Data Warehousing: Myths, Pitfalls, and the Secret Weapon

Save to myBoK

by Rich Moyer

Is your organization using its data warehouse to its fullest potential? The author provides some pitfalls to avoid and some tips that may help improve care and use resources more efficiently.

Most healthcare organizations have invested in building a data warehouse. But once the warehouse is built, not all organizations take the necessary steps to ensure they are using the data to its best benefit. In many cases, they have only marginally improved their ability to analyze data and take actions based on it—so they may have missed opportunities to improve care and use resources more efficiently.

This article will discuss common pitfalls to avoid in managing a data warehouse over time and explore how HIM professionals can assist in these processes.

Three Problem Areas

Difficulties in managing data warehouses are not unique to the healthcare industry. Some studies report up to 70 percent of data warehouse projects didn't provide the expected benefits.[1]

In recent decades, issues that affect the success of data warehousing have stayed remarkably constant. These issues can be grouped into three categories:

- Data quality and completeness are inadequate for decision making
- Organizations lack the skills and expertise to make data driven decisions
- Data warehouses are not developed to perform specific analyses

Each of these three categories affects and magnifies each of the other categories. For instance, lack of organizational expertise about how to make data driven decisions affects perceptions about data quality. Often organizations have adequate data quality to be able to make specific decisions but lack the expertise to determine that the quality is adequate. Let's look at each of these categories and find out how the issues can be resolved.

The Myth of Absolute Quality

Healthcare organizations have recognized the need to improve and capture the right data in a data warehouse. Many of them have developed work groups and processes charged with improving data quality. Still, most perceive that data quality has only marginally improved and is still not sufficient for decision making.

Absolute Quality

This is primarily because these efforts focus on absolute data quality instead of the relative quality needed for most decision making. When absolute quality is the goal, improvement efforts are too diffuse and investments are made on areas with diminishing returns.

When absolute data quality is the goal, a system may collect too much data that is not useful for data analysis. These may include data elements that may be necessary for operational processes but aren't useful for analysis. Data quality work groups often get bogged down trying to improve data elements that aren't critical for analytic needs. For instance, they may focus on improving data elements that contain long textual data or a large number of similar categorical variables, which are almost universally not useful for analysis.

By focusing on absolute data quality, work groups often strive for unattainable levels of data quality. Analysis requires a level of data quality that is relative to the problem being answered. For instance, if the analysis focuses on highly aggregated trend data, lower absolute data quality is usually acceptable. In fact, this type of analysis is more adversely affected by changes in data quality over time, even when the change was an improvement.

Most data quality work groups have done a poor job of explaining and documenting what level of data quality is needed for specific decisions. As a result, organizations don’t use data that could have been used to improve their performance. Even when data is incrementally improved, word rarely gets out. This freezes most analysis users’ data quality perceptions at an early point in time, before any improvements are even made.

The Completeness Conundrum

Completeness is another data quality issue. Think about the completeness of claims data for health plans and patient accounting data for providers. Because of lags in capture, coding, processing, extracting, and loading into data warehouses, claims or patient accounting data for a specific time period may be incomplete. But analysts frequently analyze data without recognizing that it is incomplete. This leads to problems likely to be related to the completeness of the data.

Data lags that affect completeness are complex and are highly dependent on manual processes, such as medical record coding. Better communication between data entry staff, HIM professionals, and analysts would provide better understanding of these data lags.

Completeness issues also occur when data warehouses are built without integrating many data sources. Too often, data warehouse projects focus on primary data sources without planning to incorporate secondary sources such as departmental clinical systems or external comparative data. When these other data sources are not incorporated, all the data in the data warehouse is less valuable. For example, one organization incorporated electronic medical record data into a data warehouse. Because this data included items such as height, weight, blood pressure, and lab results, the organization was able to develop disease management programs more precisely and rapidly.

To improve data quality and completeness, develop well-defined and measurable data quality targets. These targets should reflect the relative quality needs of important organizational analyses, not strive for absolute perfect quality. Communicate data quality assessments and improvements across the organization so that data can be used more confidently. And create data quality workgroups that are diverse, including data entry staff, HIM professionals, analysts, and analysis users. HIM professionals must play a more active role in educating others about data nuances.

Learning to Ask the Right Questions

Expertise in data-driven decision making is essential, but healthcare has been slow to develop it. Too often, data warehousing training programs have focused on specific technical training, emphasizing the skills needed to use querying and reporting tools.

A few organizations have broadened the training to include more detailed knowledge about the data, but even this rarely includes discussions of the current level of data quality and what level is needed for specific analyses.

Unfortunately, courses that teach how to frame questions that can be answered by data are virtually nonexistent. Because of this lack of training, data suppliers, analysts, and analysis users have struggled to ask questions that can be answered by data.

One symptom of this lack of overall expertise is the constant reorganizing of analytic functions. One reorganization centralizes analytic staff and the next decentralizes it back into business and clinical processes. The underlying issue is the overall expertise and the development of networks of staff working together to solve problems. This network of staff can be developed regardless of how the analysts are organized.

Organizations must continually train staff to build their knowledge in data warehousing tools, data, and data-driven decision making to build their analytic expertise. One effective approach is to involve a wide variety of staff in training preparation and delivery.

Data entry staff and HIM professionals can explain the operational processes that create and code the data, analysts can explain how to use the tools and data, and analysis users can run through analysis scenarios designed to answer specific healthcare questions using the tools and data.

The training can also include a discussion of relative data quality and which types of analyses are appropriate given current data quality. This training should be given to a wide audience with more in-depth training on certain modules for some groups, such as more involved tool training for analysts.

Analytic Applications: The Secret Weapon

Most healthcare data warehouses haven't been designed to meet specific analytic needs. Data warehouse projects have focused on developing databases and deploying general data warehouse querying and reporting tools. This provides the components to be able to do analysis but often requires sophisticated analysts to add logic to turn raw ingredients into useable information.

The need to create logic to do analyses means analyses projects take longer, the breadth of analyses are generally narrower, and redundant work is often performed to replicate logic. This creates a bottleneck of analysis by relying on a small subset of sophisticated analysts to build analytic logic.

The additional logic needed to turn data into useable information can be categorized in several levels, which are explained in "Levels of Logic," below.

Levels of Logic

Analytic Concept	**Definition**	**Examples**
Technique	A piece of logic that is used to analyze a specific set of data	• DRG case mix adjusting • Rolling DRGs into product lines
Metric	A measure constructed from the data. Often the metric is a rate consisting of a numerator and denominator	• Diabetic retinal eye exam screening rate • Hospitalization readmission rate • Nosocomial infection rate
Analytic Application	A collection of techniques, metrics, and reports that are used to support a specific process	• Disease management analysis application • Cost analysis application

Analytics: Turning Data into Information

Collectively, techniques, metrics, and analytic applications can be referred to as analytics. Analytics turn raw data into useable information. Unfortunately, most healthcare organizations don't build the analytics they need into their data warehouses.

Some analytics are general to business processes that span multiple industries. Consider the financial management industry, where techniques such as debits and credits and metrics such as "days cash on hand" are used across industries.

In healthcare, techniques such as case-mix adjusting hospitalizations and metrics such as diabetic retinal eye exam screening rates are industry-specific analytics. A consensus is developing over the analytics needed to manage healthcare business and clinical processes. Most healthcare organizations are interested in the same analytics.

Some analytic standards and pre-packaged analytic applications have been developed (such as standards developed by the Joint Commission and the National Committee for Quality Assurance). These standards include specific instructions on techniques and metrics and even include the certifying of vendors with pre-packaged analytic applications. The implication for organizations is that they don't have to invent analytics on their own and that they can be more easily integrated into their data warehouses.

A robust application, however, should support many pieces of the clinical or business process. Current analyses from data warehouses usually provide partial analytical support for business or clinical processes, but organizations have limited resources to develop analytics and often target only narrow opportunities. Analytic applications should be designed to support the full analytical process and be designed using the same rigor as operational systems are designed and implemented.

A disease management analysis application is an example of an analytic application for healthcare. A robust disease management analytic application would identify organizational opportunities to manage diseases, help develop and measure improvement strategies, and provide patient registries of individuals who haven't received specific interventions instead of performing analyses that focus on support for a small portion of an entire process. A data warehouse without pre-developed methods, techniques, and applications will either take longer to do specific analyses or will miss out on opportunities entirely.

To avoid this pitfall, organizations should either develop or purchase analytic applications that integrate with their data warehouses. These applications should support industry standard analysis processes and provide support for all analytic phases of a clinical or business process.

It's particularly important that these applications be available to the entire organization. Departmental specific applications create redundancies and reduce the ability for multiple perspectives.

HIM professionals should be involved in the implementation and use of these applications. They can help with the understanding of specific analytic techniques, such as case-mix adjusting, and helping interpret data to capture specific diseases in a disease management application.

It's not enough to build databases and deploy querying and reporting tools. To receive the promised value from investments in data warehousing, organizations must carefully target data quality opportunities and increase mutual understanding by data suppliers and users. They must also improve their expertise in making data-driven decisions by training on a full range of techniques that include framing simple questions answerable by data.

And they need to build or buy analytic applications to automate routing sophisticated analytic techniques. HIM professionals have a significant role in realizing these improvements by fully participating in data warehouse projects.

Note

1. Zornes, Aaron. "Report on the DCI Data Warehouse World Conference, New York, July 28-30." *The On-Line Executive Journal for Data-Intensive Decision Support* 2, no. 32 (1998). Available at www.hpcwire.com/dsstar/98/0811/100254.html.

***Rich Moyer** (rich.moyer@milliman.com) is MedInsight product manager at Milliman USA in Seattle, WA.*

Article citation:
Moyer, Rich. "Data Warehousing: Myths, Pitfalls, and the Secret Weapon." *Journal of AHIMA* 74, no.10 (November 2003): 46-50.