Challenges Faced by E-Healthcare Comparative Data Warehouses

Vincent Dols, MBA

ABSTRACT

More and more healthcare facilities are equipping themselves with DSS systems and looking at their own data in a benchmarking manner. Some are considering accessing data from other healthcare facilities to decrease costs while improving quality of care. Many vendors claim to have that information. Some redistribute publicly available data; others actively seek information from a multitude of facilities. Those vendors are faced with many challenges as they try to provide information that is meaningful and usable for their clients. This article explores some of these challenges, particularly the data and technical challenges. It addresses the issues of standardizing data, as well as the technology to manage and secure that information, and provides examples that illustrate solutions that some vendors have implemented.

KEYWORDS

• Standardizing data
• Data infrastructure
• Quality checks
• Firewalls

More and more healthcare facilities are equipping themselves with DSSs and looking at their own data in a benchmarking manner. Because there is now an industrywide push for benchmarking, hospitals are looking beyond internal sources and considering whether to access data from other healthcare facilities. Many vendors claim to have the information hospitals need. Some vendors redistribute publicly available data; others seek information from a multitude of facilities.

Vendors face many challenges in their effort to provide information that is meaningful and usable for their clients. This article explores some of the
challenges related to standardizing data and putting technology in place to manage and secure that information. We provide specific examples of vendors’ solutions. The article is targeted to healthcare executives and DSS analysts who are concerned about the challenges faced by e-healthcare comparative data warehouses, and who want to know how some vendors overcome them. We will cover the functionality of software applications in a future article.

The Challenge: Different Input, Same Data

Healthcare facilities have difficulty standardizing their data elements across different internal computer systems. Some hospital systems do not interface with others. For example, an outdated admission system may not be compatible with the pharmacy system used to process orders, which may not be compatible with the billing system within the same facility. With this in mind, try to imagine what would happen if a multitude of healthcare facilities were to combine their patient information in a data warehouse. Would there be any standardization of data elements? We hope so. Healthcare comparative data vendors are faced with this challenge on a daily basis. Their information comes from many different hospital systems. Some have centralized ISs; some have DSSs; and some have none of those. One hospital may name the same data element exactly the same as another hospital. For example, Attending Physician is clearly defined and possibly used to describe the same data element. However, some may refer to another data element with two or three different names, such as Charge Code or Billing Code, which may also be referred to as Procedure Code. “Either data sources will increase and data formats will become standardized or the healthcare industry will be crushed under its own collective weight.”1 When looking at public hospital data, such as the data reported to the Health Care Financing Administration (HCFA) under the MedPar data, this challenge of difference in data has very little impact. The data submitted to Medicare are standardized and have very little detail because of the diversity of the data. Hospitals looking for more detail and a broader population than the Medicare population are forced to look beyond the public data to the proprietary data warehouses.

With regard to proprietary data warehouses: some vendors who are faced by this challenge have addressed it by mapping their clients’ data elements to standard data dictionaries. For example, Premier Inc. uses a staff of trained clinical consultants to map each one of their participating facilities’ charge masters to their proprietary standardized charge master that contains over forty thousand items. This method uses the concept of “common units,” that is, all data are converted to common units to support comparison.2 By doing so, Premier Inc. ensures that any facility in their data warehouse can compare its resource utilization with any of the other participating facilities. Mappings are common in the effort to bring unique data together. To make data comparable and create segments of data that are comparable, individuals are placed into
standard groupings: physician specialties, patient types, and payers. Grouping allows finding a common ground. If an orthopedic surgeon wants to identify a group of peers, it would make sense that all orthopedic surgeons for all hospitals be grouped in the same category. As the industry changes and the need for different groupings increases, additional standards will evolve.

The use of industry standards addresses that challenge; UB92 codes are common across all entities and do not need to be mapped. Some vendors collect standardized data from an organization’s legacy information systems to populate their data warehouse. If all data were standardized, it would be much easier for all hospitals to compare themselves; they would be speaking the same language. In the future, with the push for standards, this burden of mapping and translating data should be alleviated. Some good news is on the horizon with the arrival of the administrative simplification provisions of the Health Insurance Portability and Accountability Act (HIPAA), which requires healthcare players to implement uniform electronic data interchange (EDI) by October 2002. The HIPAA final rule mandates the implementation of industry specifications based on the X12 standards. In this case, all hospitals will be faced with the challenge of mapping their unique data elements to those standards, which could be mapped to vendors’ specific groupings. “An effective methodology provides encyclopedic standards that support a step-by-step, rules-based approach for mapping the data from the source to the warehouse, as well as rules for transformation, standardization, and summarization.”3

**Different Facilities, Same Results.** We hear over and over that it makes little sense to compare one facility to another. The phrase “We do things differently” actually says something good. If everyone practiced medicine the same way, there would be no examples of best practices. To make patients from one facility comparable to another patient from a different facility, some companies go beyond the DRG level and group their patients based on common methodologies. Premier Inc. and others use the well-known 3M APR-DRG grouper to classify their patients by severity levels. This methodology allows an entire patient population to be severity adjusted, based on the same algorithm. Doctor Smith can compare his patients to those of Doctor Jones without having to worry about how much healthier Doctor Jones’s patients are. It becomes an apple-to-apple comparison. Using open methodology that is easily accessible allows buy-in from the physicians who usually champion any benchmarking effort.

A multitude of factors make facilities hard to compare—factors often related to facilities’ size, geographical location, case mix index, or practicing physicians. All pose a challenge to a vendor wanting to make data comparable. The only remedy to this challenge is the number of hospitals that reside in the database. As hospitals enter into agreements with comparative data warehouses, all participants benefit. Premier Inc. uses the concept of peer group hospitals; users create their custom groups based on their preferences. Hospitals can choose their peer groups based on a variety of variables, including the
location of the hospitals, bed size, the percentage of managed care patients, or all of these. Having dynamic peer groups allows hospitals to choose peers for specific data analysis. One hospital may create a peer group with orthopedic hospitals to compare its physicians’ utilization to that specific peer group. The same hospital may have another peer group of hospitals with the same bed size that serves a large managed care population. This method allows patterns in outcomes to be analyzed.

**How Much Detail Is Enough?** The amount of detail is crucial to e-healthcare comparative data warehouses. More important is the ability to get end users’ questions answered, which means a meaningful combination of drills (down, across, and up). Imagine the relationship with a Rubik’s cube to understand the complexity of dicing out information. Data elements are that way, too. You need enough detail to get your unique questions answered. The need for the amount of detail is dictated by the pertinence of the question. For example, the simple question, How does my DRG 209 patient cost compare to the benchmark? can be answered by identifying only three data elements: hospital, DRG code, and cost. As the question refines to differentiate older patients from younger patients, a fourth element is included: age. If the physician wants to compare at the department level, a new level of data element is required. Each face of the Rubik’s cube breaks into multiple choices, and the database gets bigger. With over one hundred descriptive fields on a single patient encounter, the possible combinations are endless. One may want to look at utilization by departments by payer groups for a specific DRG; another may want to look at the same detail of information classified by attending physician specialties, admission type, and discharge disposition. It is no longer the number of data that is important, but the ability to slice and dice them in any possible way that makes them powerful.

**The Technical Challenge**

**Infrastructure of the Data: Space and Speed.** With all that information (data), e-healthcare comparative data warehouse vendors need a lot of processing power. Some data warehouses may contain millions of patient discharges for one year. All those have charge-code-level detail; most have them broken down by day of stay. Being highly efficient in data storage and allowing for quick retrievals is critical. Some vendors may process the data on a separate data transformation system and have a completely separate, dedicated data warehouse. Technology is evolving so rapidly, with hardware price dropping so drastically that what might be perceived as a challenge is easy to solve. Adding terabytes of drive space is no longer a dream or a luxury; it is becoming a daily adventure. It appears that every time a PC catalog is printed, the price of hard drives remains the same, while their size increases drastically. For example, in 1994 a 200 MB hard drive sold for about $150. In 1996, for the same price, the first 1 GB drive was available; now a 30 GB hard drive is
the same price. With price per byte dropping so quickly and computers getting so much faster, e-healthcare comparative data warehouses should be delighted by the improvements in technology.

**Integrity of the Data: Quality Checks.** Each data transformation comes with a potential for errors. Missing patients during a submission of data is a common occurrence. Even more common is a missing field for a patient record, which may make the entire record questionable. How good are data if they are not accurate? How much effort is required to regain confidence in the data once their validity has been questioned? We all know the answer to those questions and know that integrity is of the essence. Most vendors create rules to check for data integrity. Having a list of required fields that is as complete as possible is one way to address this issue. For example, all patients should have demographic information such as Gender, Age, Race, and Marital Status. They should also have Principal Diagnosis and Attending Physicians. Having required fields is the first level of data integrity checks. The second level is related to the data logic. Does it make sense to have a male patient delivering a baby? Does it make sense to have an obstetric physician be the attending physician on an orthopedic case? All those rules are set up to identify potential data issues. Most vendors develop their own rules for their logical tests. It has been shown that using clinicians to review a sample of the data in a similar way that hospitals do their medical record review is an effective quality check. To name a few of those quality checks, some may match the length of stay of the patients to the room-and-board charges. They may also check for the presence of surgical implants on hips and knee replacements and match the anesthesia time to the surgery time. All those processes in place address the challenge of data integrity.

**Security of the Data: Firewalls and Data Encryption.** “As hospitals increasingly use the Internet and similar technologies to exchange patient information, the vulnerability of that information has become a major concern.”\(^4\) The Privacy Act of 1974 mandates that federal information systems must protect the confidentiality of individually identifiable data. Section 5 U.S.C. 552a (e) of the Act is clear: federal systems must “establish appropriate administrative, technical, and physical safeguards to insure the security and confidentiality of records and to protect against any anticipated threats or hazards to their security or integrity which could result in substantial harm, embarrassment, inconvenience, or unfairness to any individual on whom information is maintained.”

To ensure data confidentiality, most e-healthcare comparative data warehouses do not contain any information that is individually identifiable. Most transactions are encounter-driven and contain no personal information. Identifiers such as patients’ names, addresses, and Social Security numbers are stripped from the files the hospitals submit to those companies. All transmissions of raw files are usually done via SSL, S-Mime, or encrypted FTP sites to prevent any potential breach of confidentiality. In addition, to protect the
privacy of the data in the warehouse, vendors create business rules against displaying any data elements if the sample size is too small and could lead to a logical guess of the identity of the submitter. This rule complies with FTC Antitrust Safety Zone conditions. “There are at least five providers reporting data upon which each disseminated statistic is based, no individual provider’s data represents more than 25 percent on a weighted basis of that statistic, and any information disseminated is sufficiently aggregated such that it would not allow recipients to identify the prices charged or compensation paid by any particular provider.” Access to that proprietary information is done through a password-protected environment, also complying with HCFA guidelines.

**Conclusion**

So they can compare their practice to that of other facilities, some hospitals are turning to e-healthcare comparative data warehouses. How good those warehouses are depends on the way vendors handle the many challenges they face. Because hospital data are seldom standardized, it is important to evaluate the processes that the vendors put in place to address this issue. The integrity of the data is the most important factor when looking at benchmarking. At the hospital level, it is crucial to understand that those vendors are continuously trying to prevent the “garbage in, garbage out” effect by putting quality checks in place. Reducing the number of data discrepancies at the hospital level can only be beneficial to both parties.

**References**


**About the Author**

Vincent Dols, MBA, is a senior consultant with Premier Inc., based in Charlotte, North Carolina. He has been involved in internal and external healthcare benchmarking for more than seven years.