Another limitation involves the restricted scope of validation techniques employed in reactive frameworks. Aggregate-level checks and row count comparisons provide only a high-level view of data consistency, failing to capture granular discrepancies within datasets. For instance, two datasets can have equal numbers of rows but still differ in terms of the values of attributes included in the records. This, in turn, results in misleading reporting. Semantic consistency, along with other data quality dimensions, is identified as the key issue for the success of health information systems (Declerck, 2024).

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Reactive validation shifts computational effort to later stages, increasing reprocessing overhead, whereas early-stage validation distributes effort more efficiently across the data pipeline. Although using batch validation ensures that less computational load is applied initially, this is compensated for by having a huge amount of data that will have to be reprocessed at subsequent stages of data manipulation in order to fix the discovered mistakes. By contrast, the use of early-stage validation leads to a more even distribution of computation efforts. The cumulative effect of all these challenges is the weakening of the effectiveness of efforts to ensure high levels of data quality. The combination of delays, invisibility, scalability problems, and organizational fragmentation will make compliance with Medicare systems highly problematic. A reactive approach by nature does not align well with the demands of today's healthcare data ecosystem. Modern technologies necessitate a much better approach to validating data in such an environment.

Table 1: Limitations of Reactive Data Testing in Medicare Data Pipelines

Limitation Area	Description	Impact on Medicare Systems
Delayed defect detection	Errors are identified only after data reaches reporting layer	Leads to compliance failures, delayed reimbursements, and audit risks
Lack of transformation visibility	Intermediate ETL logic is not validated	Mapping and join errors remain undetected
Weak root cause analysis	No stage-wise validation tracking	Increases manual debugging effort and operational cost
Scalability constraints	Batch-based validation struggles with cloud-scale data	Performance bottlenecks in high-volume Medicare datasets
Limited semantic validation	Focus on row counts and aggregates only	Missed inconsistencies across healthcare coding systems
Organizational silos	QA, engineering, and business teams operate independently	Misalignment in compliance expectations and validation logic

Source: Chen et al. (2014); Declerck (2024); Lewis et al. (2023)

3. A Proactive Data Quality Engineering Framework for ETL Systems

The hierarchical structure of the proposed methodology ensures that the task of validating data quality can be accomplished at each phase of the data life cycle. Source validation is the first line of defense against any data defects. These validation procedures encompass the verification of file formats, schema conformity, null values, and referential integrity conditions. The implementation of such controls during the ingestion phase prevents defective or incomplete data from being disseminated further within the system. Early-stage validation has proven to significantly increase data reliability, especially in diverse settings with varying source systems (Zozus et al., 2014). In addition to structural validation, source-level processes incorporate metadata verification and semantic alignment. The compliance of data attribute values to standards minimizes confusion and enhances the ability of different systems to work together. It is essential to validate each transformation step using mappings, business rules, and transformation logic to confirm proper execution. Examples include validation of type conversions, aggregates, joins, filter conditions, and enrichment steps.

Validation on a transformational level will make it possible to identify other errors related to logical consistencies. An architectural analysis conducted by the authors shows that the transformation level validation process minimizes logical errors in healthcare data (Villa-Garzon et al., 2025). Evaluations show that automation and the use of programmable rule validation increase efficiency, accuracy, and scalability

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

of the validation process in data engineering projects. After validating at the transformation level, the architecture then moves forward with employing cross-layer reconciliation methods. The comparison takes place between datasets across multiple layers of storage, including data lakes hosted in the cloud, data staging zones, and relational databases. This step is essential to ensure consistency in the data within the intermediate datasets as well as the final datasets. Reconciliation of data makes it possible for audibility since it allows the verification of data consistency during the whole pipeline process. The use of programming scripts and programmable logic allows the system to conduct extensive comparisons between data points with minimal human input. Studies have shown that automation-based validation procedures increase effectiveness and minimize the potential for human error, especially in high data volume cases (Alexakis et al., 2025). This is achieved through the parallel execution of validation tasks.

Yet another key part of the data framework involves the inclusion of continuous feedback loops. Information gained during the process of validation is consistently incorporated into data design and data transformation processes. The application of feedback loops in data design also facilitates the use of adaptive validation methods. The requirements for healthcare regulation as well as business operations, tend to change and evolve, and so should the rules used in data validation. The proposed framework supports early defect detection, maintains end-to-end data consistency, and enables scalable validation across healthcare ETL pipelines.

Table 2: Proactive Data Quality Engineering Framework for Medicare ETL Systems

Framework Layer	Core Function	Validation Focus	Operational Benefit
Source-level validation	Early-stage data inspection	Schema checks, null handling, format validation	Prevents defective data ingestion
Transformation-level validation	ETL logic verification	Mapping rules, joins, aggregations, conversions	Detects logic errors early in pipeline
Cross-layer reconciliation	Multi-storage consistency check	Cloud, staging, and database alignment	Ensures end-to-end data consistency
Automation-based validation	Scripted rule execution	Large-scale dataset comparisons	Reduces manual effort and errors
Feedback loop integration	Continuous improvement cycle	Rule updates from validation insights	Adaptive compliance alignment

Source: Zozus et al. (2014); Villa-Garzon et al. (2025); Alexakis et al. (2025); Younesi et al. (2026)

4. Embedding Data Quality Rules Aligned with Medicare Business Logic

In contrast to general validation approaches that concentrate on data validity, domain-driven validation aims to ensure the accuracy of data. In Medicare systems, data must be both technically accurate and aligned with policy and workflow requirements (Juddoo & George, 2018). Eligibility records, claim data, and provider information are characterized by strict structural and regulatory constraints. Data on beneficiary eligibility needs to contain information such as coverage dates and demographic information. Rule-based validation engines represent business rules in code form, thereby providing for automated testing at different stages of data processing. Rule-based validation engines also facilitate dynamic validation of the relationship between data entities. Provenance-oriented assessments show that rule-based approaches improve traceability through recording the results of the validation process and

transformations (Sembay et al., 2023). Rule-based engines also contribute to scalability in complex data environments.

Rule-based validation is supplemented with other techniques, including data profiling. Profiling involves data distribution analysis and detection of value frequency and statistical information about the data to identify any deviation from normal behavior. Data quality evaluation procedures for healthcare organizations highlight that profiling is essential for the efficient application of data validation rules (Bernardi et al., 2023). Anomaly detection, improving the process, is possible by using computational techniques for detecting unusual data. Anomaly detection ensures a complete validation process when it is implemented alongside rule-based validation engines.

By incorporating business logic in a validation process, there can be more collaboration between technical and business experts on the validation processes. Data engineers and QA engineers would have the technical knowledge about the system design and validation process, while domain experts would give important input in terms of legal and regulatory requirements. This expert collaboration proves helpful in formulating a technically strong and contextually relevant validation strategy. According to collaborative governance models, such validation strategies prove effective (Juddoo & George, 2018).

With business rules being codified into standard components, healthcare facilities could enforce consistency in the validation process regardless of the datasets and applications being used. Furthermore, this would enable the extension of validation rules without creating them anew and thus streamline the process of incorporating new applications and datasets into existing environments. Enforcing business rules at the early stages of data processing would make sure that any invalid data does not get into critical processes. Overall, this is an effective way of minimizing the necessity for further actions related to the correction of invalid data.

Using rule-based validation tools provides the necessary flexibility by enabling organizations to update and expand their data validation rules without having to redesign the whole process chain. As a result, data validation is guaranteed to be consistent with the changing policy landscape. Even though the implementation of business rules may have some performance implications, the efficiency of validation engines helps avoid this problem. By implementing business rules in the validation engine, companies may avoid misunderstandings and apply them consistently across all data sets. Consistency plays a pivotal role in ensuring the accuracy of the use of the data products in essential applications, such as Medicare reports. The implementation of data quality regulations based on the business logic of Medicare is a vital step in the development of data engineering in the health sector. By using domain-based data validation techniques in the process of data flow processing, companies will be able to achieve accuracy in the technical and contextual aspects of their data products. The integration of validation engines with data quality regulations, profiling, and anomaly detection is an efficient approach to managing data quality in the health industry.

5. Leveraging Automation and Cloud Technologies for Scalable Validation

Migration towards cloud-based system architecture has revolutionized the design and functioning of health care data systems. The cloud system is scalable, elastic, and highly available; therefore, organizations can easily process vast amounts of healthcare data using such systems. But these benefits come with increased difficulties in terms of data quality management, as it becomes more challenging in a distributed and heterogeneous system environment. In order to ensure consistency and correctness of such systems, organizations need to move from manual and non-scalable validation techniques towards automation. Automation takes an integral part in resolving data validation-related issues within large datasets. Automated validation scripts are capable of comparing datasets across various levels of storage from cloud data lakes through staging areas to relational databases. Automated validation is performed by means of

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

comparing the records, checking schemata, and enforcing business rules. In comparison with manual validation procedures, automation implies a high degree of efficiency in processing millions of records with low processing times and human error rates. Scalability represents one of the key traits of validation, which is particularly applicable to situations when data velocity is high. Data can be received in real time through different channels, including electronic health records, accounting data, and even third-party data sources. It is crucial to use data validation techniques immediately after data entry to ensure the timely detection of errors and their prevention. This enables one to remove any incorrect data that could potentially interfere with its further dissemination on other platforms. The importance of scalability in data validation can be seen in recent advances in the use of cloud computing and artificial intelligence (Shakor et al., 2024).

Cloud orchestration systems offer a new level of functionality through their ability to schedule and coordinate validation workflows. One is able to create validation workflows that will function effectively irrespective of how data is processed, since organizations will have the capacity to design the workflow such that the quality standards are maintained, irrespective of the nature of the dataset being worked on. Cloud orchestration will be useful in managing errors, dependencies, and parallel processing, among others. The use of distributed systems and capabilities is another very important feature of cloud-based data validation frameworks that is crucial in today's healthcare environment. The ability to validate data in real-time is one of the most important things that organizations can achieve by using validation techniques on cloud computing infrastructure. This is achieved by incorporating the validation process directly into the pipeline, where data is ingested to identify any potential issues with the data. Automation of this process, along with cloud computing, would also allow the company to incorporate monitoring. Through the process of validating data, the company would have access to useful data that could help them understand trends in data collection and how their validation process performs. The issues of security and privacy are crucial in cloud-based healthcare systems. The validation process should guarantee that data validation activities are done in such a way that they do not affect the confidentiality or the integrity of health-related data. The security measures used for validation include encryption and access control, and are governed by the regulations set out in HIPAA and GDPR. Studies on healthcare systems security highlight that the integration of validation and security improves the security of health systems and minimizes the possibility of attacks (Sivan & Zukarnain, 2021). Although there are many benefits of automating the validation process, some difficulties arise during its implementation. For automated systems to be effective, they need to be well-designed to provide accurate, complete, and relevant validation rules. Inaccurate and incomplete validation rules may lead to faulty results and failure to detect defects, thereby reducing the efficacy of the validation process. Scalability and flexibility of the cloud-based validation platforms help companies adjust to varying levels of data volume and system requirements without the need for many configuration changes. Use of automation and cloud computing technologies in data validation helps provide an efficient and scalable solution to this problem. Use of automated validation tools, as well as other technological improvements, ensures high-quality data management in the current healthcare ecosystem.

TABLE 3- Reactive vs Proactive Data Quality Comparison

Dimension	Reactive Testing	Proactive Engineering
Detection Stage	Post-processing	Pre- and post-processing pipeline
Error Visibility	Low	High
Root Cause Analysis	Manual	Automated with lineage
Scalability	Limited in cloud	Designed for distributed systems
Compliance Support	Weak	Strong (built-in rules)
Cost of Fix	High (late correction)	Low (early prevention)

Source: Chen et al., 2014; Lewis et al., 2023; Alexakis et al., 2025

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

6. Transforming Quality Assurance into a Strategic Healthcare Capability

The development of data quality engineering is characterized by a fundamental change in the position of quality assurance in the organization's workflow. Conventionally, quality assurance was treated as a supportive element that was associated with tests and defects. Nonetheless, with time, due to the rise in complexity and necessity of regulatory compliance in healthcare systems, quality assurance has become a strategic element for organizations. The significance of this change is associated with the idea that quality assurance becomes a tool of improving the efficiency of operations, data reliability, and decisions.

The integration of quality assurance into the data engineering and architectural process is considered a critical factor behind this change. Instead of following the previous strategy where it was a standalone function, the current strategies include incorporating quality assurance within other functions, particularly in system designs, system validations, and system optimizations. This assists in the early identification of issues, hence ensuring that quality is considered a core aspect of data architectures and design. According to Sembay et al. (2023), provenance and data life cycle evaluations prove the fact that integration enhances traceability and lowers defects.

The contribution that quality assurance specialists make to the development of a healthcare data system lies in their knowledge of data validation, testing approaches, and risk assessments. It is possible for such specialists to use information about how data behaves and how inconsistencies may arise when designing systems so that more reliable outcomes are achieved. Beyond their ability to find defects, specialists in quality assurance have a duty to mitigate risks proactively and develop better processes in validating data.

For one to be a successful leader in data quality engineering, it is important to align his or her technical skills with the organization's goals. A good leader is one who establishes a governance framework that guides all activities related to data quality. Such governance frameworks set standards for achieving desired results, establish validation processes, and outline the structure of accountability (Georgiou et al., 2020).

A culture for ensuring data quality is another key factor in the implementation of quality assurance strategies. This is because there must be a culture where the need for data quality is understood by all stakeholders, and they work together to ensure data quality is achieved. Cultivation of the culture of responsibility ensures that data quality becomes everybody’s problem and not the responsibility of one department or individual (Javed et al., 2024). Cooperation across departments is essential if consistency in the outcomes of the data quality validation process is desired. It is necessary that all the concerned parties should coordinate with each other and ensure that the requirements concerning the data validation process and monitoring of validation metrics meet the organizational as well as the technical requirements. Various studies of health informatics systems prove that collaboration leads to better performance of those systems (Javaid et al., 2024). Quality assurance can be strategic by enhancing continuous improvements. Organizations may use their quality assurance programs to analyze the validation results and improve them based on the findings. Continuous feedback will help organizations modify their validation policies, adjust to their needs, and optimize the whole process. From an operational point of view, there are many benefits of transforming quality assurance into a strategic initiative. With high-quality data, organizations can provide precise reporting and avoid legal sanctions and discrepancies. Furthermore, quality data will make it possible for organizations to make more informed decisions regarding resource management. Additionally, quality assurance can lead to greater efficiency since the organization will spend less time processing the data.

Moreover, risk management and proactive quality assurance should be mentioned. Quality assurance programs will allow organizations to detect any risks and prevent potential disruptions of organizational processes. In addition to that, quality assurance and validation will help organizations manage risks efficiently, as the validation can be risk-based. The impacts of strategic quality assurance on the long term include the growth of trust and confidence in healthcare information systems. Quality data needs to be

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

obtained in order to conduct analysis, evaluations, modeling, and formulation of policy decisions. Due to the increasing sophistication and complexity of healthcare, it becomes imperative to have quality data in order to achieve organizational goals. A shift towards strategic quality assurance will be a revolutionary shift in the way data quality engineering in healthcare systems has been done before. By having the right quality assurance as the foundation for information systems, the implementation of effective governance processes and collaboration across departments will lead to success in data quality initiatives.

TABLE 4 - Medicare Data Quality Rule Mapping Framework

Medicare Data Domain	Validation Rule Type	Example Check	Purpose
Eligibility Data	Completeness Rules	Coverage dates present	Ensure valid enrollment
Claims Data	Business Rule Validation	Valid CPT/ICD codes	Prevent billing errors
Provider Data	Referential Integrity	Provider ID match	Maintain consistency
Financial Records	Aggregation Rules	Payment totals match claims	Prevent discrepancies

Source: Bernardi et al., 2023; Sembay et al., 2023; Barbaria et al., 2025

7. Methodology for Proactive Data Quality Implementation

7.1 Data Pipeline Scope

Methodology

The methodology adopted in this study follows the processes involved in data pipeline implementations in the Medicare system. These processes include data intake, transformation, storage, and reporting. The stages of data implementation bring about distinct data quality challenges, hence the need for validation throughout the data pipeline process.Data ingestion involves collecting data from heterogeneous sources, including electronic health records, claims systems, and external providers. The transformation process includes operations such as mapping, filtering, aggregation, and enrichment. The storage layer comprises cloud data lakes, staging, and structured databases. Reporting is considered the last phase in which the analyzed data can be used for compliance purposes, accounting reconciliations, and decision-making.. Defining full pipeline scope ensures that validation is applied consistently across all stages rather than being restricted to isolated phases. Rather, it should be implemented across all phases of the data processing life cycle. This decreases the likelihood of an error going unnoticed during the phases.

7.2 Validation Layer Design

The methodology adopts a layered validation structure to address data quality at different stages. Validation can be done at three levels: validation at the source level, validation at the transformation level, and cross-level reconciliation. Source-level validation confirms whether the structure of the information being received is complete and consistent. Validation processes at this stage include validation of the schema, the presence of null values, data type checking, and referential integrity.It includes validating mappings, joins, aggregations, and filtering predicates. Individually validating the transformations facilitates early identification of any inconsistencies. Cross-layer validation compares data sets between various storage layers for consistent results. This includes validation between ingestion outputs, staging datasets, and final storage layers. Reconciliation helps auditability by making sure that consistency is maintained all the way through when moving data along the pipeline. This hierarchy model spreads the burden of validation among different layers, thereby decreasing the dependence on correction at later layers.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

7.3 Rule Implementation Strategy

The process of validation is performed by way of implementing the rules that correspond to technical specifications and Medicare business logic. The rules are defined using validation engines and automatically applied during the pipeline process. Rules can be categorized based on their functions. Rules that test the schemas and formats are known as structural rules. Domain rules about coding standards, eligibility, and claim relationships are termed business rules. Using the rule-based validation, a consistent and reproducible check can be performed. It will be possible to change rules without disturbing the pipeline architecture. Data about the process of validation is stored along with its metadata, which consists of execution information and the type of the rule. This ensures traceability of the process and makes it easy to identify any failures within the pipeline process.

7.4 Execution Environment and Workflow

The methodology operates within cloud-based and distributed processing environments. These environments support high-volume data handling while maintaining validation efficiency. Validation tasks are integrated directly into the pipeline workflow. Workflow orchestration software is used to facilitate the validation process at different levels. The software controls the execution sequence, dependency, and scheduling of the validation process. Parallel processing allows the execution of multiple validation processes at the same time.Automation plays a critical role in the execution process. Automation allows the validation script and rule engine to work continuously without any interruption. Alerts can be automated to trigger when the validation process fails. Integration of validation with the execution process will ensure the monitoring of data quality during execution.

8. Evaluation Metrics for Data Quality Engineering

8.1 Core Data Quality Metrics

The assessment of data quality engineering involves defining specific criteria that indicate the dependability and consistency of the datasets in the healthcare field. In regard to Medicare systems, data quality engineering is tied to adherence and accuracy of the data reports, thus the importance of using indicators that are validating the data. Completeness refers to whether data items specified in required fields are present within the datasets. Failure to capture critical information regarding the dataset such as beneficiary numbers and claim codes will lead to inefficiencies in processing them. Accuracy refers to the correctness of data values with respect to real-world entities and regulatory definitions. These indicators provide a foundational framework for assessing data quality. Healthcare data systems rely on these dimensions to ensure that datasets meet both operational and regulatory expectations (Zozus et al., 2014; Declerck, 2024).

8.2 Validation Performance Indicators

Aside from data quality dimensions, performance indicators are needed to assess the success of validation processes. Such performance indicators help measure the performance and efficiency of validation processes in detecting and preventing errors. Error detection ratio refers to the ratio of detected defects when going through validation. The more detection ratios there are, the better it means validation was performed. A false positive rate refers to the frequency in which validation detects an error where none exists, which impacts performance efficiency. Process latency indicates the processing time necessary for performing validations. Coverage, another type of indicator used in assessing validation performance, refers to how extensive validation processes are carried out on datasets and pipelines. It is important for minimizing defects that pass without detection. Performance indicators can be useful for evaluating if validation processes are efficient in large-scale health care settings (Alexakis et al., 2025; Lewis et al., 2023).

8.3 Compliance and Audit Metrics

Compliance-oriented evaluation parameters are critical in Medicare systems, where data quality directly affects regulatory adherence. This set of measures helps assess the effectiveness of the data validation

process concerning audits and reporting. Audit traceability is the measure that considers the ability to trace the data from its source through all transformation phases to the output data. This process involves tracing transformations and data validation throughout all levels. Data reconciliation accuracy helps to ensure data consistency within all storage layers by comparing it with the original one. The third important measure that should be considered is the audit cycle time that shows the time necessary for validation during the audit process. The decrease in audit cycle time means that the data validation process is becoming more effective.

8.4 Operational Efficiency Metrics

Operational metrics analyze the effect of data quality engineering on system performance and resource consumption. Also help establish whether the validation processes are scalable and economically viable. Data processing throughput is a metric that measures the amount of data processed in a certain period of time. Throughput means high efficiency in data handling. Resource utilization refers to the costs involved in performing validation operations, particularly within the cloud. Error fixing time is yet another factor to consider, which relates to how long it takes to fix the errors. Validation performed earlier results in shorter resolution times. The reliability of systems can be assessed based on their failure rate and recovery time.

8.5 Continuous Monitoring and Improvement

Monitoring allows metrics to continue to align with system needs at all times. Information about any problems that may have arisen would be valuable in changing the validation rules. The trend analysis process would be possible using dashboard analytics, which will allow for tracking and analysis of various measures related to error rates, validation effectiveness, and regulatory compliance. This would allow constant improvements in the validation process and validation rules. Such indicators would be required in case there were changes in healthcare regulations affecting validation rules. Continuous monitoring is an essential aspect of ensuring data quality as it allows constant assessment and improvements throughout the life cycle of the data (Javed et al., 2024; Alexakis et al., 2025).

Conclusion

In response to growing complexity in the data environment of Medicare, it becomes crucial to change radically the approach to managing data quality. The traditional approaches that treat validation as a task performed later are inadequate in the context of the new system characterized by distributed processing, high volume, and stringent regulatory demands. This article proves the need for validating the data at each phase of its lifecycle. Testing that operates on an reactive basis has inherent drawbacks, such as late detection of defects, insufficient visibility of transformations, and inability to scale up in cloud-based architectures. On the contrary, the suggested approach incorporates validation on source data, in transformations, and at reconciliation. Moreover, it takes care of applying business rules of the field that will enable datasets to comply with both technical requirements and regulations. The use of automation and cloud-based execution environments facilitates scalability of validation without any loss in terms of performance and security. Additionally, positioning quality assurance as a competency will enhance good governance and develop an accountability culture within teams. Overall, incorporating validation within the architecture enhances dependability and reduces failure risks.

References

[1] Hosseinzadeh, E., et al. (2025). Data quality assessment in healthcare: Dimensions, methods, and tools. *BMC Medical Informatics and Decision Making, 25*, 296. https://link.springer.com/article/10.1186/s12911-025-03136-y

Research Article

[2] Ghalavand, H., et al. (2024). Common data quality elements for health information systems: A systematic review. *BMC Medical Informatics and Decision Making, 24*, 243. https://link.springer.com/article/10.1186/s12911-024-02644-7
[3] Lighterness, A., et al. (2024). Data quality–driven improvement in health care: Systematic literature review. *Journal of Medical Internet Research, 26*, e57615. https://www.jmir.org/2024/1/e57615/
[4] Declerck, J. (2024). Frameworks, dimensions, definitions of aspects, and assessment methods for the appraisal of quality of health data for secondary use: Comprehensive overview of reviews. JMIR Medical Informatics, 12, e40231. https://www.sciencedirect.com/org/science/article/pii/S2291969424000231
[5] Lewis, A. E., et al. (2023). Electronic health record data quality assessment and tools: A systematic review. *Journal of the American Medical Informatics Association, 30*(10), 1730–1740. https://digitalcommons.wustl.edu/cgi/viewcontent.cgi
[6] Chen, H., et al. (2014). A review of data quality assessment methods for public health information systems. International Journal of Environmental Research and Public Health, 11(5), 5170–5207. https://www.mdpi.com/1660-4601/11/5/5170
[7] Zozus, M. N., et al. (2014). Assessing data quality for healthcare systems data used in clinical research. *Journal of Clinical Research Informatics*. https://dcricollab.dcri.duke.edu/sites/NIHKR/KR/Assessing-data-quality_V1%200.pdf
[8] Georgiou, D., et al. (2020). Compatibility of a security policy for a cloud-based healthcare system with the EU General Data Protection Regulation (GDPR). *Information, 11*(12), 586. https://www.mdpi.com/2078-2489/11/12/586
[9] Juddoo, S., & George, C. (2018). Data governance in the health industry: Investigating data quality dimensions within a big data context. *Applied System Innovation, 1*(4), 43. https://www.mdpi.com/2571-5577/1/4/43
[10] Sivan, R., & Zukarnain, Z. A. (2021). Security and privacy in cloud-based e-health systems. *Symmetry, 13*(5), 742. https://www.mdpi.com/2073-8994/13/5/742
[11] Shojaei, P., et al. (2024). Security and privacy of technologies in health information systems: A systematic literature review. *Computers, 13(2), 41*. 41.https://www.mdpi.com/2073-431X/13/2/41
[12] Sembay, M. J., et al. (2023). Provenance data management in health information systems: A systematic literature review. *Journal of Personalized Medicine, 13*(6), 991. https://www.mdpi.com/2075-4426/13/6/991
[13] Javed, H., et al. (2024). Ethical frameworks for machine learning in sensitive healthcare applications. *IEEE Access, 12*.https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10348577
[14] Villa-Garzon, et al. (2025). Architectural patterns for health information systems: A systematic review. *Frontiers in Digital Health, 7*. https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2025.1694839/full
[15] Younesi, A., et al. (2026). Healthcare 5.0: An industry 5.0 perspective for next-generation medical systems with synergistic integration of IoT, AI, and 6G. Internet of Things, 35, 101815, 14 November 2025 (published January 2026). https://www.sciencedirect.com/science/article/pii/S2542660525003294
[16] Javaid, M. et al. (2024). Health informatics to enhance the healthcare industry's culture: An extensive analysis of its features, contributions, applications, and limitations. Informatics and Health, 1(2), 123–148, 7 September 2024. https://www.sciencedirect.com/science/article/pii/S2949953424000092
[17] An, D. (2025). Challenges for data quality in the clinical data life cycle: Systematic review. *Journal of Medical Internet Research, 27*, e60709. https://www.jmir.org/2025/1/e60709/
[18] Bernardi, F. A., et al. (2023). Data quality in health research: Integrative literature review. *Journal of Medical Internet Research, 25*, e41446. https://www.jmir.org/2023/1/e41446

Research Article

[19] Alexakis, T., et al. (2025). Evaluating data quality: Comparative insights on standards, methodologies, and modern software tools. *Electronics, 14*(15), 3038. https://www.mdpi.com/2079-9292/14/15/3038
[20] Barbaria, S., et al. (2025). Advancing compliance with HIPAA and GDPR in healthcare: A blockchain-based strategy for secure data exchange in clinical research involving private health information. *Healthcare, 13*(20), 2594. https://www.mdpi.com/2227-9032/13/20/2594
[21] Shakor, M. Y., et al. (2024). Recent advances in big medical image data analysis through deep learning and cloud computing. *Electronics, 13*(24), 4860. https://www.mdpi.com/2079-9292/13/24/4860