Interactively Visualizing Data Warehouses

Yoshitaka Bito, MS; Robert Kero, MS; Hitoshi Matsuo, MS; Yoichi Shintani, PhD; Michael Silver, MD, FACP, FCCP, FCCM

ABSTRACT
As healthcare costs continue to rise, along with the urgency to control these costs, timely analyses of operational data stored in data warehouses have become critical. Although healthcare organizations have begun to amass large volumes of raw data within data warehouses, many have yet to capitalize on that valuable information. Many clinical and financial events are recorded each day, yet few of these data are leveraged to increase organizational awareness and performance. Because the development and implementation of a data warehouse requires significant time and capital, as well as highly skilled labor, the need to more fully leverage this considerable investment demands techniques that facilitate detailed data exploration, analysis, and the subsequent communication of findings. This article describes how the effective display of complex relationships in data can be used to discover areas of variance in large, changing data warehouses.

KEYWORDS
• Visualization
• Interactive
• Care process
• Data analysis
• Multidimensional data
• Critical pathway
• Communication

Increases in the intensity of services over the past thirty years have helped to drive healthcare costs to all-time highs.¹ Health economics literature suggests that a large portion of this healthcare activity is superfluous.²

Note: The authors would like to thank Patricia Skarulis, chief information officer at Rush-Presbyterian–St. Luke's Medical Center, for her continued direction and support of the Hitachi/Rush Enterprise Data Warehousing Research Project.
Although physician-based practice reporting can be an effective tool to encourage more resource-efficient decision making, the focus on care process change must be supported to truly benefit cost-containment and reduction efforts. One way to help meet these objectives is to develop and use critical pathways. These are “best practice models or guidelines” used to guide clinicians toward appropriate, cost-effective, and consistent treatments of homogeneous patient groups. Variance analysis can, in turn, be used to find the differences between the planned care process for a homogeneous patient group and the actual care processes followed for individual patients. Insights gained through variance analysis can subsequently be used to refine existing critical pathways. The concept of variance analysis is also effective, even if critical pathways have not been stipulated. In this case, the word variance means the difference between the average or standard care process for a homogeneous patient group and the actual care process followed for an individual patient.

Efforts to develop, monitor, and manage critical pathways are expensive because of the clinical and administrative participation required. Although the potential return from fully implemented pathways can be high, healthcare provider organizations (HCOs) find it difficult to bear the costs (time and labor) necessary to fully develop and continuously monitor a comprehensive suite of pathways. Hence, a subset of diagnoses and procedures is normally chosen, based on specific criteria (for example, high-volume, high-cost, high-variability procedures); associated pathways are then developed in as much detail as possible.

Even before an organization can develop, implement, and benefit from pathways, it must first determine which care factors are both influential and controllable. The timely analysis of disparate healthcare data sources, integrated within a data warehouse, greatly supports the pathway development process. Advanced visualization technologies can further support this process by efficiently identifying care process variations. To present an example of this complementary approach, a subpopulation of homogeneous surgical patient encounters will be used to model surgeon care processes via time-stamped, billable service events.

The interactive visualization technique to be introduced uses two cooperative methods to handle a three-dimensional projection of patient, service, and time: (1) a one-dimensional bar chart to show the aggregate values (for example, quantity or cost) of services by a surgeon and (2) a two-dimensional matrix to show a time-based pattern of services. The pattern is represented on the matrix using a coordinate (time, service) and gray level to correspond to the coordinate value (for example, difference between two surgeon population averages). The interactive visualization method provides users with a synchronized means to identify and understand variance. The identification of clinical efficiencies, cost reductions, best practices, and the purpose of knowledge sharing are all supported by this approach.
Social Structures

HCOs clinically partition their areas of practice and assign the responsibilities for each area to specific clinical groups. When the composition of these groups is fairly homogeneous, they tend to share a basic set of requirements, assumptions, and motivations. Over time, these groups tend to develop normative structures that create communities-of-practice. These communities-of-practice usually evolve in a way that provides increased organizational efficiency and effectiveness, as well as a specific set of abilities, expertise, and expectations. Complex business practices and processes defining how an organization conducts the business of care are the results. Therefore, a data warehouse must be used to provide direct feedback to both clinicians and their associated communities-of-practice to improve overall awareness and future performance.

Traditionally, physician performance has been measured by way of profiling. The primary objective sought through the use of profiling has been the reduction or control of utilization and costs associated with the delivery of care. Comparative physician analyses were generated, usually with financial data, to pressure physicians to meet projected results. Because relevant clinical data tend not to be used to profile care delivery, physicians end up spending considerable time explaining why subsequent profiles are meaningless or unrelated to the quality of care provided. Current physician profiling efforts often become adversarial, failing to produce the impact that HCOs have hoped for. Physicians are not provided with the types of actionable information needed to support true self-assessment and correction. Physician-driven performance management approaches must begin to replace previous profiling methods. The communities-of-practice must become primary stakeholders in overall performance management and in continuous learning processes.

Analysis Process

Traditional analysis technologies do very little to provide clinicians with an effective contextual understanding of the data being analyzed. The burden of creating cognitive problem representations is normally left up to each user, as the design and development of technologies to support this process are difficult. An interactive visualization technique, through the use of new display paradigms, was developed to support the analysis of difficult cognitive problem representations, for example, of care processes. Because the explicit graphical representation of care processes can be easily displayed and manipulated with this new technique, analysis tasks previously left to decision makers require considerably less cognitive effort to be accomplished. In contrast, if tabular-based textual representations were used, decision makers would require significant cognitive efforts to arrive at the same decisions, as they would be forced to rely on their own visualization capacities to perform these tasks.9
Although care process analyses are believed to be effective in identifying ways to improve care and reduce costs, these types of analyses are difficult because they rely on extremely detailed data. Hence, the complexity of care process analyses increases according to the level of detail represented. For example, complexity differs relative to the number of distinct service classes used and time resolution applied to each rendered service. The difficulty of these analyses, however, is not only based on the volumes of data but on the complex sets of data relationships to be handled. Care process data are inherently multidimensional because they include information pertaining to who rendered services, when and what services were rendered, number of services applied, and so on. Because subsequent analyses must be able to find characteristics and outliers from this multidimensional data, the commonly used one- or two-dimensional graphing approaches are not sufficient. These traditional approaches fall short for this purpose due to their lack of ability to display relationships among multidimensional axes and to slice and categorize data based on attributive information. Applications that employ multidimensional databases enable the handling of these data to some extent—for example, drilling down to detailed information and categorizing according to attributive information. However, most of those applications use one- or two-dimensional graphing representations that could not efficiently show the temporal pattern of the care process. Although another technique has been proposed to handle multidimensional data, it cannot efficiently visualize the detail of the care process.

**Approaches to Progressive Care Process Analysis**

This section introduces the limitations of traditional analysis approaches and the reasons behind the development of the interactive visualization method. To better explain the effectiveness of each approach, we describe how they were used to perform inpatient care process analyses.

The first, simplest approach was to display a list of all services applied to specific patients using only textual representations. Although we could identify how physicians took care of patients to some extent, it was extremely difficult to show the process clearly. For example, it was difficult to identify relationships among services or to know whether specific services were used when employing the list. Additional disadvantages were that groups of patients could not be handled, rendering it impossible to see the average care process for specific patient groups, compare care processes applied to patients, and identify outliers. This approach forced the authors to rely significantly on their own visualization capacities to perform the analyses.

The second approach concentrated on the visualization of service intensities through the use of standard graphs, for example, total and average cost of each service provided. Although standard graphs could display totals and ratios of service utilization, they could not show the timing of rendered services
across care processes. This lack of temporal representation caused the authors to exert considerable effort to perform analyses.

The third approach was to visualize both the timing and type of services. We used a two-dimensional color matrix (2D-matrix) of which dimensions were “elapsed days from admission” and “rendered service types”; color represented the quantities or costs of the rendered services. A subgroup of inpatients was preselected, for example, by diagnosis-related group (DRG), and the aggregated values were displayed on the 2D-matrix. Although this approach could provide for standard care processes of inpatient subgroups, we realized that we were more interested in specific outliers. Our analyses started to identify patients who were provided nonstandard services; we, in turn, became interested in the reasons behind these occurrences. This approach, however, did not provide a means to dynamically select patients for further analyses nor did it provide a way to compare rendered services quantitatively. It was also difficult to compare the two values represented by two colors on the matrix.

We discovered five important points as we applied these limited approaches:

• Patients, timing, and types of services are the three most important factors. Attributive information of patients and services (for example, the attending doctor of the patient or the cost center of the service) were also important to support analyses.
• The categorization of patients and the ability to drill down to individual patient levels must be supported.
• Qualitative visualization (a 2D-matrix) is an effective way to interpret characteristics and identify outliers. However, qualitative visualization was not sufficient, so quantitative visualization must be employed to interpret details.
• Patients and services should be categorized dynamically, which means patients and services should be categorized not only by their attributive information but on the data as well.
• Patients and services should be selected dynamically, which means the display of patients and services should be restricted, based on both their attributive information and on the data.

Based on these points we have proposed a new care process analysis method using an interactive visualization technique.

Interactive Visualization Technique

The care process analysis method supported by the new interactive visualization technique utilizes three-dimensional data representations: (1) patients, (2) elapsed days from admissions, and (3) service types. Figure 1 shows how the method works on three-dimensional data.
The technique uses two graphs—one a two-dimensional color matrix (2D-matrix) and the other a one-dimensional bar-chart graph (1D-graph). The two dimensions of the 2D-matrix represent the days from admissions and the service types. The three-dimensional data are projected onto the matrix with aggregated values (for example, total or average quantity and cost) displayed using colors or gray levels. The one dimension of the 1D-graph represents patients, and as three-dimensional data are projected onto the 1D-graph, aggregated values represented by the height of the graph are calculated. In order to dynamically slice and categorize data, the method uses a new interactive visualization technique that allows users to select regions of interest (ROI) on either graph and then observe as the other graph is dynamically updated using data associated with the ROIs. That is, when users select groups of patients via the 1D-graph, the care process patterns for the groups are then displayed on the 2D-matrix. Contrarily, when users select various services on the 2D-matrix, the amounts (for example, quantities or costs) associated with these selected services are then displayed on the 1D-graph. This bidirectional functionality provides interactive advantages through visual synchronization capabilities. Upon direct selection of a ROI, users can observe restricted data redisplays on the opposing graph. The 2D-matrix has an advantage by displaying qualitative
patterns and correlations between the selected two dimensions. The 1D-graph has an advantage by quantitatively comparing the selected one dimension.

**Visualization**

We have developed a prototype analysis tool based on this interactive visualization technique. Furthermore, we have implemented additional features to improve the tool, for example, sorting and categorizing items on the dimension in order to systematically observe the data, providing a double 2D-matrix to simultaneously compare the pattern of two patient groups, and so on.

The prototype discussed is now being used on a data warehouse containing over one hundred thousand inpatient encounter records, with over ten million associated billable service items. We have implemented procedures to easily create three-dimensional care process data for specific DRGs. Additionally, Web pages have been implemented connecting the data warehouse and client prototype software. The Web pages begin the procedures on the data warehouse, retrieve resultant care process data, and ship them to the client machine via the Internet. This process enables users to begin an analysis by just accessing specific Web pages.

**Care Process Analysis Using Interactive Visualization**

We have applied this tool to care process analyses for various DRGs. Figure 2 shows an example of its application to a synthetically generated surgical DRG population.

The left window shows the double 2D-matrix; the right upper window shows a control panel to control appearance of the graphs; and the right lower window shows the 1D-graph. Although Figure 2 is displayed in black and white, the 2D-matrix and the 1D-graph normally use color to direct attention. The ROIs on the graphs are also drawn with colors. For example, the rectangles on the 2D-matrix representing the ROIs are drawn using red and green, and the corresponding bar charts on the 1D-graph are drawn using the same colors, respectively. The dimension of the 1D-graph represents patients and is categorized by surgeon to estimate the service intensity variance between surgeons. The bar chart shows the average cost of surgeons. The leftmost bar represents the average cost of all services, and the rightmost two bars represent the average costs of the services found in the two ROIs selected on the 2D-matrix. The meaning of the two ROIs on the 2D-matrix is described later. Two ROI rectangles are selected on the 1D-graph, one of which included the leftmost surgeon; the other included the remaining three surgeons.

The reason these two ROIs were originally selected is that the average cost of the leftmost surgeon is considerably higher than the other three surgeons. Our analyses focused on the identification of where service variation occurred between the two groups throughout their care processes. The double
2D-matrix displays the two temporal-based average care processes related to these two surgeon groups. The upper window corresponds to the leftmost surgeon, and the lower window corresponds to the three others. The horizontal axis represents the elapsed days from the admissions, and the vertical axis represents the applied service types. The vertical axis is sorted by the amount of average cost, which means the top row is the service type that is most frequently used for this DRG. The two ROIs are selected on the 2D-matrix: one consists of the top two cost-consuming service types and the other consists of the third through ninth service types. The rightmost two bars of the 1D-graph show the cost fractions of these aggregated service groups.

The 2D-matrix clearly shows a difference in the service utilization time patterns for the average care processes of the two surgeon groups. The 1D-graph shows the difference of the average costs among the surgeons (left bar), the similarity of the usage of the top two service types (middle bar), and the difference of the usage of the third through ninth cost-consuming service types (right bar). The difference of the total costs associated with the two groups was not highly correlated to the usage of the top two service types. Instead, the difference is mostly attributed to the usage of the third through ninth cost-consuming
service types. Further analysis could show the difference in timing, duration, and amount of various services rendered. This tool could easily be used to find average care processes, the difference between average care processes, and the dominant service types attributing to such differences. Furthermore, the tool could be used to visualize additional attributive information.

We believe that this interactive visualization technique can be easily adapted to many other analyses, especially ones that use multidimensional data and need to see correlations between more than two dimensions. For example, it would be quite effective for evaluating profitability based on DRGs, payers, and the ages of patients.

Conclusion

HCOs that disregard the need to be constantly analyzing and improving their accepted care processes and the activities that make up those care processes will not remain competitive. The interactive visualization technique presented here was proven to be a powerful tool to support timely care process analyses and to reduce the overall cognitive effort associated with such analyses.

References

About the Authors

Yoshitaka Bito, MS, is a research scientist, Medical Systems Research Department, Central Research Laboratory, Hitachi, Ltd., Tokyo, Japan.

Robert Kero, MS, is chief of business development and senior manager, Information Technology Laboratory, Hitachi America, Ltd., Buffalo Grove, Illinois.

Hitoshi Matsuo, MS, is unit leader and senior researcher, Medical Systems Research Department, Central Research Laboratory, Hitachi, Ltd., Tokyo, Japan.

Yoichi Shintani, PhD, is director and laboratory manager, Information Technology Laboratory, Hitachi America, Ltd., Brisbane, California.

Michael Silver, MD, FACP, FCCP, FCCM, is associate professor of medicine at Rush Medical College, associate director of the Section of Pulmonary and Critical Care Medicine at Rush, and vice president of medical affairs at Oak Park Hospital. He is board certified in internal medicine, pulmonary medicine, and critical care medicine.