MAYO CLINIC

EDITORIAL

September 2012
Volume 87
Number 9

MAYO CLINIC
PROCEEDINGS

Assessing Automated Extraction of Prognostic Information for Intensive Care Unit Patients

See also page 817

In the current issue of *Mayo Clinic Proceedings*, Singh et al[1] report the feasibility of using an automated digital algorithm to extract prognostic information from electronic medical records (EMRs). Electronic medical records are currently widely used in the hospital inpatient setting and are increasingly being adopted in the office setting,[2] possibly in part due to the potential financial incentives in the 2009 Health Information Technology for Economic and Clinical Health Act.[3] Electronic medical records are expected to increase the transparency and efficiency of patient care, improve the ability to study health care delivery, and improve individual health outcomes and population health.[4] Studies of EMRs and other health information systems have shown beneficial effects of varying magnitude on medication management, preventive care, management of specific health conditions, data quality, and the processes and outcomes of care.[5] Studies of EMRs in the office setting have also shown some benefits.[6]

One basic task for physicians in managing hospital patients using EMRs is extracting relevant information from patients' medical records. Traditionally, information about patients' medical and surgical histories is recorded in a separate past medical history and past surgical history (PMH/PSH) section of the medical record. This practice was likely started to facilitate the efficient use of physicians' time and was apparently of such self-evident value that it was widely disseminated and adopted in hospital medical records. However, reviewing the medical record and the PMH/PSH requires physicians' time, and the information may not be recorded in ways that are most useful to physicians. There are few or no published studies to validate the value of organizing clinical information in the PMH/PSH section of the EMR in the hospital setting. One study that evaluated the value of the components of EMRs in the office setting was limited in scope, did not show improvements in outcomes, and did not specifically evaluate the value of the PMH/PSH section.[7]

Despite these aforementioned weaknesses in the indexed literature, there is still optimism that the summary information in the PMH/PSH section of the medical record may be of value in managing patients and contributing to improving outcomes. Thus, the PMH/PSH section of the medical record can serve as a source of information for studies of prognosis. The Charlson Comorbidity Index (CCI), a weighted index based on the presence of 19 medical conditions, was developed to predict mortality after hospital admission. It is one of the earliest and most widely studied prognostic indices.[8] The CCI is, thus, an ideal vehicle to test the value of automated extraction of information from the hospital medical record.

In the study reported by Singh et al,[1] an automated digital algorithm was developed to extract the 19 components of the CCI from the PMH/PSH section of the EMR in a 1-year sample of 1447 intensive care unit patients; the algorithm was later evaluated in a second sample of patients. The automated digital algorithm was developed using Boolean logic. (The functionality of this tool relies on the presence of any term from a list of diagnoses, acronyms, and similar terms and the absence of any words from a list of negated terms.) The algorithm was developed and refined for each condition to achieve a 95% or greater sensitivity and 95% or greater specificity compared with the criterion standard ("gold standard") on the basis of review of 5 previous years of medical records by trained research fellows.

The automated digital algorithm for extracting comorbid medical conditions from the PMH/PSH was compared with extraction of *International Classification of Diseases, Ninth Revision,* codes (developed by the World Health Organization), which were presumably assigned by medical record coders using hospital discharge diagnoses. These 2 approaches were evaluated in a second, smaller sample of 240 patients with severe sepsis, in which the prevalence of comorbid conditions varied from 0% (AIDS) to 32.1% (malignancy). The 2 methods were evaluated individually by estimating the diagnostic test characteristics (sensitivity, specificity, positive predictive value, and negative predictive value), and the confidence limits around these estimates, for each of the component comorbid conditions of the CCI.

The automated digital algorithm was developed successfully, with uniformly high sensitivity and specificity for each of the CCI components. In the separate validation group of patients, the sensitivity and specificity of each of the 19 components of the CCI were higher than the corresponding measures for the search of the *ICD-9* codes from the hospital discharge data. The validation is limited by the absence of prespecified statements about the degree of improvement in performance of the automated digital algorithm compared with the extraction of *ICD-9* codes that could be considered clinically important and by the absence of any formal statistical testing of the differences in the diagnostic test characteristics of the 2 approaches. The impact of even small improvements in sensitivity and specificity can be substantial and will vary by the prevalence of the specific comorbid conditions. The validation sample of patients with severe sepsis was smaller than the derivation sample and had a higher prevalence of comorbid conditions, suggesting that this was a sicker group of patients. It is possible that the automated digital algorithm would perform better in sicker patients with more comorbid conditions or with more severe comorbid conditions. One useful summary measure of the 2 approaches would be the magnitude of change in the CCI prediction of 1-year mortality.

A strength of the study is the detailed description of the search terms and negated terms used in developing the automated digital algorithm. This should stimulate other researchers to replicate the work and extend the work to other areas. The ability of other institutions to implement this or similar automated digital algorithms to extract information from hospital EMRs will depend on the quality and completeness of their data in the EMR and on the availability and ease of use of EMR functions for the complex logical queries that were developed using the Mayo Data Discovery and Query Builder. The query functions are likely to vary in ease of use across platforms and vendors. The automated digital algorithm extraction of the 19 comorbid conditions in the CCI is a rather simple task compared with other more ambitious approaches to extracting information from EMRs.[9] Nevertheless, it is reasonable to expect incremental advances in applications of these approaches in the future.

How else might the automated digital algorithm for comorbidity be used? One obvious extension would be to develop queries of any sections of the medical record for a variety of clinical conditions, not just the 19 components of the CCI. Another extension would be to use the automated digital algorithm for the 19 components of the CCI in decision support systems to provide prognostic information to physicians caring for intensive care unit patients or other hospital patients.

The automated digital algorithm may be less useful in extracting information about the PMH/PSH from office-based EMRs. In the primary care setting, physicians tend to have an ongoing relationship with patients and acquire a more comprehensive understanding of patients' medical conditions and past surgical care over time. However, the automated digital algorithm may still have value in extracting information from records of past visits that may have been forgotten, in circumstances where there is limited time during the office visit to review medical records extending over many years, or in situations in which patient care must be provided by a surrogate (eg, cross-covering) or replacement physician.

Important issues not addressed in the present study relate to the actual impact of the information extracted by the automated digital algorithm (and other decision features) on the processes and outcomes of care. These features may alter the processes of care by improving decisions, adding prognostic information that may alter treatment decisions, and potentially improving outcomes by increasing more appropriate care or decreasing harmful or futile care. Alternatively, the additional information extracted by automated digital algorithms may be distracting, contribute to "alert fatigue," or increase costs by leading to additional testing and evaluation of the comorbid conditions. In hospitalized patients, how will busy intensive care unit physicians, hospitalists, consulting specialist physicians, and others use the time saved by the automated digital algorithm? Will more time be devoted to visits with patients and their families? Will less time be spent directly eliciting information about medical history? The SUPPORT (Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments) study (a trial designed to improve care in seriously ill hospitalized patients) reminds us that even in the best of settings, improved prognostic information and advance directives may not lead to improved care or better outcomes.[10] Furthermore, well-controlled studies are needed to ascertain whether the automated digital algorithm results in measurable improvements in clinical decisions, processes of care, and patient outcomes.

Marc D. Silverstein, MD

Abbreviations and Acronyms: **CCI** = Charlson Comorbidity Index; **EMR** = electronic medical record; **ICD-9** = International Classification of Diseases, Ninth Revision; **PMH/PSH** = past medical history and past surgical history

Correspondence: Address to Marc D. Silverstein, MD, PO Box 544, Mashpee, MA 02649 (mds02@comcast.net).

REFERENCES

1. Singh B, Singh A, Ahmed A, et al. Derivation and validation of automated electronic search strategies to extract Charlson comorbidities from the electronic medical records. *Mayo Clin Proc*. 2012;87(9):817-824.
2. Hsiao CJ, Hing E, Socey TC, Cai B. Electronic Health Record Systems and Intent to Apply for Meaningful Use Incentives Among Office-Based Physician Practices: United States, 2001-2011. Hyattsville, MD: National Center for Health Statistics; 2011.
3. Blumenthal D. Stimulating the adoption of health information technology. *N Engl J Med*. 2009;360(15):1477-1479.
4. Blumenthal D. Launching HITECH. *N Engl J Med*. 2010;362(5):382-385.
5. Lau F, Kuziemsky C, Price M, Gardner J. A review on systematic reviews of health information system studies. *J Am Med Inform Assoc*. 2010;17(6):637-645.
6. Lau F, Price M, Boyd J, Partridge C, Bell H, Raworth R. Impact of electronic medical record on physician practice in office settings: a systematic review. *BMC Med Inform Decis Mak*. 2012;12:10.
7. Keyhani S, Hebert PL, Ross JS, Federman A, Zhu CW, Siu AL. Electronic health record components and the quality of care. *Med Care*. 2008;46(12):1267-1272.
8. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373-383.
9. Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF. Extracting information from textual documents in the electronic health record: a review of recent research. In Geissbuhler A, Kulikowski C, editors. *Yearb Med Inform*. 2008;47(suppl 1):128-144.
10. SUPPORT Principal Investigators. A controlled trial to improve care for seriously ill hospitalized patients: the study to understand prognoses and preferences for outcomes and risks of treatments (SUPPORT). *JAMA*. 1995;274(20):1591-1598.