Measuring and Improving Data Quality
Part IV: Conclusions and Specific Suggestions

Written by Mike Martin DVM, MPH, Clemson University

The first three parts of this paper gave a brief introduction to the science of data quality improvement. This approach has drawn heavily from the field of manufacturing quality improvement, which uses three basic steps. First, the definition of quality is based upon customer requirements. Second, specific measurement techniques are developed to monitor quality on each of the customer-defined attributes. And finally, the results of these measurements are used to drive and monitor a quality improvement cycle based on root-cause analysis and process improvement. These conclusions about data quality improvement within Veterinary Services (VS) include each of these three steps.

Conclusions:

The first article in this series gave examples of data quality, or the lack thereof. Quality is defined as the ability of the data to meet the requirements of all internal and external customers. The first step toward data quality is understanding who these customers are and what they require of the data. VS is very program oriented. Therefore, there is a natural tendency for data collection activities to focus on immediate program management needs. On a strategic level, however, secondary uses such as national surveillance may actually be more important than the primary application. Secondary uses of the data in other programs will increase the intensity of data usage, which, as discussed in part three and below, is a key factor in data quality improvement. These secondary uses will be possible only if the data structure and quality are defined in such a manner that the data are appropriate for those uses.

The second part of this paper addressed the measurement of data quality. An essential step in any quality improvement program is the development of consistent, reliable measures of each quality attribute important to one or more customers. The design of these measures must ensure that they accurately reflect quality, as perceived by the customer. Space limitations did not allow for a detailed discussion of the analysis of these measures. However, the mathematical steps involved in establishing ranges of "normal" and monitoring for abnormal events or trends is so similar to population health monitoring that VS should already have the necessary expertise. Only when quality attributes are defined and measured can process improvement begin. Without these steps, any changes made are nothing more than "shots in the dark." Measurement makes it possible to assess the effectiveness of each change and then either make the change permanent or move on to other efforts based on evidence of effectiveness.

Part three discussed general principles of data quality improvement. Quality data result from quality data production processes, which should be based on sound engineering principles. The continuous feedback loop is the most important concept. The quality of the data needs to be reported back to the source in a meaningful way, just as the result of a pilot's input to the ailerons is fed back to him on an artificial horizon instrument. The best way to provide this feedback is through intensive data use and re-use. Any data re-entry represents a lost opportunity to create a feedback loop.

Recommendations:

Establishment of requirements for data collection and data quality is an essential part of strategic planning by senior leadership. Everything VS does ultimately depends on data quality. Surveillance is only as rapid as the data are timely, and disease incidence regionalization can be no more precise than the incidence reporting upon which it is based. Only the senior leadership is in a position to understand the totality of these dependencies and to define the global requirements for data quality. Workers in the field, whether at regional or area offices, have a natural and appropriate tendency to focus on the task or program at hand. By defining and communicating the importance of secondary data applications and the resulting quality requirements, senior leaders can lead the rest of the data quality improvement process in the right direction.

Establishing positions for data quality managers will greatly facilitate data quality improvement, but only if the mangers are used appropriately. A data quality manager's job should not be to try to "inspect quality into the data."[1] Rather, he or she should establish and monitor metrics of quality on each attribute—as defined by senior leadership—and facilitate process improvement. Perhaps the most challenging aspect of a data quality manager's job will be to follow W. Edwards Deming's eighth point, "Drive out fear."[2] Prior to Deming, the traditional approach was to find a mistake, find the person who made the mistake, and reprimand or retrain that person. This strategy of treating the immediate cause has proven ineffective in every application. Instead, a strategy should be implemented that identifies the process that is not producing adequate quality, builds a team tasked with improving the process, and applies process engineering to bring it back under control. For this to be effective, the data quality manager must be seen as a facilitator, not an adversary.

Clear definitions of quality attributes and the systems that measure them will form the essential foundation for data quality improvement. Application of the engineering principles discussed in part three will also be facilitated by a number of recent improvements in the VS' information infrastructure. Improved networking infrastructure now provides the ability to enter a piece of data once and immediately begin re-use, not only in the local area, but throughout the system. Feedback can now take place in near real time. The Emergency Management and Response System, National Animal Health Laboratory Network messaging, BSE Web-based system, and the Standardized Premises Registration System are all examples of applications where data re-entry has been replaced by data re-use. These are all obviously first efforts and significant additional work is needed to continue improving interoperability and re-usability of all data elements. The direction, however, is clearly positive.

This series of articles has given only a broad overview of the field of data quality improvement. Readers with an interest in data quality are encouraged to read the many excellent primary and secondary sources available, many of which have been referenced in the endnotes of each article.

Please contact Dr. Michael Martin at mmarti5@CLEMSON.EDU with any questions or comments pertaining to Data Quality.

---

[1] Deming WE, *Out of the Crisis*, MIT Center for Advanced Engineering Study, 1982, pp 28-31.
[2] Walton M, *The Deming Management Method*, Perigee Books, 1986, pp 72-73.