

Using Data from Electronic Medical Records: Theory versus Practice

Tezeta F. Mitiku and Karen Tu

The Issue

In Canada, the measurement of quality of healthcare has historically focused on specialized hospital-based care. Considerably less is known about the quality of care provided in the offices of primary care physicians. Primary care research has relied on data collected manually from physicians' offices or from administrative databases. Manual data collection from paper-based patient charts in primary care physicians' offices is costly and time consuming, and often only a small portion of the information in the charts is useable due to the lack of uniform documentation. Although data from administrative databases are more readily accessible and encompass the entire population, they are limited in their depth of clinical information.

The increased use of electronic medical records (EMRs) by primary care physicians presents an opportunity for the efficient extraction and use of large quantities of clinical information. EMRs capture comprehensive longitudinal information on individual patients not available from other sources, including important risk factors for health outcomes such as smoking status, family history and clinical and laboratory measurements (e.g., blood pressure, body mass index and cholesterol levels).

EMR use in Canada has expanded rapidly, with many provincial governments providing funding for primary care physicians to adopt EMRs into their practices. In 2006, approximately 22% of primary care physicians in Canada (24% in Ontario) were using EMRs (Forster 2006). Canada Health Infoway, an organization established to spearhead the movement toward a national electronic health records system, has estimated that \$10–12 billion dollars are needed to establish basic EMR infrastructure in Canada by 2015 (Canada Health Infoway 2007).

Researchers in the United States and the United Kingdom have demonstrated the utility of EMR data for chronic disease surveillance, management and prevention and for health services research (Holt et al. 2008; Ornstein 2001; Roskell et al. 2004). However, research organizations in these countries obtain the data from a common EMR format. In Canada, there are multiple provincially accredited EMR formats in use (currently 11 in Ontario and 10 in Alberta) (Alberta Netcare 2008; OntarioMD 2008). There is also no single EMR vendor

that has accreditation across all provinces. Although Canada Health Infoway and the provincial agencies are setting standards for interoperability between EMRs, the multitude of software vendors represents a particular challenge for research. In theory, gathering data that are already in an electronic format should be uncomplicated; but in practice, extracting comprehensive information from even a single EMR poses many challenges.

The Findings

Over the past two years, we have been working with a leading EMR software vendor in Ontario to conduct a pilot study to evaluate the feasibility of using data from an EMR to measure cardiovascular-related primary care quality indicators. In so doing, we have gained a better appreciation of the challenges inherent in this undertaking.

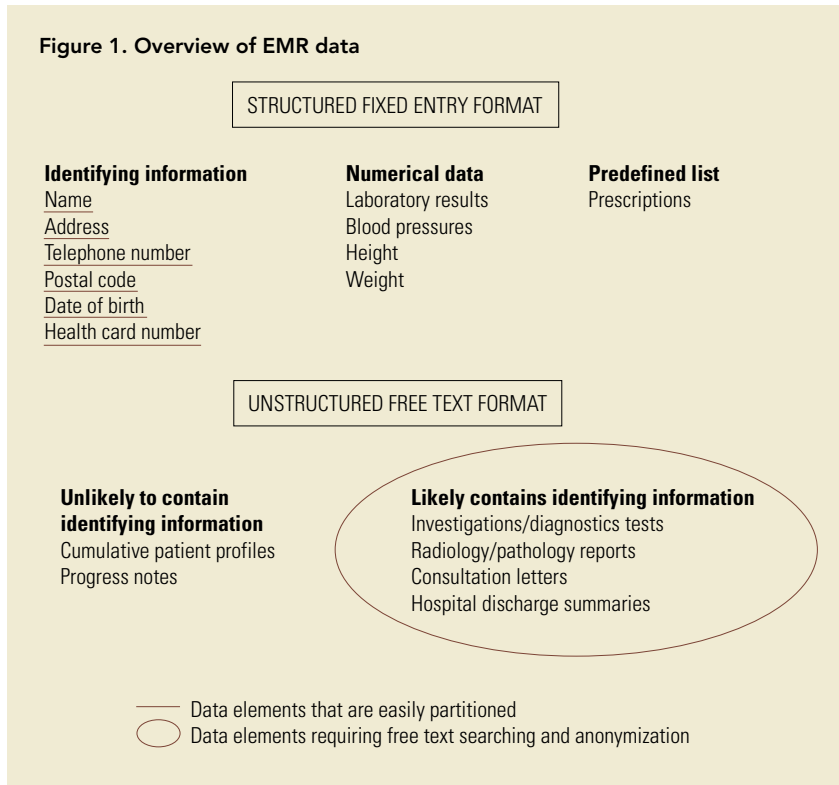
Data Extraction

To extract data from an EMR, it is necessary to engage an EMR software vendor to develop methods for getting information from the dynamic database environment of the EMR into a format that contains specific clinical information under exact variable headings. Data-extraction programs developed by vendors are designed at a specific point in time on a particular version of the EMR software. However, these software programs are constantly being modified and upgraded, thus requiring ongoing engagement between users and vendor software technicians. Further, physicians' offices have different versions of the EMR software depending on when it was installed and whether they have chosen to upgrade their software. This presents a challenge to establishing automated data-collection procedures.

Data De-identification

Data contained within the EMR is entered in either a structured or unstructured format (Figure 1). Structured data may be numerical (e.g., blood pressure readings, lab results) or single words or finite word combinations (e.g., prescriptions). This information can easily be linked with a study ID number and analyzed without compromising patient confidentiality, since it does not contain any identifying information. Access to this

Figure 1. Overview of EMR data



EMR database is necessary to measure the presence of disease-specific quality indicators and to examine patterns of practice; this requires the development of automated techniques. While some disease conditions can be deduced from prescription profiles or laboratory values, many require examination of the free text. Simply searching for the occurrence of a particular disease condition is not sufficient because the inclusion of the condition in the free text does not necessarily mean the patient has that condition. For instance, the free text might note “no evidence of a myocardial infarction,” “mother had an MI,” “rule out MI” or “patient had a heart attack.” All these phrases relate to the same disease condition – myocardial infarction – but each conveys a different message. Furthermore, physicians often use abbreviations and acronyms that are unique. We are currently developing methods to address variations in physician documentation to better identify patients with specific disease conditions.

comprehensive clinical information exceeds what is currently available in administrative data and facilitates the ability to answer important research questions. Data contained in an unstructured/free text format can also add to the research capabilities of EMR data, but unstructured data also has the potential risk of containing personal identifying information. Data that is unstructured/free text can be classified as unlikely to contain identifying information (e.g., progress notes) or highly likely to contain identifying information (e.g., consultation letters). Free-text data must be subject to automated searching mechanisms that can anonymize or strip personal identifying information. A further challenge for de-identification is that in some EMRs the free text (e.g., a consultation letter) is captured in a file format such as PDF or TIFF that requires further processing to convert the text into a searchable format that can be edited.

Data Transfer

Once data have been extracted and anonymized, they must be transported to a central location where they can be analyzed. This is a costly but necessary step that requires sophisticated technological security and expertise along with regular risk assessment.

Disease Identification

Identifying patients with particular disease conditions in the

A Pragmatic Approach to EMR Data Collection

Rather than requiring physicians to perform additional data-entry steps or adopt special coding practices, we are designing methods to make use of the existing data generated by those physicians’ already using EMR software. We adopted this pragmatic approach to lessen the disruption to the established clinical flow in physicians’ offices, in hopes of increasing cooperation and response rate for study participation. In addition, this approach minimizes the selection bias that is inherent in prospective research studies that require physicians to perform additional data-entry steps.

Conclusions

In theory, accessing and analyzing data contained within an EMR appears to be a straightforward process. In reality, there are many barriers and challenges that need to be overcome to set up an EMR database that preserves the richness of information contained within the EMR.

Implications

Although we anticipate that the initial investment for developing an EMR database will be substantial, ongoing maintenance and upkeep with software changes and dealing with multiple software vendors are also likely to incur substantial costs. In addition, the development and validation of automated ways

of processing and extracting information from EMR records is costly and time consuming. The large estimated expenditure for EMR implementation do not take into account costs for the extraction and use of data for research and evaluation purposes. This initial investment of time and finances is necessary to use this rich data source in Canada to its fullest. To our knowledge, this retrospective approach to capturing and evaluating EMR data is a leading initiative in Canada.

Recommendation

If seamless portability of data between accredited EMR vendors cannot be attained, then the number of such vendors within provinces and across the country should be considerably reduced so that the substantial investment in EMRs maximize their research potential in addition to achieving enhanced patient care goals. **HQ**

Acknowledgements

This work was supported by a Canadian Institutes of Health Research Team Grant in Cardiovascular Outcomes Research to the Canadian Cardiovascular Outcomes Research Team (CCORT).

References

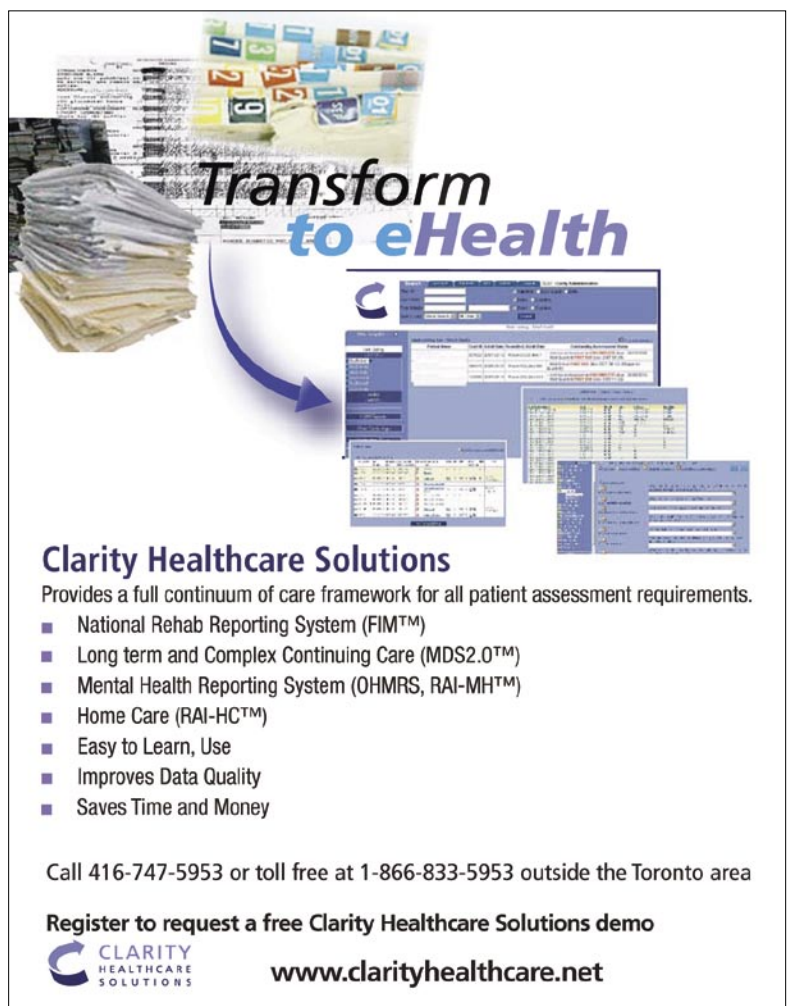
- Alberta Netcare Physician Office System Program. 2008. *VCUR Product List*. Edmonton, AB: Author. Retrieved : July 25, 2008. <http://www.posp.ab.ca/files/VCURProductList-080318v3.11_000.pdf>.
- Canada Health Infoway. 2007. *2015: Canada's Next Generation of Healthcare at a Glance*. Toronto: Author. Retrieved : July 25, 2008 <<http://www.theglobeandmail.com/partners/free/infoway/pdf/2015%20Health%20care%20at%20a%20glance%20EN.pdf>>.
- Forster, B. 2006. "Engaging Physicians in Ontario's eHealth Vision." *Waterloo Smarter Health Seminar Series*, September 27, 2006. Retrieved : November 7, 2007. <<http://infranet.uwaterloo.ca/infstalks/2005-2006/2006-09-27/default.pdf>>.
- Holt, T., D. Stables, S. O'Hanlon, J. Hippisley-Cox and A. Majeed. 2008. "Identifying Undiagnosed Diabetes: Cross-sectional Survey of 3.6 Million Patients' Electronic Records." *British Journal of General Practice* 58: 192-96.
- OntarioMD CMS Vendor Rankings for June 2008*. 2008. EMR Advisor. Retrieved : July 25, 2008 <<https://www.emradvisor.ca/node/253>>.
- Ornstein, S.M. 2001. "Translating Research into Practice Using Electronic Medical Records. The PPRNet-TRIP Project: Primary and Secondary Prevention of Coronary Heart Disease and Stroke." *Topics in Health Information Management* 22: 52-58.

Roskell, N.S., J.W. Logie, M. Stender and M. Feudjo-Tepie. 2004. "A Systematic Approach for Describing Comorbidities Using the UK General Practice Research Database." *Pharmacoepidemiology and Drug Safety* 13: S41-42.

About the Authors

Tezeta F. Mitiku, BSc, MSc(c), is a graduate student in the Department of Community Health and Epidemiology at Queen's University, Kingston, Ontario, and a research assistant at the Institute for Clinical Evaluative Sciences.

Karen Tu, MD, MSc, CCFP, FCFP, is a scientist at the Institute for Clinical Evaluative Sciences. She is a practicing family physician at the University Health Network Toronto Western Hospital and a funded researcher and associate professor in the Department of Family and Community Medicine at the University of Toronto, Toronto, Ontario.



Transform to eHealth


Clarity Healthcare Solutions

Provides a full continuum of care framework for all patient assessment requirements.

- National Rehab Reporting System (FIM™)
- Long term and Complex Continuing Care (MDS2.0™)
- Mental Health Reporting System (OHMRS, RAI-MH™)
- Home Care (RAI-HC™)
- Easy to Learn, Use
- Improves Data Quality
- Saves Time and Money

Call 416-747-5953 or toll free at 1-866-833-5953 outside the Toronto area

Register to request a free Clarity Healthcare Solutions demo

 **CLARITY HEALTHCARE SOLUTIONS**

www.clarityhealthcare.net