What are the ways in which healthcare data is misinterpreted? — and how do you read it for yourself?

There's no shortage of data in healthcare these days; dashboards, reports, summaries, charts, KPIs (key performance indicators) are part and parcel of almost every decision made by hospitals, start-ups, insurance companies and public health departments. The issue with this abundance of information, however, is not that there is a dearth of data, but that the data is often misinterpreted.

Misinterpretation does not typically occur due to carelessness or incompetence. Instead, it occurs when there is an absence of context within the healthcare data itself. While many teams have clinicians who are skilled at working with data, and many have individuals with clinical experience who also work with data, it is still relatively rare for cross-functional or non-clinical teams to understand the nuances of healthcare data.

Misinformation Hiding Behind a Veil

Most of the times, when we look at data, it appears to be precise; therefore, people trust the data to provide answers. However, the data in healthcare is rarely precise. Metrics used to measure healthcare performance are typically built upon assumptions and approximations. Many times the data being used comes from incomplete records, outdated data, or indirect measures. For example, a decline in documented cases or visits may not necessarily indicate an improvement in the world outside of the hospital. It could mean something entirely different such as a change in documentation requirements, or a staffing shortage, or simply a delay in reporting.

To begin reading data safely, one needs to resist the desire to view numbers as absolute truth. More importantly, data in healthcare should be viewed as signals — not conclusions.

When Definitions Shift Invisibly

Another way in which data can be misinterpreted is when definitions are changed. Data in healthcare is highly dependent on the definition of the terms, codes, and groupings of the data. When definitions are changed, even slightly, the interpretation of the data will change along with them.

Examples of changing definitions include "readmission rates" or "adverse events." Although these types of data appear to remain constant, the criteria under which the data is collected can change from organization to organization and/or over time. Changes to documentation standards or coding practices can create an appearance of improved performance or reduced performance, without any change to actual patient outcomes.

In the end, as a safe reader of healthcare data, one thing always needs to be asked: I can see something being counted here. What exactly is being counted, and are the definitions same throughout the process?

Aggregating Hides Important Variation

The numbers in healthcare are very often grouped into categories (aggregated) so that they are easy to comprehend; however, the process of aggregating is useful, but it also obscures variability. By averaging numbers across departments, regions, and/or populations, the underlying variability becomes less apparent. An individual number that represents an aggregate can seem perfectly fine on its own. There will undoubtedly be instances where there is a large outlier within a group that an overall average fails to reveal. Unless data is broken down in a thoughtful manner, decision-makers may fail to see where the true problems reside. Therefore, safe interpretation of data includes exercising caution when using averages and always consider whether segmentation or aggregation will alter the story.

Correlation Looks Like Causation

This is a well-known data issue, but it can pose significant risks in healthcare. When two trends occur at the same time, many people believe that one trend is responsible for the other trend. In healthcare, making such an assumption can be misleading. A trend toward a higher frequency of a specific event may correlate with a recent initiative, policy or intervention, but correlation does not imply causation. Many external variables (e.g., seasonal trends, demographic shifts, access issues) likely influence the relationship between the two trends. Using restraint with your intellect, you should ask yourself "what possible explanations are available?" when reviewing data, and "what evidence do I actually have?"

Context is Not Displayed by Dashboards

Dashboards are used primarily to provide a summary of data, not to provide explanation about the data. Dashboards typically do not include the operational, clinical, and/or organizational context that influenced the data collected. For example, workforce shortages, changes to workflows, or documentation backlogs may significantly impact the metrics that are reported. Numbers can appear either alarming or reassuring without the proper context. Safe data interpretation requires awareness of what a dashboard cannot display. Engaging in conversations with subject-matter experts, examining documentation processes, and evaluating system constraints may be as important as the numbers displayed in the dashboard itself.

Small Errors Add Up Rapidly

Healthcare systems are made up of many different pieces, and small errors when collecting data can grow into bigger problems when you look at them all together. Small errors like one misspelled code or one misplaced timestamp or duplicate record might appear insignificant on their own but they can greatly impact the overall trend and summary of a group of thousands of other similar records.

It's especially important to recognize this pattern in analytic environments that move data from multiple systems before visualizing it as an error may have started somewhere along the way by the time it shows up on your dashboard.

A healthy reader will always view data outputs with some level of skepticism rather than blindly trusting them; if the output seems unusual then the first step should be to question how the data was moved along its path, not because reality just changed.

Better Ways to Interpret Healthcare Data

Reading healthcare data safely doesn't require clinical authority or advanced statistical training — it just takes discipline and humility.

Here is a simple, practical approach to safe interpretation of healthcare data:

1) Determine what the data represents and what it does not represent.
2) Understand how the data was collected and processed.
3) Be cautious when using averages and summary metrics (especially if you don't know the distribution).
4) Avoid making causal claims without strong evidence.
5) Look for context beyond the dashboard.

Most importantly, safe interpretation means you are comfortable saying "we need more context before we draw conclusions." In healthcare, that pause can prevent costly mistakes.

Imagining yourself as a detective to solve a case is very close to the mindset that a person handling data should have.

Why This Matters

Healthcare decisions often affect real people, real resources, and real systems. Even if data is used only for internal planning or business strategy, misinterpretation can lead to misplaced priorities, efficient work flows, or unintended consequences. Clear, responsible data interpretation helps teams make better decisions without overstepping into clinical judgment; it also creates trust between technical, clinical and non-clinical stakeholders by keeping conversations grounded in what the data can actually support.

Illustrative Example:

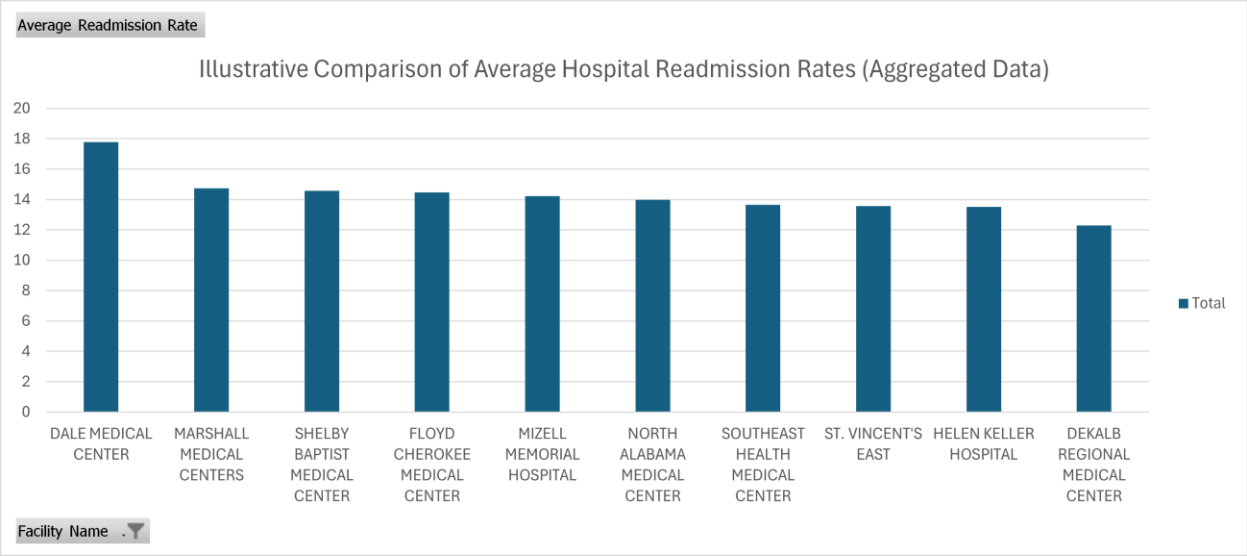

Figure: Illustrative comparison of aggregated hospital readmission rates. For illustration only; to raise awareness of potential pitfalls when interpreting health care data.

An illustrative example was created to demonstrate how slight numerical differences can create the appearance of significance when comparing aggregate hospital level readmission data. The example is intended to emphasize how easily health care data can be misunderstood when the definitions of data being interpreted, method(s) of aggregating the data or reporting the data are not fully understood.

Open Data Source (de-identified)

This document refers to publicly accessible, de-identified health care performance data from the Centers for Medicare and Medicaid Services (CMS) specifically relating to hospital level readmission data published as a result of the Hospital Readmissions Reduction Program (HRRP). Only aggregate data regarding hospital-level measures were provided in the dataset and no patient identifiable information was included in the dataset. As an example of the types of common risks associated with interpreting health care analytic data this data will be referenced here.

Limitations:

This document represents a conceptual framework and is based on high-level principles and publicly available, aggregate data. The document does not address organizational specific workflows, documentation methods, population differences or operational limitations which may affect an individual's data set. The examples contained within this document are for illustrative purposes only and are intended to represent common risk factors related to the interpretation of health care data rather than evaluating outcomes, quality of care or an organizations performance.

Disclaimer:

This work is intended for informational and educational purposes only. It addresses general principles of health care data interpretation and analytics. It does not contain clinical advice, diagnostic direction, treatment recommendations or regulatory positions and is not intended for use in providing patient care or making clinical decisions.