DELMARVA FOUNDATION

Human Factors and the Design of Everyday Data Collection Tools

Authors: Howard Townsend, PhD, and Sandra Lesikar, PhD

June 2004

INSIGHTS INSIGHTS INSIGHTS INSIGHTS INSIGHTS

INSIGHTS

Human Factors and the Design of Everyday Data Collection Tools

Executive Summary

An essential part of a successful healthcare quality improvement culture is the continuous improvement of the tools used for measuring quality. In this paper, we discuss the quality of different data collection systems; the use of a human factors approach for improving the quality of data collection tools, with the Resident Assessment Instrument-Minimum Data Set (RAI-MDS) as an example; and how this approach may be applied to other healthcare data collection systems.

The RAI-MDS is an example of a tool in which thorough data are collected that enable us to view a comprehensive picture of healthcare quality. The development of similar tools and widespread use for additional healthcare settings (other than nursing facilities) would be a boon to the advancement of healthcare quality. Along with development, efforts to continuously improve the quality of such tools are an essential aspect for continuous improvement of healthcare quality.

To advance a quality improvement culture in healthcare, tools that ensure thoroughness, accuracy, and reliability are necessary. The human factors approach to the development of data collection tools that can be used to improve detection and, ultimately, prevention of errors in data thus prevents erroneous information from being used to measure healthcare quality. By reducing the error in healthcare data through considerations of the design of data collection tools, the "picture of healthcare quality" becomes clearer, quality of care can be measured more accurately, interventions can be designed more effectively, and resulting improvements can be estimated with greater confidence.

Introduction

Human culture has made advances through time not because of biological or physical advances in the composition of a human, but through the development and refinement of everyday tools that serve as extensions of our biological and physical composition. Groups that have continued to refine tools and invent new ones have become the predominant cultures on the earth, whereas groups that have retained the use of Stone-Age tools have gone extinct or remain in small, isolated areas of the globe. If a healthcare quality improvement culture is to proliferate, then the tools of improvement should be continuously refined. An essential part of a successful healthcare quality improvement culture is the continuous improvement of the tools used for measuring quality.

In virtually all quality improvement paradigms—from Baldrige to Six Sigma—measurement of quality is considered to be a critical component of quality improvement. However, in the arena of healthcare quality improvement, the accuracy and completeness of data for measuring quality are not always given the requisite attention. As public reporting of quality measures and the use of provider incentives to improve quality become more common, the use of accurate, reliable data for calculating results and indicators is imperative, and standardized data collection systems used in similar settings should be the norm. In this paper, we discuss the quality of different data collection systems and the use of a human factors approach for improving the quality of data collection tools.

Largely, administrative data—data used for payment of providers (e.g., claims, encounters, and enrollment data) from a variety of settings (inpatient, outpatient, professional offices, etc.)—have been successfully used to measure quality (Asch, Sloss, Hogan, Brooks, & Kravitz, 2001; Black & Roos, 1998; Goldfield & Villani, 1996; Johantgen, Elixhauser, Bali, Goldfarb, & Harris, 1998; Schwartz, Gagnon, Muri, Zhao, & Kellog, 1999). In many cases, success has hinged on systematic review of data collection systems and verification of data accuracy. Administrative data are an attractive source of information for monitoring healthcare quality because they are 1) readily available from computerized healthcare databases, 2) relatively inexpensive to acquire, and 3) encompass large and diverse populations, lending themselves statistical and interpretive power (*sensu* Iezzoni, 1997). These attributes allow vast pictures of healthcare quality to be developed with broad sweeping strokes. However, often the shortcomings of using administrative data to assess quality are overlooked. Limitations in the scope of data recorded in administrative data systems may prevent the capture of sufficient data to accurately assess quality (Hsu, Go, & Selby, 2001; Keating et al., 2003; Kieszak, Flanders, Kosinski, Shipp, & Karp, 1999; Majoor, Ibrahim, Cicuttini, Boyce, & McNeil, 1999; Schneider, Wiblin, Downs, & O'Donnell, 2001). Although administrative data lack the thoroughness of medical records data, their major advantage is their accessibility.

Medical record abstraction and other data tools based on assessment of patients may provide a more complete view of the quality of care provided to a patient, but these tools are labor intensive, expensive, and introduce more subjectivity in measurement than the use of large administrative data sets. While medical record abstraction data enable a finer-grain view of healthcare quality, they lack the breadth and statistical power of administrative data (Peabody, Luck, Glassman, Dresselhaus, & Lee, 2000; Steinwachs et al., 1998). Lack of proper documentation within medical records is another difficulty that makes medical record abstraction data less comprehensive (Peabody et al., 2000).

The use of these tools separately poses major threats to the continued proliferation of a quality improvement culture. By using administrative data sources (and other tools not directly associated with healthcare quality), the risk of "fuzzy" assessment of quality is high. The perceived fuzziness of measures from administrative data makes quality improvement a hard sell, because changes in quality may be interpreted as being attributable to noise in the system rather than actual changes in quality. Conversely, medical record abstraction, while providing a more detailed picture of quality, may lead to quality being viewed as a parasite on healthcare. Cultures that deal effectively with parasites have a greater chance of successful propagation than those with a high parasite burden.

Ultimately, for a tool to be effective, it should be associated with the measurement of payment (an essential part of a healthcare system) but in a symbiotic relationship. The tool should allow capture of complete patient diagnostic and assessment information, which is necessary for accurate and complete payment for services as well as for developing a finer grain view of the quality of care provided. In most recent quality improvement paradigms, the quality of goods and services is associated with value, i.e., quality improvement is positively correlated with increases in return on investment. Logically, quality of assessment would be positively correlated with measurement of costs and values. A quality improvement culture that is able to construct effective tools will have a higher probability of successful propagation than one that uses Stone-Age tools.

A significant advancement in cost and quality data collection tools is embodied in the Resident Assessment Instrument-Minimum Data Set (RAI-MDS) developed by the Centers for Medicare & Medicaid Services (CMS) and its partners for measuring costs and quality of care for nursing home residents. Developers of the most recent version (version 2.0) of the RAI-MDS have made significant strides in interweaving the measurement of payment and cost-related data with quality of life and care for residents. In addition, the developers have recognized and acted on the need for continuous quality improvement in tool development. It is in this spirit of continuous quality improvement that we have written this paper and use the RAI-MDS as an example of how a human factors approach can be used to improve the design of healthcare data collection tools and ultimately the quality of health and healthcare for every American.

In this paper, we demonstrate how using human factors thinking can aid in minimizing errors in the collection of data for healthcare quality and costs. We outline the human factors approach to understanding human error; demonstrate how human factors and human error may affect the accuracy and effectiveness of the measurement of healthcare quality, using the RAI-MDS as an example; and discuss how this approach may be used for other healthcare data collection systems.

Human Factors Approach to Understanding Error

The human factors approach to understanding error is founded on understanding the underlying psychological functioning and how that makes us likely or unlikely to execute a sequence of actions correctly. Some of the inherent abilities that humans possess make us able to easily deduce information from a busy world full of stimuli, to make sense of this information, and to gather this spotty data and make generalizations. However, these abilities also leave us prone to error making; hence, the familiar phrase—to err is human. These same abilities that have always set us apart, also make it easy for us to see patterns of similarities and differences that may not really exist, to generalize information too rapidly, and to jump to conclusions.

The first delineation of error types is the intention of the person taking action and making the mistake. The clearest explanation of the distinction of slips and mistakes comes from *The Design of Everyday Things* (Norman, 2002, p. 106): “Form an appropriate goal but mess up in the performance, and you’ve made a slip…Form the wrong goal, and you’ve made a mistake.” Slips are generally smaller actions and are easier to discover by monitoring the process or procedure that failed to achieve the intended goal. In contrast, mistakes may be major events, with flawless processes and procedures. Mistakes may be impossible to detect. Table 1 provides a summary of the major error types.

Table 1. Distinctions between error types adapted from Reason (1990)

	Slips and Lapses (Skill-Based Error)	Mistake (Rule-Based Error)	Mistake (Knowledge-Based Error)
Type of Activity	Routine actions	Problem-solving actions	
Focus of Attention	Not on task at hand	Directed at problem-related issues	
Control Mode	Automatic processors (schemata)	Automatic processors (stored rules)	Limited, conscious processes
Predictability	Largely predictable		Variable
Ratio of Error to Opportunity for Error	The absolute number is high, but these types represent a relatively small proportion of the opportunities for error		The ratio is high, but absolute number is low
Influence of Situational Factors	Low to moderate; intrinsic factors (frequency of prior use, i.e., experience) likely to exert dominant influence		Extrinsic factors likely to dominate
Ease of Detection	Usually fairly rapid	Difficult and often only achieved through external intervention	

Slips and Lapses

Errors are either intentional or unintentional. This distinction is important for the detection of errors and results in different types of interventions and corrective actions. Slips and lapses are unintentional errors. They usually result from practiced behaviors overtaking our current actions that unintentionally and subconsciously cause us to veer off from the path that would get us to our intended goal. An example of a slip that occurs in everyday life is all too familiar: we get in the car after work with the intention of stopping at the store on the way home and suddenly find ourselves in our driveway in front of our house. We had set the appropriate goal—of going to the store—but well-practiced routines (i.e., driving home from work) took over our behavior and caused us to make a slip.

Slips are often separated into six categories: capture errors, description errors, data-driven errors, associative activation errors, loss of activation errors, and mode errors. These categories help us to understand the specific underlying cause of the slip and guide us to the appropriate corrective action. The underlying cause for slips is usually a result of a lapse in attention or the similarity of actions.

Capture Errors

Capture errors occur when a well-practiced activity "captures" your attention and takes over the task. The previous example of intending to drive to the store and instead ending up at home is a classic example of a capture error.

Description Errors

The most common type of errors, description errors occur when the mental description of the action to be taken is too vague or when there are two similar response stimuli adjacent to each other. For example, if a box of baby cereal is adjacent to similar box of powdered automatic dishwashing detergent, the mental description of the activity may be to pour one cup of powder from the box on the shelf into the food bowl and add water. In error, we might feed the baby soap powder. The function of the design of many things makes them prone to slips, i.e., long rows of switches and forms with columns of identical check boxes.

Data-Driven Errors

Data-driven errors happen when new information arrives that intrudes on the current actions. For example, if someone is trying to remember a phone number by mentally rehearsing the string of digits and a paging system intrudes and announces another string of digits, he or she can no longer distinguish the phone number from the numbers heard over the paging system.

Associative Activation Errors

Associative activation errors occur when a stimulus requires a similar but not identical action to another stimulus and we mistakenly substitute the wrong, but similar, action. For example, a receptionist for a busy corporation may frequently answer the phone at home with the corporate greeting she uses at work.

Loss of Activation Errors

These errors occur when we forget the purpose of our goal. We have completed part of the sequence of activities but suddenly forget the ultimate goal and are therefore unable to compete the task and fulfill the goal. The classic example of this is walking into a room and forgetting the reason for being there. These are common errors but we are usually able compensate by walking back into the room and "re-activating" our memory of the goal of our original actions.

Mode Errors

Mode errors occur when a single device has multiple modes and the user is required to remember what mode the device is in to operate it properly. Calculators, cameras, computers, watches, and other electronic equipment are excellent examples of devices with features that have multiple modes.

Mistakes

Mistakes are intentional errors. These types of errors result from conscious choices. For example, we get in our car to go home with the intention of arriving as quickly as physically possible, but we find ourselves on the roadside giving our driver's license and registration to a police officer. We set the wrong goal—arriving home as quickly as possible; we should have set the goal of arriving home as quickly as *legally* possible.

Mistakes have been further broken down into two categories: 1) rule-based mistakes and 2) knowledge-based mistakes. Rule-based mistakes arise from the use of rules for an action in which the situation has changed from the typical situation. For example, a person accustomed to making a right turn at a red light (when no traffic is observed to the driver's immediate left in the intersecting lane) may have an accident if he or she continues operating according to that rule in a five-way intersection where traffic may be coming from different lanes. Knowledge-based mistakes result from situational changes for which a person has no experience and has not anticipated the changes. For example, young drivers may fail to reduce speed appropriately while driving on an icy road, because they lack knowledge of how their vehicle handles.

Error Detection and Correction

The capacity of the human brain to handle complex informational tasks is remarkable; however, this capacity is imperfect and as a result, errors are inevitable. To mitigate the consequences of error, resources have to be dedicated to the detection and correction of error. In this section we consider the ways in which slips, lapses, and mistakes can be detected and corrected, and how these can be used to ensure that the design of tools allows for quick and accurate collection of data.

In general, people detect an error in one of three ways: 1) they find it themselves, 2) something in the working environment provides evidence of an error, or 3) someone tells them about the error. The modes of detection range in complexity and the level of effort required to execute them, thus the mode of error detection determines promptness and efficiency of error correction. Conversely, the energy and effort required for different modes of error detection may decrease the efficiency with which a task is accomplished. Thus, to maximize the efficiency with which a task is performed correctly, optimal allocation of effort to different modes of error detection is necessary.

Ideally, to maximize efficiency of data collection tools, to the extent possible, error detection and correction should be relegated to machines, so that an operator's mental workspace can be dedicated to the completion of the task at hand.

Human Factors and the RAI-MDS

Our objective in this paper is to give examples of the types of human error that are likely to be made when administrators and clinicians use the RAI-MDS tools and suggest methods for preventing and detecting such errors. There are different questions to consider when evaluating the set-up of a data collection tool:

1) Is the data collection tool a paper tool or an electronic tool?
2) Are the questions related to easily verifiable information or a subjective opinion or judgment?
3) How specific are the guidelines for responses?

Answers to these questions can guide the detection and prevention of human error in collecting data. This approach may be applied to any data collection tool. The first few sections of the MDS tool (AA, AB, AC, and A) are primarily for the collection of administrative data (name, identification numbers, etc.). Recording of such information would be a routine action and may result in a slip or lapse, such as recording a phone number where a Medicare identification number is required. Two main techniques for error correction and detection could be used for these sections. The first would be for the person completing these sections to double check entries made on the form to ensure that the correct responses were recorded. Of course, double checking may easily become a routine action, in which case it may also be subject to slips and lapses, especially in a chaotic or high-pressure environment. In this case, error detection based on information technology may be desirable. For example, an electronic version of the MDS form may force the value of a Medicare identification number to conform to the specifications of a Medicare identification number. The sooner this type of slip is detected, the more quickly and effectively it can be corrected, thereby reducing the burden and cost of recording this information.

Other sections of the MDS tool are based on clinical assessment of nursing home residents and, in general, require a higher level of cognitive processing than the administrative portion. As a result, these sections are more prone to rule- and knowledge-based errors (i.e., mistakes). In addition, responses to some of the MDS items are based on clinical judgment. To the extent that some of the responses are subjective, error detection may be difficult. Two different individuals may respond to the same MDS item differently, and neither individual is necessarily incorrect.

For example, consider item C6 on the RAI-MDS (residents' ability to understand others). The response options for this item are "understands," "usually understands," "sometimes understands," and "rarely/never understands." Clinician A may respond that the resident "usually understands," whereas Clinician B may respond that the resident "sometimes understands." The absolute truth about how well a resident understands lies within the resident, who may not be able to express how well he or she understands. Thus the "correct" response to item C6 cannot be known and the extent to which this item has been answered erroneously cannot be identified.

Although this type of error is difficult to detect, data collection tools can be designed to reduce errors attributable to differences in clinical judgment. Essentially, these are rule-based errors. In the above example, the two clinicians may have developed internal rules for making a determination on item C6, i.e., they may have internal definitions for the codes for the item, but their rules might differ, as shown by Table 2. Because both of the examples of internal rules are consistent with the descriptions for coding this item, both clinicians' responses may be considered accurate. However, because their responses differ, neither response can be considered reliable. The fault of this inconsistency does not necessarily lie with the clinicians, but rather is a result of a shortcoming in the tool design.

Table 2. Examples of internal rules for coding MDS item C6 (residents' ability to understand others)

Code	Description	Definition (Internal Rule)	
		Clinician A	Clinician B
0	Understands	Resident rarely stares blankly in response to my questions or instructions (i.e., never more than 5% of the time during an interaction)	Resident never stares blankly in response to my questions or instructions (i.e., 0% of the time during an interaction)
1	*Usually understands*—may miss some part/intent of message	Resident occasionally stares blankly in response to my questions or instructions (i.e., 5–50% of the time during an interaction)	Resident occasionally stares blankly in response to my questions or instructions (i.e., 1–40% of the time during an interaction)

Code	Description	Definition (Internal Rule)	
		Clinician A	Clinician B
2	*Sometimes understands–* responds adequately to simple, direct communication	Resident frequently stares blankly in response to my questions or instructions (i.e., 51–89% of the time during an interaction)	Resident frequently stares blankly in response to my questions or instructions (i.e., 41–94% of the time during an interaction)
3	*Rarely/Never understands*	Resident always stares blankly in response to my questions or instructions (i.e., more than 90% of the time during an interaction)	Resident always stares blankly in response to my questions or instructions (i.e., more than 95% of the time during an interaction)

To improve the design, item responses or codes should be based on observable, measurable data. As much as possible, objective rather than subjective data should be used for determining costs and assessing quality of healthcare, so that information is consistent and reliable. In the example of differing judgment on item C6, the tool could be improved by providing specific, quantifiable definitions for each code to standardize internal rules used by clinicians.

In addition to being prone to rule-based errors, MDS items that depend on clinical or professional judgment are prone to knowledge-based errors. Knowledge-based errors, like rule-based errors, are difficult to detect, and, in some circumstances, may be detected but disregarded, because the detector is assumed to be knowledgeable. When judgment is necessary for determining correct responses to data collection tools, then error detection becomes more difficult. Although no foolproof methods for detecting knowledge-based errors are available, mechanisms for detecting some knowledge-based errors can be implemented. Logical consistency checks (i.e., edits) may be used in an electronic tool to detect possible error. For example, if an MDS assessment is coded that the patient is comatose (item B1) but also independent in all activities of daily living (section G—walking, dressing, eating, etc.) this is a clinical logical inconsistency. An "edit" could be programmed into the software tool that collects that data to warn the user that a logical inconsistency exists, and recommend re-checking items involved or re-evaluating the patient.

This type of error detection may also prove useful in determining whether errors are truly the result of a misapplied rule (or a lack of knowledge) or if the intent of the person making the error is to "game the system." A pattern of misapplications of rules for data tool use in which the mistake consistently favors the tool user (whether by increasing financial gains or "painting a better picture" of the quality of care provided) may be indicative of attempts to abuse the system.

Conclusions

The RAI-MDS is an example of a tool in which thorough data are collected that enable us to view a comprehensive picture of healthcare quality. The development of similar tools (and widespread use of such tools) for additional healthcare settings (other than nursing facilities) would be a boon to the advancement of healthcare quality. Along with the development of such tools, efforts to continuously improve the quality of such tools are an essential aspect for continuous improvement of healthcare quality.

To advance a quality improvement culture in healthcare, tools that ensure thoroughness, accuracy, and reliability are necessary. The human factors approach to the development of data collection tools can be used to improve detection and, ultimately, prevention of errors in data—thus preventing erroneous information from being used to measure healthcare quality. By reducing the error in healthcare data through considerations of the design of data collection tools, the "picture of healthcare quality" becomes clearer, quality of care can be measured more accurately, interventions can be designed more effectively, and resulting improvements can be estimated with greater confidence.

References

Asch, S. M., Sloss, E. M., Hogan, C., Brooks, R. H., & Kravitz, R. L. (2001). Measuring underuse of necessary care among elderly Medicare beneficiaries using inpatient and outpatient claims. *Journal of the American Medical Association, 284,* 2325–2333.

Black, C., & Roos, N. P. (1998). Administrative data: Baby or bathwater? *Medical Care, 36,* 3–5.

Goldfield, N., & Villani, J. (1996). The use of administrative data as the first step in the continuous quality improvement process. *American Journal of Medical Quality, 11*(Suppl. 1), S35–S38.

Hsu, J., Go, A., & Selby, J. (2001). Overuse of administrative data to measure underuse of care. *Journal of the American Medical Association, 285,* 735.

Iezzoni, L. I. (1997). Assessing quality using administrative data. *Annals of Internal Medicine, 27,* 666–674.

Johantgen, M., Elixhauser, A., Bali, J. K., Goldfarb, M., & Harris, D. R. (1998). Quality indicators using hospital discharge data: State and national applications. *Joint Committee Journal on Quality Improvement, 24,* 88–105.

Keating, N. L., Landrum, M. B., Landon, B. E., Ayanian, J. Z., Borbas, C., & Guadagnoli, E. (2003). Measuring the quality of diabetes care using administrative data: Is there a bias? *Health Services Research, 38,* 1529–1545.

Kieszak, S. M., Flanders, W. D., Kosinski, A. S., Shipp, C. C., & Karp, H. (1999). A comparison of the Charlson comorbidity index derived from medical record data and administrative billing data. *Journal of Clinical Epidemiology, 52,* 137–142.

Majoor, J. W., Ibrahim, J. E., Cicuttini, F. M., Boyce, N. W., & McNeil, J. J. (1999). The extraction of quality-of-care clinical indicators from state health department administrative databases. *Medical Journal of Australia, 170,* 420–424.

Norman, D. A. (2002). *The design of everyday things.* New York: Basic Books.

Peabody, J. W., Luck, J., Glassman, P., Dresselhaus, T. R., & Lee, M. (2000). Comparison of vignettes, standardized patients, and chart abstraction: A prospective validation study of 3 methods for measuring quality. *Journal of the American Medical Association, 283,* 1715–1722.

Reason, J. (1990). *Human error*. Cambridge: Cambridge University Press.

Schneider, K. M., Wiblin, R. T., Downs, K. S., & O'Donnell, R. E. (2001). Methods for evaluating the provision of well child care. *Joint Committee Journal on Quality Improvement*, *27,* 673–682.

Schwartz, R. M., Gagnon, D. E., Muri, J. H., Zhao, Q. R., & Kellog, R. (1999). Administrative data for quality improvement. *Pediatrics, 103*, 291–301.

Steinwachs, D. M., Stuart, M. E., Scholle, S., Starfield, B., Fox, M. H., & Weiner, J. P. (1998). A comparison of ambulatory Medicaid claims to medical records: A reliability assessment. *American Journal of Medical Quality*, *13*, 63–69.