

The Healthcare Industry and Data Quality

by
Melissa C. Tzourakis
of
Medicode, Inc.
5225 Wiley Post Way, Suite 500
Salt Lake City, UT 84116
Internet:mtzoura@medicode.com

1. Introduction

The healthcare industry is experiencing great change caused by the advent of managed care, a consolidation of significantly large entities, additional reporting requirements from federal and accrediting bodies and a demand from consumers for accountability.

Of vital concern and importance is the data flow from the point of patient service delivery to an end point of decision support and analysis of the data derived from the encounter. The historical use of these data, their current uses and how industry dynamics have created a need for improved data quality provides the context for a discussion of healthcare data quality. Equally important is an understanding of how data quality is defined within the industry, how it is valued and what data quality processes have or have not worked.

2. Healthcare Data Defined

Healthcare data is defined as that information used to provide, manage, pay and/or report on the services used across the entire healthcare system. For the most part, the healthcare industry has been trailing most other industries in its acceptance and use of standard data models/formats, acceptance of newer technologies and the overall integration of data into executive decision-making. A primary cause is that healthcare data is much more complex than most other data sets. Unlike other industries, limited transactions have been standardized, but the basic data quality lessons learned from other industries can be applied.

The origin of healthcare data is an encounter between a patient and a healthcare provider. This provider may be a physician, hospital, radiologist, physical therapist, lab technician or a physiologist. From this encounter, the provider will record the service rendered (an office visit, a lab test, room and board for an inpatient stay), the conditions of the service (the diagnosis, date, place of service), patient information (sex, age, patient history, insurance information, height, weight, etc.) and clinical information (result of tests, prognoses, consultation notes). All of this information is useful to the provider and appears in a variety of forms, from paperwork to billing systems. Information is also divided by provider since each one documents the services provided. A provider has limited access to information logged and maintained by another provider, with the

exception of sharing limited patient history and demographic data during clinical collaboration. There is almost no automated clinical connectivity between providers.

Some standardization exists in the way data is captured. Most of this standardization has come about in the last 10 years because of the government's Medicare program. For a provider to be reimbursed by Medicare through a fiscal intermediary (i.e., the payor), the provider must submit a "claim" for services rendered. Historically, the provider submitted a nonstandard claim, which minimally included the patient name, age, sex, a unique employer/group identifier, the services provided and the date they were delivered. This information was used almost exclusively by the payor to reimburse the provider.

At times, providers report services improperly for reimbursement at a rate more than appropriate for the services provided. Studies have shown that less than 5 percent of providers actually submit fraudulent claims, but they account for most overpayments. As concern of fraud and abuse escalated from this apparent lack of control, standards began to develop. Certain data elements were submitted based on standard code definitions for the procedure or service performed, and the diagnosis. The service codes are maintained by the American Medical Association for physician services while diagnosis codes are maintained by the government's Health Care Financing Administration.

Once implemented, these coding rules become the basis for payment to the provider while concurrently being captured as "reimbursement" data by the payor. Data collection and submission falls on the provider. These professionals are trained clinicians who view administration and reimbursement requirements as a burden. Further, the complexity of correct coding creates two additional problems. First, there are approximately 8,000 procedure codes and 16,000 diagnosis codes. A physician's office staff salary is typically a minimum wage and these individuals may not have had adequate training on coding rules and methodologies. Secondly, these methodologies change annually, but the training and materials may be viewed as an unnecessary expense. While standardization has evolved further, data actually captured are recorded and resubmitted by the provider for reimbursement, but may still not be representative of the encounter since the clinical aspects are recorded on paper and stored by the provider.

3. The Changing Healthcare Environment

Before the 1980s, the payor would reimburse the provider a fee for services rendered. Patients would see providers at their discretion and be responsible for paying part of the bill. In the 1980s, the formation of health maintenance organizations (HMOs) entirely changed payor organization and healthcare delivery. Early HMOs were created by providers (physicians) who assumed some of the financial risk. They marketed themselves on the premise that they could provide comprehensive healthcare for a lower cost by better managing patient care through a Primary Care Physician (PCP). The patient paid a flat co-payment for office visits and prescriptions, and were covered at 100 percent for most other services. Access to the healthcare system became

controlled by PCPs and limited to providers within a "network" who were paid a negotiated fee rather than as billed. Often, some of the physician's fees were withheld until year-end and then returned only if the physician had met certain measures indicating he/she had managed care effectively. Other payment models emerged, including physicians being paid in advance on a per-member, per-month basis (called "capitation"). All these measures were designed to bring down costs.

With the popularity of HMOs in the 1990s, managed care was seen as a panacea for controlling spiraling healthcare costs. As healthcare inflation normalized into the single digits, greater public attention was drawn to quality-of-care and access-to-care issues. As a result, HMOs became subject to accreditation and were required to provide quality-of-care information to employers. At the same time, pressure continued for them to reduce costs to remain competitive in a rapidly consolidating market.

To administer the previously cited reimbursement methods, HMOs required additional and more accurate data from providers. However, most HMOs were dependent upon systems developed for typical fee-for-service indemnity insurers. While this expanded the information captured from providers, the information systems' main focus was still on reimbursement.

4. Data Quality Issues in the Current Healthcare Setting

To trim costs and maximize productivity and value, healthcare entities are, for the first time, turning to their data and decision support tools to validate their cost and quality initiatives. For health plans to effectively monitor activity within their networks, they must have an accurate picture of encounters that take place between provider and patient. Ideally, the correct data to analyze the effectiveness and quality of these interactions would be the original clinical data collected by the provider at the point of service. Unfortunately, the clinical data are not currently available because of a lack of automation and standard methodologies for collecting clinical information. Consequently, the healthcare industry has had to accept the fact that available reimbursement or claim data are most representative of the encounter.

Unfortunately, encounter data can further degrade from the payor's handling of the claim. Data may not be captured completely or accurately. System maintenance files and insured enrollment information may be incorrect, referral data may be missing (causing an incorrect denial), data may be keyed incorrectly or a contract may not be set up correctly. There are many other points within a payor's claims adjudication system where data may further degrade.

Claims operations areas focus on the correctness of the financial payment that results from the claim. This is true for both traditional insurers as well as managed care plans. To curb fraud and abuse in the early 1990s, most health plans began to install clinical editing systems that assess combinations of procedure and diagnostic codes, automatically notifying the entrant of a suspicious or incorrect combination. The installation

of these systems was mainly to promote cost-effective and clinically appropriate processing. Their capabilities represented the first automated initiative to address quality of claims.

Health plan analytical and reporting demands have dramatically increased during the past three years. The traditional focus on production and financial reporting gave way in the 1990s to enterprise-wide access by business users needing information on medical management, enhanced account reporting, marketing, etc. Because of these users' demand for data, health plans have begun to realize the need to install warehouses of historical data to address these business needs. As users began accessing data, the enthusiasm to provide information quickly turned to frustration. Many times the numbers pulled did not match other reports. Data were found to be incomplete and inconclusive, and there was generally a lack of understanding by analysts as to anomalies and inconsistencies in these data.

Often in managed care systems, contract variability from complex contract management structures and terms could not be accommodated in the relational data structures of the dictionaries. Because of the industry's use of new reporting and analysis tools, it was often unknown if quality deficiencies were caused by the tools, how they were implemented or the data. The result of this conflict is an overall mistrust in the reporting from these systems since they did not effectively link reporting output with data quality variables.

5. The Need for Effective Healthcare Data Quality Management

The culmination of these data quality problems points to the fact that there is no true data quality ownership. Operations areas are charged with decreasing the turnaround time from receipt of data to the date a claim is paid. They are also audited on financial accuracy. Information Systems departments maintain a legacy system, and have projects backlogged to bring the system up to the current pricing and flexible contractual requirements. It is very rare to find an organization that has thoughtfully placed a high level of commitment to ensure data quality.

This failure to manage data quality is ironic because the data are the only true asset held by these organizations. Some have accepted the notion that bad data is better than no data. While some visionaries attempt to tackle the issues affecting data quality, this daunting task is complicated by the fact that data quality issues are many and mitigating resources available to address them are few. Ensuring data quality does not neatly fall into one area of an organization -- it must to be addressed by the entire organization and senior management commitment is critical to successful implementation.

To further complicate this issue, few people have direct experience in dealing with this problem throughout an organization. Most data quality initiatives are born where the need is the greatest, usually in analytical or decision support departments. These departments establish their own quality needs, use the data to support database accuracy levels and pass reports to operations areas. Through these relationships, operations is influenced to consider its data requirements and change or upgrade their level of data capture. This activity

has proven to be ineffective since many analytical requirements directly conflict with operations processing requirements. Without controls in place at the point of entry, little will change to ensure more accurate, consistent data entry in the future.

A combined, cooperative effort involving all parties is required to adequately focus on data quality ownership and accompanying processes. Accuracy of submissions is, for the most part, beyond an enterprise's control and, for the time being, cannot be effectively addressed. Operations must be held accountable for maintaining an accurate and complete data entry process. Systems, in turn, must support operations by supplying the proper tools to periodically audit the incoming data stream to support building the necessary controls into the system. Analysis must meaningfully measure and report results to operations.

5.1 Tools to Ensure Healthcare Data Quality

Effective integration of data quality methodologies is impossible without the proper tools. To solve the problem, some plans (and vendors responding to market need) choose to build their own tools. For any of these tools to work effectively, they must be able to perform three functions: cleanse front-end input data, sample and audit data passing through the front end and cleanse back-end repositories data.

5.2 Cleansing Front-End Data

Front-end data cleansing can be performed on almost any platform, but it must be "programmable" by the business user. An interface must exist for the user to input a variety of business rules and controls tailored to unique business needs. These rules may define whether a field is complete or dictate the conditions under which a particular claim is entered by setting up logic that defines the relationship between fields. The interface must also access clinical rules that determine the appropriate use of service and diagnosis codes. Some statistical algorithms may also be used to detect whether the charge is beyond payment scope for a service indicated or is an incorrect submission. The number and complexity of cleansing rules will vary from plan to plan. A system that has more than adequate controls and the ability to incorporate contract specifications in its adjudication system will have fewer rules than one that is not as flexible or sophisticated.

To avoid reduced productivity in operations, the tools must feed an integrated workflow system that routes the claim and notifies an analyst that intervention is required, supported at the workstation itself.

5.3 Sampling and Auditing Front-End Data

The second function for a data quality tool is periodic auditing of a sample that has passed through the front end. A sophisticated methodology is required to test for front-end cleansing performance, which, in turn, identifies data anomalies and opportunities to add to the rules in the front end. Additionally, these audits serve as a quality improvement process by feeding specific information on errors being made back to processors, supervisors and trainers, describing which areas are most prone to errors.

Traditionally, claim auditing is a laborious manual process that is typically supported by a system-generated, paper audit trail. User selection of the sample is often manual and, at best, these data are captured in a spreadsheet for manipulation and analysis. With these limitations, the audit method is incapable of producing the most meaningful result.

A sampling tool must be developed to allow the user to establish a random selection of claim records by category (analysts, group, provider type, etc.). This selection would then populate an interface for the auditor to use in verifying the claim submission's completeness and accuracy. This interface supports the identification of errors and categorizes by the types of errors observed. Collected in a database, this information can then be used for reporting in a variety of ways. As an example, accuracy can be assessed at any time in the month by the processor's supervisor, providing the processor with knowledge of his/her current quality level. This allows that individual to address substandard performance early in the month by focusing on corrective action to increase quality performance. Valuable information is flowed to trainers, identifying areas where most processors are having difficulty. Targeted retraining and/or a restructuring of the training process to achieve greater efficiencies may result. As quality issues are discovered, auditors will report any new business rules identified from the audit for incorporation into the front end.

5.4 Cleansing Back-End Database Data

Back-end data cleansing of the analytical database against predefined business rules is the third function of a data quality tool. Concurrent with installing a front-end cleanser, an analysis on prior years' data is also needed to bring these legacy data to the same standard as the current data being entered. With an increasingly consolidated marketplace, data from these once-separate entities require an enormous amount of mapping from one data set to another. There are also definitional differences that must be agreed upon. The cleanser must also produce and map fields so they are more meaningful to the analyst, and use standard terms and definitions for categorizing fields rather than system-generated mnemonics. The cleanser may also group the data into categories by product, benefit or episode of care. The process supports a more accurate and timely pre-analysis of certain aspects of the data, which, in turn, may offer opportunities to enhance the front-end edits prior to full data quality implementation.

5.5 Other Data Entry Points

Accounting for the three previously mentioned functions of a data quality tool would provide an effective framework to ensure accuracy in claims data. The next step is to look at all other significant points of data entry into the system and propose a similar process. Entry points include provider, enrollment/eligibility, utilization information and authorization data. These areas also contribute to data quality problems because the data are

constantly changing and the maintenance requirements are often unsupported with adequate resources. While these data directly affect financial transactions, often the controls were never built into entry screens and system audits.

6. Conclusion

The organization and delivery of healthcare services in the U.S. is under dynamic revision -- a revision that is redefining the relationship between providers, payors and the public. Market forces and initiatives mandating quality healthcare delivery challenge all healthcare participants to make more positive and responsive contributions during this evolutionary process.

A critical element in this healthcare scenario is a commitment to ensuring data quality. These complex data drive accurate, comprehensive and appropriate medical reimbursement, as evidenced by the Medicare program's focus on evaluation of \$159 billion currently being spent on elder care in this country. Healthcare entities, participating in a market characterized by increased competition and consolidation, must use their data to validate cost and quality initiatives. Automated claims editing systems that assess appropriate processing mark the first information systems attempt to deal with the quality of claims processing using quality data-based information. In the future and with the support of senior management, healthcare data quality initiatives such as these will grow because of their recognized importance in the process. Operations, systems and analysis functions will continue to define data quality requirements and work together to implement these requirements. As progress continues, front-end data cleansing, in-process audit sampling and back-end database cleansing will be baseline elements that characterize an effective approach to ensuring data quality.

Melissa C. Tzourakis is the Director of Payor Product Development for Medicode, Inc. A graduate of Long Island University with a B.S. in Biology, Ms.Tzourakis has eight years of experience managing and administrating health insurance claims data quality, analysis and processing issues in development and operations environments. Her background as a database and software developer contributed to her extensive experience in implementing clinical editing systems on PC-based and mainframe platforms.

Medicode, Inc. is a healthcare information company that combines its databases, software, clinical expertise and technical publications to help claims payors, providers and employers manage cost-effective healthcare delivery and reimbursement. Founded in 1983, Medicode offers its products and services to major commercial health and accident insurance companies, workers' compensation organizations, Blue Cross/Blue Shield organizations, managed care companies, third-party administrators, academic medical centers and others.