Data Warehousing in Disease Management Programs

Denise C. Ramick

ABSTRACT

Disease management programs offer the benefits of lower disease occurrence, improved patient care, and lower healthcare costs. In such programs, the key mechanism used to identify individuals at risk for targeted diseases is the data warehouse. This article surveys recent warehousing techniques from HMOs to map out critical issues relating to the preparation, design, and implementation of a successful data warehouse. Discussions of scope, data cleansing, and storage management are included in depicting warehouse preparation and design; data implementation options are contrasted. Examples are provided of data warehouse execution in disease management programs that identify members with preexisting illnesses, as well as those exhibiting high-risk conditions. The proper deployment of successful data warehouses in disease management programs benefits both the organization and the member. Organizations benefit from decreased medical costs; members benefit through an improved quality of life through disease-specific care.

KEYWORDS

- Data warehousing
- Warehouse implementation
- Warehouse design
- Disease management
- Population management

Disease management programs have proliferated with the advent of managed care and advancing data management technologies. The purpose of these programs is to identify ill or potentially ill members in a system and apply programs to prevent the progression or appearance of the illness, thus maintaining the health of the population.[1] The challenge lies in identifying the individuals at risk for a targeted disease or diseases. A useful tool borrowed from

the traditional business world to be used in the disease management process is the data warehouse.

This article presents recent warehousing techniques from healthcare organizations; critical issues relating to the preparation, design, and implementation of a successful data warehouse are mapped. Discussions of scope, data cleansing, and storage management are included in the depiction of warehouse preparation and design; data implementation options are contrasted. Comprehensive examples from leading HMOs, including design alternatives, illustrate the benefits of proper warehouse deployment for both the organization and the member.

Data Warehouse Overview

The data warehouse is a repository of integrated information from any number of sources with analysis and querying functions.[2] The enormous numbers of data and the powerful analytical capabilities that a data warehouse embodies allow users to identify members, plan programs, track patient progress, evaluate programs, and complete an overall financial analysis.[1] Successful implementation of a data warehouse requires planning, designing, and maintaining the data—a daunting task, considering the vast amounts of information inherent to healthcare organizations. However, the benefits of doing so outweigh the drawbacks. Lower disease occurrence, increased quality of patient care, and lower healthcare costs[3] drive the success and marketability of organizations with a data warehouse to aid in the development of disease management programs.[1]

Two main types of data warehouses are (1) operational data stores or data marts and (2) enterprise data warehouses.[2] Enterprise data warehouses contain data from all areas of the organization and can be used by all departments within the entity. A typical healthcare organization may compile data from various sources such as claims, providers, pharmacy, laboratory, and materials management.[1] An enterprise data warehouse generally stores three to five years of information and is usually not accessed by the user because of its size and complexity.[2] The primary purpose of a data warehouse is to store the cleansed historical information that is the standard source of data from which various, specialized data warehouses are populated.

The spin-off data warehouses, known as data marts or operational data stores, are available to users with a querying and decision support tool. Data marts are more department-oriented because of their smaller scope. It is recommended that data for disease management programs be placed in a data mart because this warehouse type will improve accessibility for users and reduce the number of maintenance issues for the database administrator.[1] Some organizations may opt to begin their data warehouse project with a data mart because it requires significantly less effort and time, as well as fewer financial resources, to deploy.[2] If necessary, the scope of the warehouse can then be broadened to meet other business needs.[4]

Planning. Once the scope of the warehouse is defined, the planning, designing, and implementing process can begin. The planning process must include consideration of data sources, data cleansing, warehouse growth rates, future expansion, data inconsistencies, data semantics, storage management, and external data sources.[2]

Data Sources. Large healthcare organizations have access to voluminous data, so the organization can control the format. Smaller organizations may have to consider forming partnerships to acquire the information integral to their particular disease management program. Unfortunately, data obtained from outside sources may require more data cleansing and translation because the organization does *not* have control over the format. The benefit is that a smaller organization can choose data wisely. For example, it can consider factors such as the time lag associated with claims data, or choose to pursue pharmacy data, which is "real time," or look for detailed information found within provider data, according to the organization's needs.[1]

Another ideal data source that may play a more active role in the future is real-time information transmitted from home healthcare devices.[5] These data will be crucial to member-care management once members are identified to participate in a disease management program. Potential data sources should be extensively researched early in the warehouse planning process so that critical data elements can be included in storage plans.

Data Cleansing. Cleansing data involves removing unnecessary fields, fixing formats, and completing related tasks. Before undertaking this step it is important to consider how the data will be used. Data in a disease management warehouse are used for two important functions. First, the information stratifies patients by risk level for targeted medical conditions. Second, the data track patients' progress through the disease management program, according to Linda Reeder at Advanced Research Systems in Seattle.[1]

What data are important in these two disease management processes? Much of the patient information such as occupation and address is significant because it can reveal important health hazards.[1] Proceed with caution because eliminating data may be like discarding pieces of a jigsaw puzzle; the big picture will never be complete without each interlocking piece. For instance, an organization may want to analyze patient data by residence and diagnosis due to environmental hazards in a particular area. An organization with many members concentrated in this location can help itself and the public by using the data to create a disease management program for any resulting condition. Hence, future uses of data must be considered before discarding pieces of information.

Data Storage. Storage management and expansion potential of the data warehouse are topics that should be discussed as the warehouse application is being planned. Data will be voluminous and will multiply as the disease management programs gain members, data sources, and tracking mechanisms.[1] The key is to store all data and make them accessible to users. Generally, compensating for storage expansion can be handled by system architectures that support parallel processing and operation.[2]

Design. The major design goal of a data warehouse is to meet the analysis information requirements of the disease management program or programs. In other words, data warehouses rely on decision-support and knowledge-based applications rather than on data transaction processing only. The need for information often creates complex relationships among the tables within the warehouse, making the design even more complicated.[2] In fact, ordinary relational databases may not meet analytical needs. Data used for the analytical part of the disease management program may prove more useful in a multidimensional structure or "hypercube."[2] A design of this caliber will support more sophisticated analysis and identify elaborate relationships within the data, which are usually related in several ways, some of which are obvious; others are not. Other concerns to be addressed in the design process include the timing of data updates, whether the architecture is sufficient to deal with incomplete or inconsistent data, and whether the query capabilities are flexible enough to be useful.[2]

Implementation. Once the planning and designing of the data warehouse are complete, an implementation process must be chosen. Basically, implementation can be completed using an *incremental* or an *enterprisewide* approach. The consensus among database administrators seems to favor the incremental implementation strategy—one that entails assembling individual data marts for various departments. Then an enterprisewide data warehouse can be created by integrating all of the organization's data marts.[2]

Maintenance. Maintenance of the data warehouse should be controlled by the database administrator, who is responsible for ensuring the security and integrity of the database, developing user views, and setting user access to the database.

Security. Security issues are critical at healthcare organizations due to current Health Insurance Portability and Accountability Act of 1996 (HIPAA) regulations. The data will contain confidential information such as patient test results and diagnoses. As a result, stringent security measures must be taken to ensure that access is solely provided to a predefined user group. Database administrators must also maintain the metadata, or "knowledge about the data."[2] Metadata define the structure of the data warehouse, the meaning of the data, and the source and life cycle of the data. This information provides a consistent view of the data within the warehouse and the architecture of the warehouse itself.[6]

Data Recovery. Other maintenance requirements include the backup and recovery plans for the data warehouse. A backup strategy should be selected in the design phase because of the time required to complete the process. Recovery plans should be determined and tested for every possible system failure. The wealth of business and patient information contained in data warehouses is enough to guarantee both short- and long-term protection with tested backup and recovery processes.[2]

Data Analysis. Now that the data are accessible in the warehouse or disease management data mart, the information can be analyzed to identify patients for disease management programs. Before analyzing the data, the types

of conditions that will be targeted for disease management programs should be determined. The most common diseases that respond well to intervention programs are diabetes, cardiovascular disease, and asthma. Algorithms can be developed for the target conditions that use the formatted data in the warehouse.[1]

The complexity of the data analysis determines the number of members in the at-risk group for a particular disease. Disease risk levels must be set and adjusted on a regular basis to ensure the coverage of members most in need of care management. A basic analysis may yield too few patients, whereas a detailed analysis may yield many more than the program can manage. For instance, a search of members based on their previous diagnosis may not identify patients with habits that could lead to the onset of the disease. However, simplicity is important in the beginning stages of patient identification for programs. The sooner the organization can get patients into a disease management program, the sooner they will have measurable results on outcomes and healthier members. Positive results will prove beneficial in the recruitment of patients and providers for the programs. Patients will have superior quality of life, which will also benefit the financial picture of the organization[1] by improving the results of member satisfaction surveys. Organizations achieving superior member satisfaction ratings will appear more attractive to employer groups and potential members seeking services in the future.

Program Enrollment. The next step is to enroll the at-risk members into disease management programs. Once these patients are in the program(s), they will generate more data worthy of tracking. During this stage it becomes crucial to collect patient data. The patient care data collected during this phase will allow for the following tasks to be completed or developed: the scheduling of consecutive parts in the program, the identification of serious incidents such as hospitalization, the identification of needs for changes in the program, and the evaluation of the outcome of the program. However, the organization must overcome the tendency to track too much patient data. According to Christine Turner of Integrity Consulting, attempting to collect across varying treatment settings over the course of an entire illness episode will prove difficult.[1]

Detail is important at the start of a disease management program, but over time the company should focus on data that are most informative about patients and about the program. Finally, reports should be generated so that appropriate individuals can monitor patient and program outcomes. The reports will indicate what kind of data are most important and what form is most useful.[1]

Data Warehousing in Practice

U.S. Quality Algorithms (USQA), whose parent company is Aetna US Healthcare, uses a data warehouse to collect administrative data from encounter, pharmacy, and laboratory claims. This business unit within Aetna US Healthcare flags members with certain ailments by using a USQA-designed algorithm that examines several data elements, including diagnoses, procedures, laboratories,

and pharmaceuticals used by the member.[7] The flagged patients are placed in Aetna US Healthcare's disease management programs or are targeted for applicable member mailings.

Empire Blue Cross and Blue Shield is another organization employing data warehousing techniques to improve patient care and monitor diseases. The Systematic Analysis Review and Assistance system (SARA) is a data warehouse that contains information from claims, pharmacy, and utilization management systems.[8] However, this particular data warehouse has unique qualities. First, it not only contains the lab test ordered but also contains the lab results. Second, the time lag associated with the data is only two or three days. Therefore, the information is analyzed when it has the greatest impact on the member and the quality of his or her care.

SARA uses a special software program developed by Active Health Management, Inc., to screen and flag patients for specific values that could signify undesirable medical conditions. Once the cases are flagged, they are compared to a set of four hundred algorithms that indicate when medical intervention is necessary. Approximately four hundred thousand Empire members are monitored by the SARA data warehouse and Active Health software.[8] Since the beginning of the SARA program, Empire has proactively improved both patient care and the organization's financial position.

Horizon Mercy Health Plan of New Jersey discovered through their warehouse that the most common diagnosis among pediatric patients was asthma. Placing these patients in an asthma management program allowed them to make key patient interventions and create education programs to decrease the high emergency room costs incurred by this segment of their member population.[9] Kaiser Permanente has also been successful in using their data warehouse to determine the demographics and effective treatments for members with chronic conditions such as diabetes and cardiovascular disease.[3]

Design Alternatives: Outsourcing

An alternative to designing an in-house disease management warehouse is outsourcing the task to a specialist. This option can save the organization the aggravation of maintenance tasks such as data handling, data cleansing, and report generating. According to Robert N. Robison, chief medical officer of Horizon Mercy, a Medicaid HMO in New Jersey, it is also difficult to develop the clinical logic or algorithms behind the warehouse. He chose to outsource his disease management data warehouse to InfoMiners of Amherst, New York. InfoMiners uses a graphical interface with Cognos analytical capabilities and an Oracle database on the back end. The system also uses InfoMiners' own clinical algorithms based on their voluminous knowledge base, which allows organizations to get their disease management programs functioning quickly. However, organizations must be aware of the drawbacks of outsourcing this function, including inflated expense, lack of system support, and lack of product development.[1]

Conclusion

Regardless of which road organizations take to develop their disease management data warehouse, the benefits of deploying a program are numerous. Data warehouses aid organizations in analyzing patient populations by geographic location, diagnosis, and service consumption to determine which disease management programs will be most beneficial to the patient and the organization.[9] Application of this strategy will improve member health status, increase physician participation, and lower medical expenses. Implemented disease management programs allow providers and organizations to actively manage patient care and educate patients so they can monitor their own condition, which will result in a decreased use of high-cost services. The organization succeeds financially, and the result is a healthier patient. Well-developed disease management efforts may even lower disease occurrence in the population. The CMO of Horizon Mercy summarized the disease management power contained in a data warehouse when he said, "The reason we feel so strongly about these interventions and programs of education is simple: We can either do it early and do it right, or do it later and pay more in cost of added resources and the cost to the patient's health."[9]

References

1. De Jesus, E. X. "Disease Management in a Warehouse." *Healthcare Informatics Online,* Sept. 1999, 1–9.
2. Benander, A., and Benander, B. "Data Warehouse Administration and Management." *Information Systems Management,* 2000, *17*(1), 71.
3. Hollis, J. "Deploying an HMO's Data Warehouse." *Health Management Technology,* 1998, *19*(8), 46.
4. Scheese, R. "Data Warehousing as a Healthcare Business Solution." *Healthcare Financial Management,* 1998, *52*(2), 56.
5. Howe, R. S., Terpening, M. B., and Wadhwa, S. "Disease Management and Clinical Decision Support." *Journal of the Healthcare Information and Management Systems Society,* 1999, *13*(2), 6.
6. Lang, R. "Data Warehousing." [http://courseinfo.sju.edu/courses/HAD_4105/coursedocuments/ _000000. . ./had4105_12.html]. Apr. 23, 2000.
7. Marietti, C. "Diamond Mining." *Healthcare Informatics Online,* Jan. 2000, pp. 1–7.
8. Rodat, J. W. "SARA Creatively Uses Data to Improve Empire's Patient Care." *Capital District Business Review,* 2000, *27*(2), 30.
9. "InfoMiners Instant Info. Horizon Mercy Knows, Understands Its Medicaid Population." [http://www.infominers.com/instantinfo/horizon_mercy.html]. Apr. 18, 2000.

About the Author

Denise C. Ramick is a senior managed care analyst at Kennedy Health Systems, Cherry Hill, New Jersey, and a graduate student at St. Joseph's University with a concentration in health information management systems.