IT ROUNDUP: PART III

Converting EHR Data into Knowledge: The Benefits of Clinical Data Warehouses in Quality and Safety Research

Daniel R. Murphy, MD, MBA, and Hardeep Singh, MD, MPH

Drs. Murphy and Singh are research scientists at the Houston VA Health Services Research & Development Center of Excellence and the Baylor College of Medicine.

Although electronic health records (EHRs) have brought several benefits to health care, many benefits have not been fully realized. For example, EHRs facilitate the storage of vast amounts of clinical data generated each day, but few systems have been able to effectively put those data to use by permitting efficient retrieval of information in a manner that facilitates quality and safety improvement and research.

The Department of Veterans Affairs (VA) has now successfully implemented and used the VistA EHR system for a decade.[1] VistA has evolved to support a variety of clinical processes and allows secure access to large amounts of clinical and administrative information. However, similar to most other EHRs,[2] VistA was not built with a research mindset; accessing a cross-section of information about many patients or conducting data mining studies in VistA can often be tedious or impossible. Furthermore, because data is spread among multiple systems in addition to VistA (e.g. scheduling and billing) and each facility manages its own data locally, analysis of data that spans multiple systems or facilities often requires manually assimilating data extracted from multiple sources.

To facilitate performance measurement, quality improvement, and patient safety research, the VA has established regional data warehouses to store data collected from multiple administrative and clinical systems across multiple VA facilities. Additionally, the VA has recently created a nationwide data warehouse designed specifically for researchers.[3] Our early experience with using VA data warehouses has shown us the tremendous potential benefits from harvesting data needed to improve patient safety and quality of care as we evolve in our EHR journey.

What is a Data Warehouse?

A clinical data warehouse is a large, centralized repository of information extracted from one or more administrative and clinical data systems (often called "operational" or "transaction" systems). While operational systems are designed to handle high volumes of transaction processes, data warehouses include features to support high volumes and varieties of analytic processes that help to make sense of patterns in data. Most data warehouses are composed of three components: software programs used in obtaining and verifying data from primary sources, the data storage repository, and an interface to facilitate access and analysis of data.[4]

Data Warehouse Benefits

Several benefits exist when data warehouses are used together with EHRs. First, data warehouses combine data from many disparate sources—including legacy and retired systems—obviating the need for time-consuming manual extraction and assimilation of data from multiple systems. Furthermore, as data are combined, they are often standardized to allow comparisons to be made (e.g. gender of "male," "M," and "1" might be converted into a uniform value). Second, the integration with tools to perform data mining and statistical analyses allows a wide variety of exploratory analyses to be rapidly performed. Third, because data warehouses are separate from the primary systems that hold the source data, performing analyses does not result in additional load and a concomitant reduction in response time to the "live" EHR system used for clinical practice. Finally, unlike EHRs that typically store data in a patient-centric manner, data warehouses store data in a format that easily allows cross-patient searches (e.g. "find all patients with systolic blood pressure greater than 140").

Quality and Safety Projects Using a Data Warehouse

The use of data warehouses has already shown benefit in hospital infection control programs,[6] efficiency of nursing staffing,[7] and quality management in oncology.[8] In our own work, we are using a data warehouse to develop and evaluate several "trigger tools" to detect patients at high risk of harm from potentially delayed or missed diagnoses.[9] These tools are designed to scan the data warehouse for evidence of potential diagnostic errors (e.g. patients with an unscheduled hospitalization occurring within several days of a primary care visit) or delayed follow-up after an abnormal cancer screening test (e.g. lack of timely colonoscopy after a positive fecal occult blood test). Because these triggers rely on combing through data contained in multiple operational systems (e.g. lab system, patient visit records, clinician scheduling), they would be extremely difficult to conduct without a data warehouse.

Many SGIM members are either already involved or will be involved in EHR initiatives. As many of them might have already realized, few EHRs natively incorporate a data warehouse, greatly limiting the flexi-

continued on page 2

continued from page 1

bility and efficiency of analyzing the data they hold. The use of data warehouses is likely to evolve in health care as clinicians, educators, administrators, and researchers learn about their tremendous potential to meaningfully measure and improve quality and patient safety.

References

1. Brown SH, Lincoln MJ, Groen PJ, Kolodner RM. VistA—US Department of Veterans Affairs national-scale HIS. Int J Med Informatics 2003; 69(2-3):135-56.
2. Ebidia A, Mulder C, Tripp B, Morgan MW. Getting data out of the electronic patient record: critical steps in building a data warehouse for decision support. Proc AMIA Symp 1999:745-9.
3. Department of Veterans Affairs. VA Informatics and Computing Infrastructure (VINCI). Available at: http://www.hsrd.research.va.gov/for_researchers/vinci/. Accessed January 24, 2012.
4. Park H-A, Cho I, Chung E. Exploring use of a clinical data repository containing international classification for nursing practice-based nursing practice data. Computers Informatics Nursing 2011; 29(7):419-26.
5. Pedersen TB, Jensen CS. Research issues in clinical data warehousing. In: Tenth International Conference on Scientific and Statistical Database Management, 1998. Proceedings IEEE 1998:43-52.
6. Wisniewski MF, Kieszkowski P, Zagorski BM, et al. Development of a clinical data warehouse for hospital infection control. J Am Med Informatics Assoc 2003; 10(5):454-62.
7. Park H-A, Cho I, Chung E. Exploring use of a clinical data repository containing international classification for nursing practice-based nursing practice data. Computers Informatics Nursing 2011; 29(7):419-26.
8. Holzer S, Tafazzoli A, Altmann U, Wachter W, Dudeck J. Data warehousing as a tool for quality management in oncology. Stud Health Technol Inform 1999; 68:432-5
9. Singh H, Thomas EJ, Khan MM, Petersen LA. Identifying diagnostic errors in primary care using an electronic screening algorithm. Arch Intern Med 2007; 167(3):302-8.

SGIM