CHAPTER 5

Data Quality and Governance

It is a capital mistake to theorize before one has data.

—Sir Arthur Conan Doyle

Healthcare leaders and quality improvement (QI) teams rely on having the best possible evidence on which to base decisions and evaluate quality and performance. The best possible evidence requires effective and accurate analytical tools able to provide understanding and insight into quality and performance based on data. In other words, without good data, analytics and the evidence it provides is likely to be suspect. Having good data for analytics and QI begins with effective management of data. This chapter will focus on how IT and QI teams can work together to ensure that a data infrastructure is available to support quality and performance improvement teams with the high-quality and highly accessible analytics they need.

Data is not a static asset within a healthcare organization (HCO). People unfamiliar with how data is managed in an organization may only consider data to be something that is entered into a computer system and sits in a database until subsequently reported on for management purposes. In addition to being a very valuable asset to an HCO, however, data is in fact a very dynamic asset within an HCO. As new information systems are adopted, and analytical requirements evolve, the underlying infrastructure and management of data must also evolve. The four main activities associated with maintaining a data system that supports the needs of an HCO (or any large organization) consist of data modeling, data creating, data storage, and data usage.[1]

1. **Data modeling.** No healthcare analytics is possible without data, and no data storage is possible without a data model. A data model is a

"wayfinding tool for both business and IT professionals, which uses a set of symbols and text to precisely explain a subset of real information to improve communication within the organization and thereby lead to a more flexible and stable application environment."[2] Data models have also been described as "what ties human thinking with computer processing [and] provides a blueprint for database and application system design."[3]

2. **Data creating.** Data is created through the day-to-day operations of providing healthcare, but can also be obtained from other sources including other organizations. Data at this stage can be considered the raw material that becomes information after additional processing and analysis.

3. **Data storage.** When data is obtained or created, it must be stored in a database. Data may be reformatted or otherwise transformed to improve the efficiency of the database, or to make data more accessible to information users. With data storage comes the need for protection of the data; healthcare data must be securely stored to prevent unauthorized access and to protect the privacy of the individuals whose information is stored.

4. **Data usage.** The main purpose of creating and storage data, of course, is to use it. Analytics is one of the primary ways in which HCOs can use data to enhance decision making and to help achieve quality and performance improvement goals.

These four key activities do not, and cannot, operate in isolation; to ensure that a high-quality and secure data infrastructure is available for decision makers and other information users throughout an HCO, an effective data management culture and structure within the HCO is required.

The Need for Effective Data Management

The adoption of healthcare information technology (HIT) in the form of electronic medical records (EMRs), electronic health records (EHRs), and other clinical systems continues to expand. For example, in the United States, 69 percent of primary care physicians reported using an EHR in 2012, compared to 49 percent in 2009.[4] This increase in EHR adoption in the United States may be in part due to government incentives (such as the HITECH Act and Meaningful Use requirements), but also because of the potential benefits of improved patient care offered by HIT. Most other industrialized countries are also experiencing increases in EMR adoption—for example, Great Britain, Australia, New Zealand, the Netherlands, and Norway all report EHR adoption rates of over 90 percent.[5] Great Britain,

long a leader in HIT use, is also a leader in the use of data for quality and performance improvement.

Although healthcare information systems are still in their relative infancy, they are generating large volumes of data. As the growth rate of HIT adoption continues to increase, the volume of data collected by these systems will also increase. Recent estimates are that healthcare data totaled 150 exabytes in 2011, and will continue to grow into the zetabytes and perhaps even yottabytes soon after.[6] To put that into perspective, consider that a gigabyte (GB) is 10^9 bytes, an exabyte (EB) is 10^{18} bytes, a zetabyte (ZB) is 10^{21} bytes, and a yottabyte (YB) is 10^{24} bytes. Many large healthcare networks have data volumes in the petabyte (1 PB = 10^{15} bytes) range.

While this very large and growing volume of data presents an exciting potential for use in quality and performance improvement activities, it is by no means a trivial task to ensure that this data is available, and usable, for such purposes. Fundamentally, data that is used for healthcare quality and performance improvement needs to be:

- **High quality**—to ensure that the information generated from analytics is valid and useful.
- **Well documented**—so that analysts and developers using the data are aware of its context and meaning.
- **Easily accessible**—and available in a data warehouse (or similar data store) to ensure that it is available for analysis when required.

To ensure that these three fundamentals are achieved, HCOs require strong and effective data governance strategies and structures. Organizations that do not employ strict data management and data quality policies run the risk of accumulating large quantities of data that is essentially unusable without expending great effort to clean and otherwise make it more usable. HCOs have a responsibility, as data owners, for:[7]

- **Data quality.** Determining the data quality levels required for different fields in the database, and how best to achieve those levels. Note that not all data requires the same level of quality. For example, data relating to demographics, clinical orders and observations, and billing information needs to be of the highest quality, whereas supplemental information not impacting patient care may not need to be of as high a quality.
- **Security and privacy.** The determination of who is able to access which data. Healthcare data is perhaps an individual's most private information, so maximum effort must be made to ensure that privacy is maintained at all times. Almost every conceivable analytics operation

can and should be done using anonymized data. In an instance where analytics is helping to predict an outcome for a particular patient, only the clinical provider should be unblinded to both the patient and the outcome.

- **Business rules.** The rules embedded in code that validate, map, calculate, or otherwise transform raw data into higher-quality and more useful data must be maintained and updated. Business rules are often used to calculate interval times, recode variables, and perform other transformations on raw data to make information easier to analyze.

- **Availability.** When information must be made available and when acceptable downtimes are possible. Interestingly, I have seen a shift toward the need for more round-the-clock availability of data systems and analytics. Clinical and management decision making does not stop at the end of the working day, so neither should the information required to assist with those decisions.

- **Periodicity.** How often the data needs to be updated (this can range from quarterly to monthly to near real time, depending on the needs of the system).

- **Performance**. The response time experienced by the user when accessing the system. The performance of analytical tools must be exceptionally high when supporting near-real-time clinical decision making, whereas slower performance is likely to be acceptable in non-real-time situations (such as when running reports).

As illustrated by the previous list, the responsibilities associated with ownership of healthcare data extend far beyond simply purchasing and maintaining database servers and software to house the data. These activities are essential to ensure that HCOs have high-quality data that can be utilized for quality and performance improvement, research, and management decision making, is accessible when needed, and protected from unauthorized access and usage.

Data Quality

The most important aspect of any analytics system is access to accurate, high-quality data. Before any reports are built, analyses performed, and dashboards deployed, ensuring that source data is trustworthy must be the first priority. Without data that is accurate, it is impossible to trust in the results of the many algorithms and other computations that constitute analytics. If the veracity of the raw material is called into question, then certainly the results of the computations using that raw data must also be suspect.

Accuracy and Quality Are Key

The most important aspect of any analytics infrastructure is access to accurate, high-quality data.

Without high-quality data, many quality and performance improvement projects may be negatively impacted—especially large-scale projects using a structured improvement methodology like Lean or Six Sigma. For this reason, healthcare QI specialists are important and necessary stakeholders in data quality. Improving quality and performance requires a solid understanding of previous and current performance, and an ability to detect changes in data that signal an improvement (or worsening) in performance. Having poor-quality data will likely increase the difficulty in detecting changes in performance, or lead to misinterpretation of data and incorrect conclusions.

HCOs need to determine their own data quality requirements. To assist with this determination, there are many dimensions that can be used to quantify the quality of data. The Canadian Institute for Health Information (CIHI), for example, uses the dimensions outlined in Table 5.1 for data quality.[8] The CIHI dimensions of data quality, identified by an asterisk in Table 5.1, are useful for gauging the quality and usability of a data set for use in healthcare analytics applications. In addition to the CIHI data quality dimensions, completeness, conformity, and consistency have also been identified as necessary dimensions of data quality,[9] and are also described Table 5.1.

Achieving Better Data Quality

Having good data cannot guarantee that effective analytics tools can and will be built, utilized effectively by an HCO, and result in the quality and performance improvements desired. Bad data, however, will most certainly mean that efforts to use information will be hindered due to a lack of trust or belief in the analytics and/or its results.

To begin with, how do we describe "good data"? Quality expert Joseph Juran states that "data are of high quality if they are fit for use in their intended operational, decision making, and other roles."[10] In this definition, "fit for use" means free of defects and possession of desired and necessary features. Achieving good data, however, is hard work. HCOs need to start with the source systems, and in particular the users of those source systems. In my experience, one of the best ways to improve end users' data entry is to share the analyses with them in the form of performance reports and other relevant forms that are meaningful to the individual. If end users can

TABLE 5.1 Data Quality Dimensions

Data Quality Dimension	Description
Accuracy*	Reflects how well information within (or otherwise derived from) data reflects the actual reality it is intended to measure.
Timeliness*	Reflects how recent and up to date data is at the time it is available for use in analytics. Measured from the time it was generated (or the end of the reference period to which the data pertains) to the time it is available for use.
Comparability*	Refers to the extent to which the data is uniform over time and uses standard conventions (such as common data elements or coding schemes).
Usability*	Reflects how easy it is to access, use, and understand the data.
Relevance*	Reflects how well the data meets the current and potential future analytics needs of the healthcare organization.
Completeness	Refers to how much of all potential electronic data (for example, from electronic health records, claims data, and other sources) is available for analytics.
Conformity	Reflects how well the available data conforms to expected formats (such as standardized nomenclature).
Consistency	Measures how well values agree across data sets and the extent of agreement exhibited by different data sets that are describing the same thing. This can range from the use of consistent acronyms to standard procedures by which to document patient discharge time.

* Denotes a data quality dimension identified by the Canadian Institute of Health Information.

see how the data is being put to use (and how the results can impact both their job and patient care), they may be less likely to dismiss accurate data entry as an unimportant and irritating part of their job.

Monitoring Data Quality

We have used our own analytics tools to detect poor data quality and automatically alert those people who can take corrective action. Rapid feedback dramatically increases the speed at which data quality issues are addressed.

When more direct measures were necessary to improve data quality within my own HCO, we have used our own analytics tools to encourage managers to provide coaching to staff when staff performance is not what is expected. For example, a project I was on utilized analytics tools to automatically measure the rate at which triage nurses were overriding a computerized scoring algorithm. It was found that the overrides were occurring primarily because nurses were not filling in all the required information appropriately, and the system was generating invalid results due to this data quality issue. By implementing automatic e-mail alerts to managers when the override rates were higher than desired, the managers could provide coaching or more in-depth training to staff so that they would complete all necessary data fields. This relatively simple intervention reduced the override rate of nurses using the tool from around 45 percent to around 10–15 percent, which was much more acceptable from a clinical standpoint. Furthermore, most of the overrides post-intervention were the result of real clinical disagreement with the results of the algorithm, not a result of poor data quality negatively impacting the calculations.

Tip

Although there are myriad possible causes of data quality problems, data quality usually begins at the source.

The best approach to improving the quality of healthcare data is to prevent data quality issues in the first place. Although there are myriad possible causes of data quality problems, data quality usually begins at the source. That is, poor data quality is most likely to be a result of the way users interact with clinical or other information systems, poorly designed user interfaces, and deficiencies in data entry validation. Less likely but still possible, poor data quality may also be the result of errors in system interface code or other instances where data is communicated between two systems.

In my experience, healthcare quality initiatives have been hindered by data quality for a number of reasons, including:

- **No data available.** Many HCOs are still not fully computerized, meaning that a lot of important data is still locked away on paper in filing cabinets. If the data needed to answer a particular question or identify a quality issue is not available to begin with, that is a blind spot that needs to be addressed. There will always be gaps in data, however—even fully computerized HCOs cannot possibly capture data on every single step of every process in care delivery.

- **Data doesn't represent workflows.** Many HCOs have had EMRs and other clinical information systems for many years. Unfortunately, processes often change more quickly (and more often) than the information systems can be updated to reflect these changes. This results in business processes becoming out of sync with the data being collected, or different meanings being attached to existing data fields. It requires commitment and effort to keep data stores up to date to reflect these changes.

- **Personal inclination.** Even if information systems accurately map to workflows, processes are well defined, and frontline staff are fully trained on the system, some staff may still choose to use the information system in an inappropriate way, which may result in the data associated with their entries being less valid. Comments such as, "Oh, I know I'm supposed to enter the data that way, but it's too many clicks, so I don't do it like that," are particularly frustrating. Obviously, frontline staff are not meant to be data collectors for QI teams, but at some point they should understand that the data being collected is used to improve care quality (and job quality) for everyone who works in the organization.

- **Data errors.** There are still plenty of good old-fashioned typos that can make analysis more challenging. For example, incorrectly spelled medications, incorrect medical record numbers, and other incorrect information can cause problems when linking data or summarizing information for analysis. Many source systems still do not employ sufficient data validation to check for the accuracy and validity of certain types of data commonly entered into the system with errors (for example, temperature, height, and weight ranges).

Overcoming Data Errors

Poor user interface design on clinical systems can be a source of poor data quality. Particularly dubious are pull-down lists that also offer the choice to enter your own data. This is where you see how creative people can get retyping information already contained on the list. For example, the Mode of Arrival field on our triage note contained an option for "Ambulance," but also had a free-text box where values for this field could be typed in. The result was that instead of having a very clean summary of how patients arrived at the emergency department, we had to contend with entries such as "arrived by ambulance," "ambulance #209," "EMS ambulance," "amblance," or simply "amb." Thankfully, the use of the free-text box was discontinued, and the data from the triage note is now much cleaner.

With growing volumes of data and increasing reliance on analytics for decision making, data quality is a major focus of research, and root causes of data errors have been studied extensively and systematically. The many possible causes of data quality problems have been grouped into 10 major causes (several of which are given in the following list, with elaboration added).[11] Addressing these root causes of poor data quality will greatly enhance the quality of data that is available for analytics.

- **Multiple data sources.** This is common in healthcare, with multiple source systems (including registration, lab, and diagnostic imaging system) storing a single patient's data; each source system may have its own data validation rules and data formats.
- **Subjective judgment in data production.** Without clear documentation and definitions of the data (both at the source *and* at time of analysis), personal interpretations of what data means impact what is recorded and how the data is eventually interpreted.
- **Security/accessibility trade-off.** If security is too tight on databases for analysis, developers may look elsewhere where data is more accessible. This is why many "rogue" copies of data are created.
- **Coded data across disciplines.** Different source systems may code data using different coding schemes (e.g., ICD-9 versus ICD-10), which may hinder comparability and compatibility if clear mappings are not available.
- **Complex data representations.** Data representing complex processes, or even simple processes that are stored iteratively, may induce errors in the extraction and analysis of the data. (Data modeling helps to alleviate this issue.)
- **Volume of data.** Large volumes of data can be challenging to work with, especially if there is limited computing power available to work with. Some desktop statistical packages can only analyze data that is in memory (as opposed to on disk), which limits the size of data that can be analyzed.
- **Changing data needs.** As the business, quality, and performance improvement needs of the organization evolve, be careful not to let these new requirements bias your understanding of the data unless the process(es) and data have, in fact, changed.

As mentioned, having accurate, high-quality data for analytics starts at the source. Analytics teams need to work together with data warehouse managers and frontline staff to ensure that all possible sources of poor data quality are identified, reduced, or eliminated. In my experience, it has been helpful for members of the analytics team to be part of system change request committees. It is likely that whenever a change to a source clinical system is required, it is because of a change in process, or because of a new process

For a comprehensive listing of data quality references and additional resources, please visit this book's web site at http://HealthcareAnalyticsBook.com.

and the need to be able to capture data from that new process. Having analytics and data warehouse team members on those change committees helps to ensure that any potential changes in data (either new fields or changes to existing data) are taken into account during change request discussions.

The full potential of healthcare analytics cannot be realized, however, if data is locked inside operational, divisional, or other information silos. One of the exciting capabilities of analytics is finding new relationships between processes and outcomes, and discovering new knowledge; this is truly possible only when data is integrated from across the enterprise. As data is integrated from multiple clinical and other systems from across the HCO, however, its management becomes an issue. How data was managed in an independent, one-off database is not suitable at all for managing data integrated from across multiple source systems. Failing to effectively manage healthcare data, across all its sources, will seriously impede the development and use of effective analytics.

Data Governance and Management

Because the quality of data is critical to quality and performance improvement activities, it is good practice to have people within the HCO who are responsible for data quality. Discussions of enterprise data quality, however, invariably raise issues of data ownership, data stewardship, and overall control of data within the organization. HCOs with very little, if any, formal data management and governance exhibit data quality management that is ad hoc and reactionary—action is taken only when it is too late and something needs to be fixed. HCOs at the opposite extreme have implemented layer upon layer of approval requirements, stewardship, and change management committees; such bureaucracy, however, can backfire and pose a risk to data quality when adhering to rules that are too strict inhibits the flexibility required to respond to changing patient care processes, changing systems, and changing analytics requirements.

Healthcare Organization Data Governance

To ensure that high-quality data is available for QI activities, HCOs must ensure that appropriate and effective data quality management processes

are in place. In addition, these processes need to be enforced, and they need to provide a balance between the rigor necessary to ensure stability and the agile responsiveness required by the evolving data needs of the HCO.

According to the Data Governance Institute, data governance is "a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods."[12] Data governance helps HCOs better manage and realize value from data, improve risk management associated with data, and ensure compliance with regulatory, legal, and other requirements.

The Data Governance Institute suggests a framework that organizations, including HCOs, can use to implement and maintain effective governance. A data governance framework should:

- Possess a mission.
- Define focused goals, governance metrics, and success measures.
- Outline clear data rules and definitions.
- Enable decision rights, accountabilities, and control mechanisms.
- Identify data stakeholders and data stewards.
- Establish a data governance office.
- Implement proactive, reactive, and ongoing data governance processes.

The key responsibilities of the data governance function within an HCO are to establish, enforce, and refine the policies and procedures for managing data at the enterprise level. Whether data governance is its own committee or a function of an existing body, the data governance function sets the ground rules for establishing and maintaining data quality, how and under what circumstances changes to data definitions or context can occur, and what constitutes appropriate use of healthcare data.

Based on input from the data owners, data stewards, analytics stakeholders, and business representatives, the data governance committee must create policies and procedures regarding how the data resources of an organization are managed. That is, the data governance function determines

Data Governance

Data governance is "a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods."

under what circumstances the data definitions, business rules, or structure can be changed. This helps prevent an unauthorized local change to a source system causing downstream data quality issues.

The data governance committee is the ultimate authority on how data is managed throughout the enterprise. Organizations without strong and effective data governance structures will likely experience major problems as changes to processes, source systems, business rules, indicators, and even the interpretation of data start to evolve, or change dramatically, without any coordination or impact analysis. Strong data governance helps to ensure that changes ranging from a clinical process to a business rule is evaluated for impact on all other components of the data system.

The personnel structure around data governance should data owners, key stakeholders (including senior and/or executive-level representation), and data stewards from across functional areas. Finally, data governance processes need to be proactive, effective, and ongoing. One of the benefits of the data governance function is that it helps ensure that the source data, and resultant reports, analytics, and insight, are held as trustworthy and valuable within the HCO.

Benefits of Data Governance

One of the benefits of the data governance function is that it helps ensure that the source data, and resultant reports, analytics, and insight are held as trustworthy and valuable within the HCO.

A data governance committee or function within an HCO has a key role in ensuring the integrity of analytics. Decisions are being made more often within HCOs that require both real-time and non-real-time but mission-critical data. When decision makers cannot afford to be wrong, neither can the data; the trust in an HCO's data must be rock-solid. Achieving this high level of trust in data is a key objective of data governance.

I have seen the impact of poor and/or nonexistent enterprise-wide data governance within an HCO. When data quality and management are left to the business intelligence and/or analytics team to manage and "enforce" without any real authority, changes made in one place (say, for example, in a process on the front line, or on a data field in a computer system) likely will not consistently or reliably get communicated to the people responsible for the data. Very often, these changes are not discovered until it is too late and manifest as errors and unexpected results in reports, dashboards, and other analytical tools. Changes in frontline processes or in the way that source system software is used should not first show up as data quality

issues in reports and dashboards because the analytics team was not notified that these changes were being implemented.

What data governance should *not* be, however, is just another layer of bureaucracy. Many HCOs have too many layers of approval required for tasks ranging from changing the design of forms on clinical systems to accessing data in a testing environment. Committees and layers of approval are not necessarily a bad thing—only when they hinder the agility of the organization to respond to actual operational needs.

Data Stewardship

As mentioned earlier, a necessary counterpart to a data governance function within the HCO is the *data steward*. Data stewardship is a necessary component of data governance to ensure high-quality and highly reliable data. The data steward is responsible for monitoring and evaluating data quality within an HCO. Specifically, the major functions associated with a data steward include:[13]

- Evaluating data quality, identifying issues, and making appropriate recommendations.
- Ensuring that any modifications to data storage and management are in line with accepted policies and procedures.
- Ensuring that data is used properly and that it is accessible.
- Helping to establish enterprise-wide standards for data quality and usage.

Within a large organization such as an HCO, the data stewardship function requires one data steward for each major data subject area or functional area.[14] In a typical HCO, this would be achieved by having assigning one data steward for each major line of business, program, and/or domain within the HCO. In a hospital for example, a data steward would be assigned for emergency medicine, surgery, cardiology, and other such functional programs.

Despite the necessity of multiple data stewards, the data stewards of each functional data set must work together and in concert with an organizational data architect to ensure that common standards and approaches are taken. This is especially important for analytics, as program and department indicators and metrics are shared throughout the organization.

The data steward works at the intersection of the business and the technology. Therefore, the data steward should have good technical skills, including knowledge of data modeling and data warehouse concepts. The data steward should also understand the business well. This is not to say that the data steward must be a clinician, but he or she must be familiar with the processes, terminology, and data required by the line of business. Finally, the data steward must possess the interpersonal and communication skills

to be able to bridge the gap in discussions between technology experts and clinical and subject matter experts from the business.

The importance of effective data stewardship cannot be understated. As mentioned, accurate output from analytical systems depends absolutely on the quality of the data that serves as input. Healthcare information technology systems are still relatively immature compared to other industries, and in my experience still undergo significant changes as HCOs evolve through their adoption of HIT. Analytics teams must work very closely with data stewards (within the guidance of the data governance function) to help ensure that when computer systems must be updated or otherwise changed, any and all impacts to the data and defined business rules are understood and mitigated.

Enterprise-wide Visibility and Opportunity

Important decisions in healthcare are becoming less localized and are taking on more of an enterprise scope. Despite this, many factions within HCOs are incredibly reluctant to relinquish control of their data, or even to share it. However, as clinical systems and the data warehouses on which information is stored become more complex, the fact is that data ownership,

THE NEED FOR AGILITY AND DATA STEWARDSHIP

The capabilities of health information technology systems are improving rapidly. For example, the emergency department information system at a typical emergency department when deployed several years ago may have started as only a patient tracking system for patients within the department. Additional capabilities can be added on over time, such as clinical documentation, computerized provider order entry, and results reporting. Other changes also occur throughout the organization that must be accounted for, such as renovations and/or expansions (which require updates to locations and business rules). Other new technology, such as radio frequency identification, might also be added. The challenge is that many of these changes can, and most often do, impact the underlying data in the system. The data stewards within an HCO must be responsible for ensuring that the data, whether it be in an enterprise data warehouse or otherwise, is up to date with the latest data definitions, processes, and business rules. If any necessary changes are not communicated to the developers of analytical tools within the HCO, reporting and analytical insight will soon lose validity, and ultimately value, within the HCO.

stewardship, and management must become a shared responsibility among all data owners. The days of a department or unit owning its own stand-alone clinical or administrative database are numbered. HCOs must work diligently to ensure the availability and trustworthiness of the enterprise-wide data and information that decision makers require.

This shared responsibility can open up whole new opportunities for HCOs to improve transparency and break down silos that have traditionally existed and that have always erected roadblocks in the efficient flow of both patients and information. As more clinical and other data become available throughout the enterprise, the opportunities for enterprise-wide quality and performance monitoring and insight are truly exciting. Provided that the responsibilities of data governance and stewardship are taken seriously throughout the HCO, healthcare departments and programs may no longer need to work to improve quality and performance in isolation.

Notes

1. Thomas C. Redman, *Data Quality for the Information Age* (Boston: Artech House), 42–43.
2. Steve Hoberman, *Data Modeling Made Simple*, 2nd ed. (Bradley Beach, NJ: Technics Publications, 2009), 36.
3. Jack E. Myers, "Data Modeling for Healthcare Systems Integration: Use of the MetaModel," www.metadata.com/whitepapers/myers1.pdf.
4. Ken Terry, "EHR Adoption: U.S. Remains the Slow Poke," InformationWeek.com, November 15, 2012, www.informationweek.com/healthcare/electronic-medical-records/ehr-adoption-us-remains-the-slow-poke/240142152.
5. Ibid.
6. Mike Cottle et al., *Transforming Health Care through Big Data: Strategies for Leveraging Big Data in the Health Care Industry* (New York: Institute for Health Technology Transformation, 2013), http://ihealthtran.com/iHT2_BigData_2013 .pdf.
7. Sid Adelman, Larissa Moss, and Majid Abai, *Data Strategy* (Upper Saddle River, NJ: Addison-Wesley, 2005), 148–151.
8. Canadian Institute for Health Information, *The CIHI Data Quality Framework*, 2009, www.cihi.ca/CIHI-ext-portal/pdf/internet/DATA_QUALITY_FRAMEWORK_ 2009_EN.
9. "6 Key Data Quality Dimensions," MelissaData.com, www.melissadata.com/ enews/articles/1007/2.htm.
10. Joseph J. Juran and A. Blanton Godfrey, eds., *Juran's Quality Handbook*, 5th ed. (New York: McGraw Hill, 1999), 34.9.
11. Yang W. Lee et al., *Journey to Data Quality* (Cambridge, MA: MIT Press, 2006), 80.

12. Data Governance Institute, *The DGI Data Governance Framework*, www
.datagovernance.com/dgi_framework.pdf.

13. Laura B. Madsen, *Healthcare Business Intelligence: A Guide to Empowering Successful Data Reporting and Analytics* (Hoboken, NJ: John Wiley & Sons, 2012), 47–54.

14. Claudia Imhoff, "Data Stewardship: Process for Achieving Data Integrity," *Data Administration Newsletter*, September 1, 1997, www.tdan.com/view-articles/4196.