Health Data Quality:

What It Is and

How To Use It

Introduction

Quality healthcare is impossible without providers, payers, labs and other stakeholders being able to share quality patient data. Unfortunately, too many healthcare organizations struggle with poor data quality in the form of duplicate patient records and overlays (in which a patient's medical information is placed in another patient's file) that can lead to clinical decisions that threaten patient safety.

Further, legacy health IT infrastructures can act as barriers to quality care. Not only do legacy systems make it difficult to securely share patient data, they make it hard for healthcare organizations to leverage innovations in patient care and technology that could produce better outcomes while helping to control healthcare costs.

To provide optimal modern healthcare to patients, we need to evolve where and why we access data. More importantly, healthcare organizations must promote and utilize quality health data. This eBook represents the original health data quality (HDQ) playbook for healthcare organizations. It will include examples of good and bad HDQ and the entities most impacted—empowering labs, hospitals, imaging centers, payers, HIE/HINs, and other key healthcare stakeholders to take meaningful action towards drastically improving their HDQ.

The Definition of Data Quality in Healthcare

What do we mean when we talk about health data quality? Health data itself encompasses both clinical data – such as immunizations, vitals, test results, and lab reports – and personal information such as names and addresses. Health data quality refers to the accuracy, consistency, relevance, and ultimate usefulness of those records. Perhaps most importantly, ideal HDQ calls for each patient to have a singular, unique record with patient data points aggregated as one comprehensive, longitudinal record across time, systems, and organizations.

Health data comes from multiple inputs across multiple interfaces. This data is subject to human error and discrepancies between systems that can result in duplicate records and inconsistent data. Crucial information may be incorrectly recorded or altogether absent. For healthcare providers to make appropriate care decisions that optimize patient safety, and for organizations to gain accurate clinical insights, it is imperative that health data be

unique, clean, and consistent—only then can data lend itself to accurate audit, useful presentation, and insightful integrated data analytics. This singularity is encapsulated in the concept of a master patient index (MPI).

HDQ encompasses the information contained within electronic medical records (EMRs) and personal health records (PHRs), but also extends to how those records are managed, how the databases which contain them are structured, and the ability of AI and ML tools to glean actionable analytical insights from the data sets.

The 5 Pillars of HDQ

HDQ includes multiple elements that can be viewed as pillars:
- Database structure and record management, such as EMRs and PHRs
- Data aggregation and augmentation
- Auditability
- Interoperability
- Analytics and ML-driven insights

Below are examples of good and bad approaches and basic success metrics for each of these pillars:

PILLAR	HIGH HDQ	LACK OF HDQ	SUCCESS METRICS
Database structure and record management	Records are unique, consistent, accurate, current and complete	Duplicate records, contradictory, incomplete, or misleading data	Number of duplicates or incorrect records
Data aggregation and augmentation	Data is collected, processed, summarized, and presented in a useful format	Data is dirty, disparate, unnormalized, and not accessible in an actionable format	Improved clinical insights, care coordination, and early detection of data quality issues
Auditability	Changes to a patient's record can be identified and traced	Changes cannot be identified and/or traced	Number of instances where data changes are unrecognizable and/or untraceable

Interoperability	Fluency between systems without human intervention	Systems cannot exchange information without error or data loss	Number of error/loss instances between systems
Analytics and ML-driven insights	Clean, high-quality data to yield accurate and actionable insights	Dirty or corrupt data that produces inaccurate or misleading insights	Number of instances where trends or insights do not reflect reality

The Real Cost of Poor Data Quality Management

HDQ for patient records across systems and organizations is essential for patient safety and for providing value-based care. While dirty data is a problem in any industry, the stakes in healthcare are higher and the costs can be immediately tangible. Inaccurate or incomplete health data creates an inherent patient safety issue which can result in suboptimal care.

But the cost of incomplete, missing, misleading, duplicated, or otherwise low-quality health data is far-reaching: Dirty data can impede the research and development process, compromise analytics and decision-making, and hamper a healthcare organization's effectiveness and viability. Poor HDQ impairs a healthcare organization's ability to:

- Access clinical data electronically
- Correctly identify the proper patient
- Make timely and proper care decisions
- Update coding principles to correlate data

HDQ has real consequences for patients, providers, and health systems, and a cumulative impact on healthcare overall.

The Cost of Data Duplication and Overlays

An overlay is a dangerous occurrence in which data from one patient is mistakenly added to another patient's chart, or where the records of different patients are mistakenly merged. The potential ill effects of overlays are obvious: They put physicians in the unknowing

position of making care decisions for one patient based on the data of an entirely different person.

Both duplication and overlays are primary causes of poor HDQ that can lead to erroneous diagnoses, inaccurate dosing, and unnecessary procedures or medication, any of which could create dangerous (or even fatal) consequences for patients. As a leading cause of medical errors, incorrect patient chart information is responsible for massive amounts of litigation and incalculable patient harm.

The risk of misidentification due to duplicate records also can result in improper or harmful care decisions. Healthcare staff may update the wrong record or may be uncertain as to which record to update. This results in significant data going unrecorded, or in outdated or misleading information infiltrating the patient record. Duplication is also a major reason for payment denial, causing billing and collection inefficiencies and resulting in revenue loss.

State of Laboratory Data Quality

When it comes to data quality the laboratory market lags behind other healthcare sectors, particularly regarding duplicate records. As much as 30% of laboratory patient data is accidentally duplicated. In addition to the immediate consequences for the affected patients and providers, poor laboratory HDQ can impact public health as well. Inaccurate lab data gives public health officials a false impression about the spread or containment of disease and the prevalence of health conditions.

The Cost to Diagnosis and Care Management

Healthcare organizations must be able to confidently assess treatment efficacy and analyze value. Duplicate, inaccurate, or contradictory data compromises this process. The confusion created by incomplete, missing, or inconsistent data impedes quality care by delaying test results, which in turn can delay crucial treatments, potentially imperiling patient safety and increasing the risk of poorer health outcomes.

For example, suppose a female patient presents at the emergency room with a possible miscarriage or ectopic pregnancy—each a potentially life-threatening condition and which

require totally different treatment plans. Absent a complete, unique, and accurate patient record, physicians would not have access to previous HCG test results, ultrasounds, etc. to inform their care management. The ability to track changes empowers clinicians to deliver better care, improves patient outcomes, and decreases the opportunity for unnecessary risk and avoidable costs to arise.

The Cost to Interoperability and Data Transparency

Stakeholder burden increases where HDQ is lacking. When providers and payers don't have the ability to access and share quality data, interoperability and data transparency is compromised, decreasing an organization's ability to improve patient outcomes, manage costs, and deliver value-based care.

The vast amounts of clinical and claims data lost to siloed systems can seriously undermine HDQ. Consider a patient who crosses the organizational borders of hospital systems, labs, providers, as many patients do. With no unique identifier firmly linking the patient with their information, portions of that patient's data will likely be scattered along their care journey. This poor data quality management often leads to uninformed clinical decisions that can endanger the patient, as well as unnecessary tests at unnecessary cost.

Resolving the data transparency challenge would mean the ability to aggregate patient data from multiple sources. Having all of a patient's data in one app should be the gold standard for data transparency and help ensure data quality.

Effects on Technological Investment

The sophisticated capabilities of artificial intelligence and machine learning can offer analysis, prediction, and convenience that enables providers to deliver better and more efficient care. But these tools are not inherently error-proof, and analyses generated from dirty data are unreliable at best and can be entirely misleading. The data produced is only as good as the data input, e.g. "garbage in, garbage out."
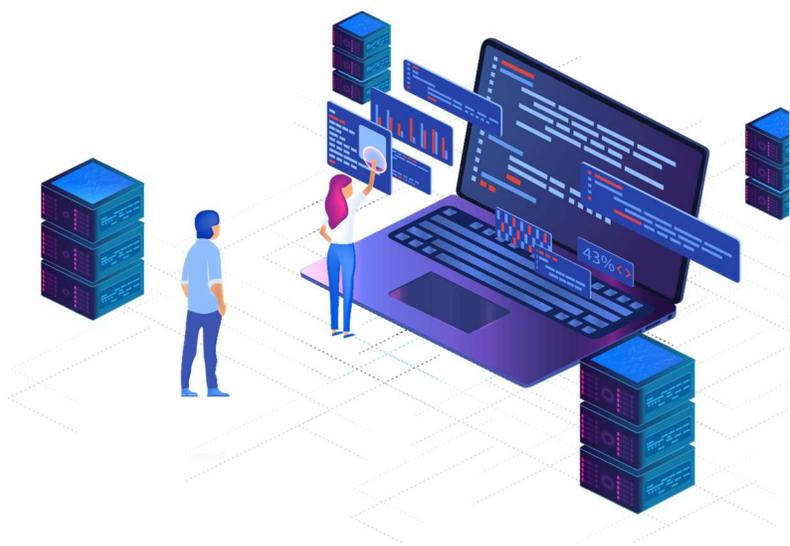
However, while poor HDQ hampers the efficacy and accuracy of these technologies, the power and potential of AI and ML is drastically improved by access to high-quality data. Ensuring clean, unique, and accurate records at the most basic level of data input increases the usefulness of more advanced data technology tools.

Benefits of improving HDQ

Clean data means better decisions, better patient care, and a more effective healthcare system. The quality of data being used directly correlates with the probability of success across all metrics, in addition to heightened organizational efficiency, patient/provider communication, and improved clinical decision-making. As hospitals, health systems, and payers increasingly rely on data analysis to make appropriate care decisions, reduce costs, and improve efficiency; high HDQ is essential to achieving these goals.

Improved Population Management and Reimbursement

Imaging centers and labs must have clean identity data sets to extract value from their data and participate in population management. When data is misleading or outright inaccurate, population management is reduced to a guessing game.

Suppose researchers were seeking to determine the percentage of women within a certain geographic cohort who have had abnormal PAP smear results. If the laboratory data is poorly managed and indexed, multiple abnormal results from multiple tests on a single patient will be treated as coming from multiple patients, thus corrupting population data.

Employing a master patient index (MPI) can help avoid and resolve data quality issues that lead to duplicate records and corrupted population health data. As health systems merge and expand to new locations, estimates now place the duplication rate at around 20% and even as high as 30%. By creating a unique and longitudinal record for each patient, an MPI approach improves the accuracy and reliability of larger data sets, even as it also improves individual care.

Clean identity resolution is particularly crucial for payers as they invest in systems to analyze population risks. Labs with a high degree of confidence in their healthcare data quality management can leverage those clean datasets to negotiate better reimbursement in payer contracts.

Improved ROI and patient/payer collections

Imagine how a poorly managed integrated data analytics system would handle a patient who comes in for a regular test four times per year. A low-HDQ database will treat that single patient as four individual patients, creating a duplicate record at every visit. According to a 2018 study, one Texas hospital with a 22% patient record duplication rate was losing $96 per duplicate. Considering the exponential increase in time and expense of the billing process alone, it's clear how quickly these unnecessary costs can accumulate.

From the perspective of patients and insurance payers, receiving duplicate bills creates confusion and uncertainty. This may lead to a lack of confidence in the legitimacy of amounts billed and can result in payment delays or claims denials. In fact, claims denials by insurers can cost a lab up to 30% of its billings. These losses can be mitigated by using an MPI to ensure high levels of data quality.

Making billing data available to patients via an app helps avoid confusion and facilitates the billing process, improving both patient experience and collections success. Because lack of HDQ directly contributes to inefficient billing and collections, healthcare organizations that address their HDQ insufficiencies see improved ROI in these areas.

Improved patient outcome and organizational efficiency

Implementing modernized systems and automated processes not only can increase HDQ, it can improve research capabilities, reduce costs, improve organizational efficiency, and lead to better patient communications – all of which can improve patient outcomes overall.

How organizations can remove poor data and make their data healthy

So how can organizations remove poor data and improve their HDQ? While realizing and analyzing the problem is paramount, it's important to remember the value of high HDQ isn't limited to the data alone—its value also lies in the needs quality data can help address and how that data can be used.

Three steps to quality health data

There are three key steps required for improving HDQ:

- **Ensure electronic access** to necessary data
- Use **identity management** to identify the patient quickly and correctly
- **Normalize the data to make it** useful and actionable

First Step

More than 50% of clinically relevant data comes from sources not structured within electronic health records (EHRs), such as physician notes, imaging, and lab reports. In addition to complicating the ability to get a complete picture of a patient's record, and therefore affecting treatment decisions, this lack of access impedes identifying actionable genomic biomarkers or matching a suitable patient with a clinical trial.

Although interoperability and data sharing have improved in recent years, truly improving data quality may mean progressing beyond legacy systems. Storing medical records in the cloud rather than within an organization's servers can enable easy and timely access by authorized users.

Second Step

Since the value of health data lies in its ability to document longitudinal changes in a patient's record, identity management is the critical next step. Without proper identity management it's difficult to connect disparate data sources, much less to associate the data correctly with a patient. Effective identity management brings enormous clinical and efficiency benefits—for example, allowing clinicians to observe the effect of treatments or disease processes on a certain patient's lab results.

Third Step

Once the right data is attributed to the right patient, and a longitudinal record is created, the third phase of HDQ improvement is to make data actionable. Health data comes from varied sources, many of which do not necessarily use the same terminology for a certain test or the same coding for a procedure. Even when records are matched to the correct patient, data must still be correlated from different clinicians, hospitals, labs, and imaging centers.

The process of data normalization creates a common terminology to resolve differences in coding, terms, and even languages, making records readable and their data useful and actionable. Providing a historic dataset for a patient allows doctors to make better care decisions.

Potential barriers, even when an organization attempts to improve data quality

Normalizing the data requires additional resources, and the speed of incoming data makes it hard to keep up, much less address core data issues. With multiple data sets streaming in, often via paper requisitions, accurate manual entry becomes unfeasible.

Another potential barrier is the lack of shared data collection standards between healthcare organizations. Where records are not consistently updated and reconciled, it can become difficult to verify whether data is current and accurate, creating a real risk of overlays or mixing the data of two different patients.

Conclusion

Quality healthcare and patient safety are dependent upon the ability of healthcare organizations to share accurate and comprehensive patient data. By making health data quality a priority, healthcare organizations can realize the promise of value-based care while keeping patients safe from errors related to duplicate records, overlays, and other poor-quality data.

How 4medica can help

With the pressures on today's health organizations—the need for exchanging and aggregating data, the need for more successful billing and collecting, the need to ensure patients are never compromised even as patient records are being exchanged—surfacing high-quality and actionable data is essential to organizational progress.

At 4medica, we're ready to help ensure you can actually deliver on the value of success for revenue, for patient safety, and for the performance of your data—driving interoperability and profitability while reducing operational losses. Reliable HDQ empowers efficient operation and quality care while providing the visibility needed to improve outcomes, strategize new initiatives, and participate in secondary data usage.

With the industry's most advanced MPI process, 4medica has revolutionized HDQ achievement: our 4-layer approach guarantees a dramatic <1% duplication rate. Our ongoing data stewardship and product design moves organizations toward full independence with easily configurable matching rules that enable self-monitoring of system health.

In fact, we recently helped the Idaho Health Data Exchange reduce their record duplication rate below 1%. Learn how 4medica resolve over 5 million messages and delivered a reliable and longitudinal health record for optimal care management, all in 90 days from start to go-live.

Ready to improve your organization's health data quality?
Contact 4medic to start today.