

Using Technology to Promote Gastrointestinal Outcomes Research: A Case for Electronic Health Records

Ashish Atreja, M.D., M.P.H.,^{1,2} Jean-Paul Achkar, M.D.,³ Anil K. Jain, M.D.,^{1,2}
C. Martin Harris, M.D., M.B.A.,^{1,2} and Bret A. Lashner, M.D., M.P.H.³

¹Medicine Institute, ²Information Technology Division, and ³Digestive Disease Institute, Cleveland Clinic, Cleveland, OH

Electronic health records (EHRs) have been shown to reduce medication errors, improve patient outcomes, and create administrative efficiencies. Numerous public and private efforts are currently underway to achieve universal EHR adoption in the United States by the year 2014. EHRs hold a great potential to integrate clinical care and research by allowing input of clinical data in a structured format, facilitating electronic data capture for clinical trials and providing linkage with genomic information. The goal of this article is to inform the academic gastrointestinal community about the research opportunities created by the widespread adoption of EHRs and present a systematic approach in utilizing EHR-derived data for observational, experimental, or translational studies.

(Am J Gastroenterol 2008;103:2171–2178)

INTRODUCTION

The United States Department of Health and Human Services declared the start of a “decade of health information technology” in 2004 and outlined a strategic framework to build a national health information infrastructure featuring joint public and private cooperation (1, 2). The plan called for universal electronic health record (EHR) adoption by the year 2014 so as to improve health-care quality, prevent medical errors, decrease paperwork, and increase administrative efficiencies. It is estimated that universal EHR adoption could cut 10% or more from the nation’s yearly \$1.7 trillion health-care spending (2, 3).

One aspect of health information technology that has received little attention is the potential benefit of EHRs for clinical research. Researchers have relied on patient records for decades to search for clues to disease, spot trends in treatment effectiveness, monitor the safety of drugs and procedures, and identify candidates for clinical trials. Electronic records could facilitate the integration between clinical and research environments, leading to great improvements in the scope and efficiency of research. The benefits range from systematically generating hypotheses for research and facilitating translational research to conducting postmarketing surveillance studies.

To successfully utilize EHRs for research, the gastrointestinal (GI) research community needs to be aware of the EHR features and the framework in which EHR-based research can be conducted. Therefore, the goal of this viewpoint article is to:

1. Provide a broad conceptual overview of various applications of a modern EHR system that can be leveraged for GI outcomes research.

2. Present a systematic approach in utilizing EHR-derived data for observational, experimental, or translational studies.

The article also provides examples of published GI research based on EHR-derived data, highlights the limitations and challenges in secondary uses of EHR data, and offers strategies for moving forward.

CONCEPTUAL OVERVIEW OF EHR

In its most simple form, EHR can be defined as computerization of health record content and associated processes. The term EHR has often been used interchangeably with electronic medical record (EMR), even though there are minor semantic differences between the two terms. We chose to use the term EHR instead of EMR to reflect the growing consensus toward a vision of interoperable electronic records.

Core Functions of EHR

As part of a national effort to encourage the adoption of computer-based health records, an institute of medicine panel has identified a set of eight core functions that EHRs should perform to promote greater safety, quality, and efficiency in health-care delivery (Fig. 1) (4). Health information and data, results management, electronic communications, and support for administrative processes are the essential components of all EHRs that most clinicians are already familiar with. Order management includes functionalities such as computerized physician order entry that allows providers to directly enter laboratory, medications, and radiology orders electronically instead of writing on order sheets or prescription pads. The electronic entry allows the clinical decision support to

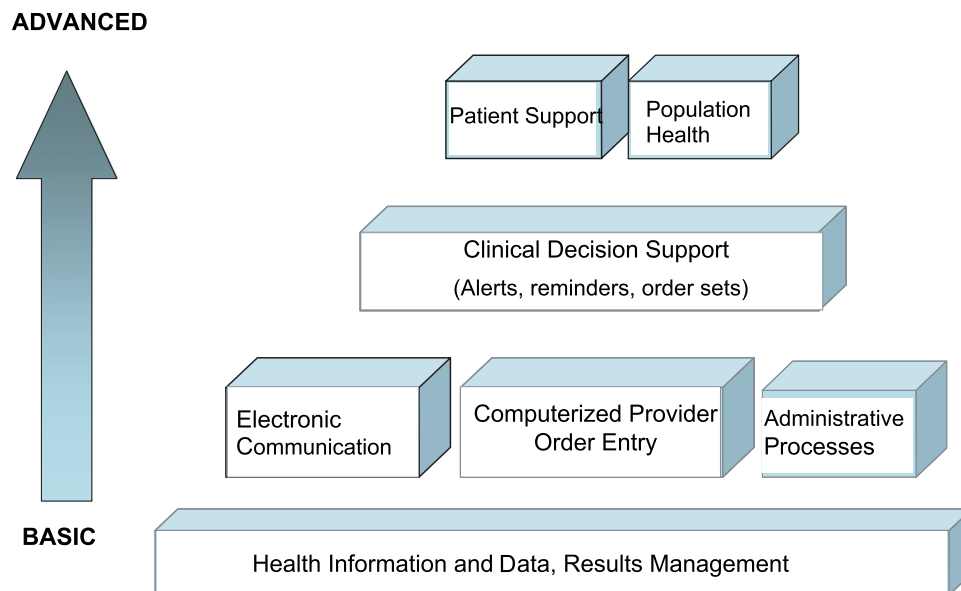


Figure 1. Core functional blocks of EHRs.

compare the order against standards for dosing, allergies, etc., warns the physician about potential problems, and provides context-specific information to the provider at the point of care (5). Patient support means providing tools that give patients access to their health records. This could be achieved through personal health records (PHRs) that allow patients' access to their health insurance information, current medications, immunizations, allergies, important tests results, and future appointments.

Clinical Benefits of EHR

Health information, results management, electronic connectivity, and administrative support activities provide 24/7 chart access and better view of clinical data and process improvement. Computerized physician order entry can lead to patient safety by increasing accuracy and legibility of the order, improving timeliness of order, and integrating decision support into the order-entry process (6). Decision support can provide alerts for drug–drug, drug–allergy, and drug–food interactions based on routinely updated drug formularies. It can also provide additional support including alerts of critical values, reminders for preventive screening, and recommendations for best practices (7–9).

Studies have shown that properly designed computerized physician order entry and decision support can lead to marked reduction in serious medication errors (7) and improve diagnosis, preventive care, and disease management (10). They can also lead to improvement in quality indicators and hard patient outcomes such as preventing pulmonary embolism in high-risk hospitalized patients (11). System-wide benefits from fully functional EHR systems also include reduced staff time spent on paperwork, reduced redundancy of laboratory and imaging services, and increased accuracy and timeliness of billing. In GI practice, EHRs can also help in streamlining guidelines-based practice by enabling disease-based tem-

plates, delivering patient education material on completion of the office visit, generating consult/referral letters based on the data entered during the patient encounters, sending recall letters for preventive services (such as screening colonoscopy), and tracking patients requiring regular follow-up (12–14).

APPLICATION OF EHR DATA FOR RESEARCH

As EHRs are widely adopted by clinicians, there will be an enormous increase in the electronic health data available for research. The reporting and population health management functionality of EHR can support secondary use of health data, which is defined as nondirect care use of personal health information (PHI) including but not limited to analysis, research, quality and safety measurement, public health, payment, provider certification or accreditation, and marketing and other business activities (15).

Information and Data Flow

To appropriately harness data in EHRs for research, it is imperative that we first understand the information and data flow in EHRs. As shown in Figure 2, EHRs can be divided conceptually into front and back ends. The front end is the part of a software system that interacts directly with the user; while in most cases, the back end is hidden from the user. In the case of EHRs, front end is the display and the functionalities that clinicians use to support clinical care, including charting notes and diagnoses, scheduling laboratory and imaging tests, and viewing alerts. The back end comprises the components that process the output from the front end. This includes clinical data repositories that integrate data entered in an EHR with data available from other health information systems such as endoscopy software, imaging systems, or tissue banks.

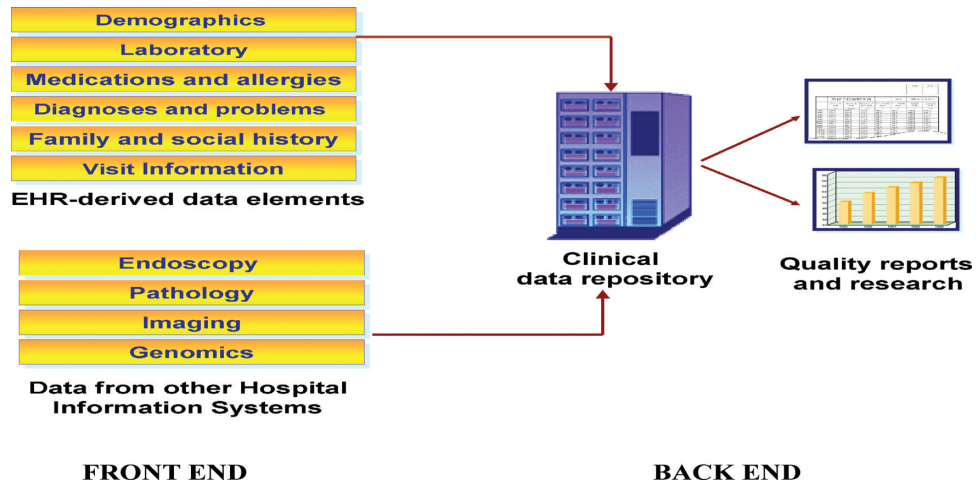


Figure 2. Overview of the information flow from front end to back end in EHRs.

Structured Versus Unstructured Data

Electronic data stored in clinical data repositories could be structured data, such as medications, laboratory results, or International Classification of Diseases (ICD) codes, or unstructured data that represent information as narrative “free text” that is not represented in any standard coding scheme (see Table 1). This distinction is crucial because structured data such as “gastric ulcer with hemorrhage” coded as “ICD-9 531.0” support consistent retrieval, reporting, and data aggregation for research using commonly available database tools. Unstructured data such as “patient has been having blood during vomiting,” on the other hand, are tough to interpret, aggregate, and analyze using automated tools. For the rest of the article, our focus will be on the use of structured data because it easily lends itself to coding schemes that can be readily utilized by the academic medical community for research.

Administrative Data Versus EHR Data

Administrative data, also called claims data, are byproducts of administering and reimbursing health-care services. The core data elements of an administrative data system are admission date, discharge date and status, ICD-9 coded diagnoses, procedures, and some demographic variables. These

data are often available as compiled research databases from federal agencies, state health departments, health plans, and private data institutions. A good example is the Healthcare Cost and Utilization Project (HCUP) databases that include longitudinal hospital care data in the United States, with all-payer, encounter-level information beginning in 1988 (<http://www.ahrq.gov/data/hcup>). Even though claims data are readily available, inexpensive, computer readable and cover large populations, they have important limitations that EHR-derived data can potentially overcome (16).

LACK OF CLINICAL DETAILS FOR RISK ANALYSIS AND RISK ADJUSTMENT. The most important limitation of claims data is the lack of clinical details. Data required for outcomes are limited in administrative data sets and can underestimate comorbidities and risk factors (17, 18). As seen in Figure 2, EHR-derived data are more comprehensive, especially with respect to clinical details and risk factors, and are shown to be superior to administrative data sets in predicting disease prognosis and outcomes (18).

INADEQUATE VALIDITY AND RELIABILITY. In general, administrative data have been shown to have fair specificity, but low sensitivity, in identifying the target population of interest. This could be because the most commonly used algorithms for the identification of target populations rely on claims-based encounter diagnoses. If the encounters are not billed or coded appropriately, patients may not be identified as having that disease and, hence, are incorrectly excluded from the study population. Studies done on administrative data sets have shown that relying on information from primary diagnosis alone may be inaccurate about half of the time (19, 20). EHRs provide more than one way of triangulating the data and identifying population of interest. For example, information from laboratory data (such as positive celiac antibody), allergies (gluten intolerance), diet (gluten-free), diagnosis (ICD-9 codes), or small bowel histology can be combined to better identify a cohort of celiac disease patients from an EHR.

Table 1. Some of the Commonly Available Structured and Unstructured Data Elements in an EHR*

Structured Data Elements	Unstructured Data Elements
Laboratory results	Chief complaint
Medications and allergy	Visit documentation
Diagnoses and surgeries	Discharge documentation/dictation
Demographics	Free text procedure and radiology reports
Appointments, hospitalizations dates	Images and graphics

*Vary among different EHR systems.

This strategy has been shown to be more accurate in identifying target population than the use of diagnostic codes alone (21, 22).

LIMITATIONS OF THIRD PARTY-CONTROLLED DATA. Many limitations of administrative data stem from the fact that they are collected and managed by third parties. Most of the information is anonymized, so if an interesting trend or pattern is seen on analysis, there are not many ways to confirm the findings or extract more relevant data of interest. Hence, the research questions have to be limited to what the core data elements of the administrative data system can provide. EHRs fundamentally differ from administrative data systems in this regard. EHR-derived data can be linked back to the patient and used to retrieve other data elements of interest that may not have been the focus of research questions before. In addition, many EHRs provide the ability to customize data entry through standardized questionnaires or other methods of electronic data capture. Outcomes metrics such as quality of life instruments can be obtained from patients via patient diaries through PHRs. The authors of the article are successfully leveraging EHRs to capture many structured data elements such as documentation of procedures not included in the EHR package (e.g., central venous line insertion).

EHRs AND GI OBSERVATIONAL STUDIES

Appropriate use of observational studies permits investigation of prevalence, incidence, associations, causes, and outcomes. EHR-derived data can be used to produce a sample for a cross-sectional or case-control study, to construct a cohort, or to identify people with certain conditions or outcomes.

Cross-Sectional Studies

Cross-sectional studies help provide data on incidence, prevalence, and demographic factors that may yield clues to the etiology of disease or help identify associations and risk factors not previously known (23). They can also provide information about the burden of illness and the variation in clinical practices that can be utilized in quality improvement studies. For example, Rubin *et al.* utilized the medical records of general practitioners in northern England to identify cases of inflammatory bowel disease and found that the point prevalence of 243.4/100,000 for ulcerative colitis and 144.8/100,000 for Crohn's disease were substantially higher than previously described in U.K. populations (24). Using the same database, Stone and colleagues found that only 65% of patients with ulcerative colitis had been prescribed a 5-aminosalicylate in the previous 6 months with significant differences between patients under specialist care and those under general practitioner care only (25).

Case-Control Studies

Case-control studies can help determine the relative importance of a predictor variable in relation to the presence or ab-

sence of the disease. They are the preferred design when the conditions are uncommon or when there is a long latency period between an exposure and the disease (23). EHRs can save time in identification of patients with uncommon diseases (cases) and also help find matched controls from a large sample using predefined computer-based algorithms. To cite an example, Graham *et al.* used a nested case-control design to determine if the use of celecoxib, ibuprofen, naproxen, rofecoxib, or other nonsteroidal anti-inflammatory drugs increase the risk of heart attack and death (26). Drug prescription files from EHR were used to identify 1,394,764 people that contributed 2,302,029 person-years of observation time to the study. The authors found that rofecoxib use increased the risk of serious coronary heart disease compared to celecoxib use, with adjusted odds ratio (OR) of 1.59 (95% confidence interval [CI] 1.10–2.32, $P = 0.015$). Aggregating and analyzing such vast amounts of data in a limited time would not have been possible if the study authors did not have access to computerized records. Similarly, Laheij *et al.* used data from EHR-fed Integrated Primary Care Information database to conduct a nested case-control study examining the relationship between the risk of community-acquired pneumonia and the use of gastric acid-suppressive drugs. The authors found that the current use of acid-suppressive drugs was associated with an increase in the risk of pneumonia (adjusted OR 1.27, 95% CI 1.06–1.54) (27).

Cohort Studies and GI Disease-Specific Registries

The cohort design allows determination of the incidence and natural history of a condition. EHRs can provide data on diagnoses, risk factors, and demographics required to construct a cohort and study its outcomes. For example, Van Soest and colleagues examined the changes in Barrett's esophagus (BE) incidence relative to the number of upper GI endoscopies performed by looking at longitudinal data from more than 500,000 persons (28). They found that the incidence of BE increased from 14.3/100,000 person-years in 1997 to 23.1/100,000 person-years in 2002, while the number of upper GI endoscopies decreased over the same time period. The authors concluded that the increase in BE incidence is independent of the number of endoscopies and predicted an increase in the incidence of esophageal adenocarcinomas in the near future. The cohort design can also be used to determine outcomes of patients undergoing specific procedures or surgeries. Levin *et al.* used the EHR-derived data to construct a retrospective cohort of patients undergoing colonoscopy at Kaiser Permanente of California and quantify the magnitude and severity of colonoscopy complications. The authors reported that biopsy or polypectomy was associated with an increased risk for any serious complication (rate ratio 9.2 [95% CI 2.9–29.0] vs colonoscopy without biopsy) (29).

The Hepatitis C case registry supported by the Department of Veterans Affairs (VA) is a good example of how EHRs can be leveraged to create prospective cohorts and registries (30). The registry contains important demographic, clinical, pharmacy, and laboratory data and is used on the national,

regional, and local levels to track and optimize clinical care of hepatitis C-infected veterans served by the VA. More details on the VA's comprehensive use of EHR-derived electronic databases and their use for research purposes are available at www.virec.research.med.va.gov.

EHRs AND CLINICAL TRIALS

Prospective controlled studies and clinical trials provide the strongest evidence regarding the effect of an intervention and are a priority for academic health centers, National Institutes of Health (NIH), and pharmaceutical agencies (23). However, they are expensive to conduct. It is estimated that a drug typically costs \$75 million over the 6 yr that it takes to go through clinical trials and approval (31). As much of the patient information collected for clinical trials already exists in the patient record, EHRs hold great promise in facilitating successful conduct of these studies (32, 33). It is estimated that each month saved by EHRs in planning, patient recruitment, and data collection can be worth \$25 million earned in drug sales (31).

Prestudy Protocol Feasibility Analysis

One of the first steps in a clinical trial after creating a protocol is assessing its feasibility, allowing study investigators and sponsors to estimate the number of research subjects that can be potentially recruited using inclusion and exclusion criteria. Structured data from EHRs such as demographics, laboratory, medications, and diagnoses can be used to determine if an adequate number of patients are available in a center for recruitment and that the study demands are reasonable, given the tight schedules of clinical trials (32).

Patient Identification and Recruitment

Once it is determined that the study is feasible, the next challenge is adequate patient recruitment to meet study goals. Many institutions have started to utilize EHR-derived databases to automate identification of patients that meet eligibility criteria and recruit them using traditional methods (letters, phone calls, etc.) or via provider alerts at the point of care (34, 35). The authors of this article have reported the use of an EHR-based clinical trial alert system in outpatient clinics that prompted physician consideration of the patient's eligibility and facilitated secure messaging to the trial's coordinator (35). The alerts were associated with a 10-fold increase in physicians' referral rate (5.7/month before, 59.5/month after; rate ratio 10.44, $P < 0.001$) and a doubling of enrollment rate (2.9/month before, 6.0/month after; rate ratio 2.06, $P = 0.007$).

Electronic Data Capture and Patient Outcomes

Today, around one-third of the clinical trials use electronic data capture systems, which are software applications to store trial data in an electronic form. Due to lack of industry-wide standards that allow an exchange of information between re-

searchers and clinicians, currently, the majority of data stored in EHRs has to be printed or hand-transcribed and re-entered into the electronic data capture system. Such duplication can be reduced once standards allow capturing additional trial-specific data from EHRs or via patient diaries in personal health records (32).

Postmarketing Surveillance and Pharmacogenomics

The Food and Drug Administration (FDA) critical path initiative to modernize the drug development process and speed up drug approval has also led to an increased recognition of the role of postmarketing surveillance and evaluation of drug safety. Most phase III clinical trials now occur over such a short duration that there is insufficient time to determine the long-term consequences of a drug (36). EHR-derived large data sets are increasingly being mined for pattern recognition to uncover correlations between particular drugs and adverse outcomes that could not be identified during clinical trials. With the mapping of the human genome, EHR-derived phenotypical data can support pharmacogenomic profiling where adverse drug reactions can be associated with specific genetic polymorphisms, and thus, help provide the missing link between molecular discoveries and human health and disease.

A leading example of using EHR data for pharmacoepidemiological research is the General Practice Research Database (GPRD) in the United Kingdom. GPRD is the world's largest computerized database of anonymized longitudinal records from general practice, comprising over 35.0 million patient-years worth of data collected from approximately 8.9 million patients (37, 38). The GPRD collects truly population-based data, has a size that makes it possible to follow up large cohorts of users of specific drugs, and includes both outpatient and inpatient clinical information. In a recently published postmarketing surveillance study using GPRD data sets, researchers looked at data from 13,556 hip fractures and 135,386 controls and found that prolonged use of high-dose proton pump inhibitor (PPI) is associated with an increased likelihood of incurring hip fractures (adjusted OR 1.44) (39). Even though this study needs to be prospectively validated, it highlights the critical role of EHRs in enabling postmarketing surveillance and phase IV trials.

LIMITATIONS AND FUTURE DIRECTION

EHR is not a panacea and leveraging it for research requires a well-rounded team of people with diverse skill sets such as programming, database administration, statistics, clinical informatics, etc. In addition, EHRs suffer from a few important limitations:

Variation Among EHRs

There are more than 200 ambulatory EHRs in the market today, with a lot of variation in the functionalities that support quality care or research. Hence, it is important to know which EHRs meet the basic requirements and provide the best fit for

one's practice or organization. However, choosing an EHR that is best fit for the organization is time-consuming and often a multistage process that involves understanding one's practice needs, identifying key decision makers, writing a request for proposal, attending demonstrations, conducting site visits, and negotiating the contract (40). In 2004, the Certification Commission for Healthcare Information Technology (CCHIT) was formed as an official agency to apply standards, test products, and award a "seal of compliance" to EHR software. The commission provides a list of certified ambulatory and inpatient EHRs so that end users can expect to leverage them for their clinical and research needs (<http://www.cchit.org/choose/>). Another resource is the American Gastroenterological Association (AGA) Center for Quality in Practice that has partnered with Medical Strategic Planning, Inc., to provide an EHR selector tool that uses over 300 practice-related criteria for more than 20 leading ambulatory EHR products (<http://www.gastro.org>).

Lack of Accurate and Structured Data

EHR-derived data are only as accurate as the data entered by the providers during clinical care. Incorrect entry of data, variability in practice style, and lack of adoption of clinically relevant vocabulary (such as no distinct ICD-9 code for indeterminate colitis) may make the data useless for research—a term commonly referred to as "garbage in, garbage out." Furthermore, many of the current EHRs store procedure documentation or operative notes in unstructured format or as free text that limits the use of the data for research or outcomes (41).

Endoscopic EMR (EEMR) or computerized endoscopic medical record (CEMR) can help overcome the limitation of many EHRs by supporting the information management in the endoscopy unit, such as capturing structured procedural data, generating endoscopy reports, supporting patient scheduling and monitoring (such as documenting vital signs during endoscopy), facilitating billing and coding, providing pathology interface, and documenting inventory for pharmaceuticals and supplies (42–44). Some EEMRs also support quality assurance and outcomes research. A noteworthy example of such EEMR is the Clinical Outcomes Research Initiative (CORI; <http://www.cori.org>) that began in 1995 under the auspices of the American Society for Gastrointestinal Endoscopy (ASGE). Members of the CORI consortium produce endoscopy reports using a computerized report generator that allows for collection of structured data. A limited data set is then sent to National Endoscopic Database (NED) weekly. The aggregate data are used for a variety of research purposes including study of endoscopic practice patterns and determination of intraprocedural complications. Fortunately, an increasing number of EHR vendors have started to include EEMR-like functions within their EHRs or are providing a tighter integration with existing EEMR systems that can facilitate research on long-term outcomes and effectiveness of procedures (42, 45).

Cost of Health Information Technology

It has been estimated that universal EHR adoption and interoperability will cost the United States \$156 billion in capital investment over 5 yr and \$48 billion in annual operating costs (46). The high upfront cost of investment is one of the main reasons cited for slow EHR adoption in the United States (47). Many public and private initiatives are currently underway to address this problem of misaligned incentives, but it remains to be seen if these initiatives will be enough to support universal adoption of EHRs by 2014.

Patient Confidentiality and Privacy

The potential for misuse of personal information has heightened concern about the confidentiality of EHR-derived data for research. Integration of genomics data to EHR data poses additional concerns regarding abuse of the data by insurers or employers to discriminate in coverage, premiums, or hiring (48). The current policies and procedures ensured to deidentify or anonymize the data may not be sufficient to prevent unintentional leaks of data due to security breaches in poorly designed EHRs (49). Proper safeguards during design and implementation of EHRs and having a monitored framework can assure that privacy and confidentiality is maintained while data from EHRs is utilized for the purpose of scientific research.

CONCLUSIONS

Numerous efforts to accelerate the adoption of EHRs are currently underway in both the public and private sectors with the goal of universal EHR adoption by the year 2014. EHR-derived data can allow researchers to focus their efforts toward research design and analysis rather than manual data gathering. EHRs have the potential to transform the research infrastructure of the 21st century by allowing input of clinical data in a structured format conforming to the national standards, facilitating electronic data capture for clinical trials, and providing linkage with genomic information. The research community needs to become aware of emerging technologies such as EHRs and get involved in adoption and advocacy efforts in order to leverage the opportunities for GI research.

Reprint requests and correspondence: Ashish Arteja, MD, MPH, Informatics Consultant, Cleveland Clinical and Translational Science Award, Associate Program Director, Medical Informatics Fellowship, Associate Staff, Medicine Institute, Cleveland Clinic, Desk A-31, 9500 Euclid Ave, Cleveland Clinic Foundation, Cleveland, OH 44195.

REFERENCES

1. Bush GW. Executive Order of the President of the United States. 2004. Available at: <http://www.whitehouse.gov/news/releases/2004/04/20040427-4.html>. Accessed July 15, 2007.

2. Brailer D. David Brailer on a private-public health information technology infrastructure. Interview by Susan V. White. *J Healthc Qual* 2004;26:20–4.
3. Brewin B. HHS pushes electronic health records. *Computerworld*. July 21, 2004. Available at <http://www.computerworld.com/governmenttopics/government/policy/story/0,10801,94665,00.html>. Accessed March 31, 2008.
4. Tang P. Key capabilities of an electronic health record system. Institute of Medicine Committee on data standards for patient safety. Board on Health Care Services. Washington D.C.: National Academies Press, 2003.
5. Cimino JJ, Li J, Graham M, et al. Use of online resources while using a clinical information system. *AMIA Annu Symp Proc* 2003;175–9.
6. Oren E, Shaffer ER, Guglielmo BJ. Impact of emerging technologies on medication errors and adverse drug events. *Am J Health Syst Pharm* 2003;60:1447–58.
7. Bates DW, Teich JM, Lee J, et al. The impact of computerized physician order entry on medication error prevention. *J Am Med Inform Assoc* 1999;6:313–21.
8. Chertow GM, Lee J, Kuperman GJ, et al. Guided medication dosing for inpatients with renal insufficiency. *JAMA* 2001;286:2839–44.
9. Poon EG, Kuperman GJ, Fiskio J, et al. Real-time notification of laboratory data requested by users through alphanumeric pagers. *J Am Med Inform Assoc* 2002;9:217–22.
10. Garg AX, Adhikari NK, McDonald H, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: A systematic review. *JAMA* 2005;293:1223–38.
11. Kucher N, Koo S, Quiroz R, et al. Electronic alerts to prevent venous thromboembolism among hospitalized patients. *N Engl J Med* 2005;352:969–77.
12. Johanson JF. Quality and outcomes management in gastroenterology. *Gastroenterol Clin North Am* 1997;26:859–71.
13. Bickston SJ, Alley J, Robin DP. Best practices: Academic gastroenterology practices. *Clin Gastroenterol Hepatol* 2006;4:1415–8.
14. Sarfaty M, Wender R. How to increase colorectal cancer screening rates in practice. *CA Cancer J Clin* 2007;57:354–66.
15. Safran C, Bloomrosen M, Hammond WE, et al. Toward a national framework for the secondary use of health. Report of a working conference of the American Medical Informatics Association. 2006. Available at: http://www.amia.org/inside/initiatives/healthdata/finalpapertowardanationalframeworkforthesecondaryuseofhealthdata_09_08_06_.pdf. Accessed July 15, 2007.
16. Zhan C, Miller MR. Administrative data-based patient safety research: A critical review. *Qual Saf Health Care* 2003;12:II58.
17. Feinstein AR. ICD, POR and DRG: Unsolved scientific problems in the nosology of clinical medicine. *Arch Intern Med* 1988;148:2269–74.
18. Jollis JG, Ancukiewicz M, DeLong ER, et al. Discordance of databases designed for claims payment versus clinical information systems. Implications for outcomes research. *Ann Intern Med* 1993;119:844–50.
19. Peabody JW, Luck J, Jain S, et al. Assessing the accuracy of administrative data in health information systems. *Med Care* 2004;42:1066–72.
20. Abraham NS, Cohen DC, Rivers B, et al. Validation of administrative data used for the diagnosis of upper gastrointestinal events following nonsteroidal anti-inflammatory drug prescription. *Aliment Pharmacol Ther* 2006;24:299–306.
21. Benin AL, Vitkauskas G, Thornquist E, et al. Validity of using an electronic medical record for assessing quality of care in an outpatient setting. *Med Care* 2005;43:691–8.
22. Tang PC, Ralston M, Arrigotti MF, et al. Comparison of methodologies for calculating quality measures based on administrative data versus clinical data from an electronic health record system: Implications for performance measures. *J Am Med Inform Assoc* 2007;14:10–5.
23. Mann CJ. Observational research methods. Research design II: Cohort, cross sectional, and case-control studies. *Emerg Med J* 2003;20:54–60.
24. Rubin GP, Hungin AP, Kelly PJ, et al. Inflammatory bowel disease: Epidemiology and management in an English general practice population. *Aliment Pharmacol Ther* 2000;14:1553–9.
25. Stone MA, Mayberry JF, Baker R. Prevalence and management of inflammatory bowel disease: A cross-sectional study from central England. *Eur J Gastroenterol Hepatol* 2003;15:1275–80.
26. Graham DJ, Campen D, Hui R, et al. Risk of acute myocardial infarction and sudden cardiac death in patients treated with cyclo-oxygenase 2 selective and non-selective non-steroidal anti-inflammatory drugs: Nested case-control study. *Lancet* 2005;365:475–81.
27. Laheij RJ, Sturkenboom MC, Hassing RJ, et al. Risk of community-acquired pneumonia and use of gastric acid-suppressive drugs. *JAMA* 2004;292:1955–60.
28. Van Soest EM, Dieleman JP, Siersema PD, et al. Increasing incidence of Barrett's oesophagus in the general population. *Gut* 2005;54:1062–6.
29. Levin TR, Zhao W, Conell C, et al. Complications of colonoscopy in an integrated health care delivery system. Cohort Studies and GI Disease-specific Registries. *Ann Intern Med* 2006;145:880–6.
30. VHA Directive 2001-009. National Hepatitis C Program. 2001. Available at: www.publichealth.va.gov/documents/directives/12002022.doc. Accessed July 15, 2007.
31. Marks L, Power E. Using technology to address recruitment issues in the clinical trial process. *Trends Biotechnol* 2002;20:105–9.
32. Hanna KE. Think research: Using electronic medical records to bridge patient care and research. 2005. Available at: http://www.fastercures.org/pdf/emr_whitepaper.pdf. Accessed March 31, 2008.
33. Shortliffe EH. The evolution of health-care records in the era of the Internet. *Medinfo* 1998;9(pt1 Suppl):8–14.
34. Harris PA, Lane L, Biaggioni. Clinical research subject recruitment: The Volunteer for Vanderbilt Research Program www.volunteer.mc.vanderbilt.edu. *J Am Med Inform Assoc* 2005;12:608–13.
35. Embi PJ, Jain A, Clark J, et al. Effect of a clinical trial alert system on physician participation in trial recruitment. *Arch Intern Med* 2005;165:2272–7.
36. Avorn J. Evaluating drug effects in the post-vioxx world: There must be a better way. *Circulation* 2006;113:2173–6.
37. Garcia Rodriguez LA, Perez Gutthann S. Use of the UK general practice research database for pharmacoepidemiology. *Br J Clin Pharmacol* 1998;45:419–25.
38. Wood L, Martinez C. The general practice research database: Role in pharmacovigilance. *Drug Saf* 2004;27:871–81.
39. Yang YX, Lewis JD, Epstein S, et al. Long-term proton pump inhibitor therapy and risk of hip fracture. *JAMA* 2006;296:2947–53.
40. Adler KG. How to select an electronic health record system. *Fam Pract Manag* 2005;12:55–62; Available

- at: <http://www.aafp.org/fpm/20050200/55howt.html>. Accessed Jan 1, 2007.
41. Wasserman H, Wang J. An applied evaluation of SNOMED CT as a clinical vocabulary for the computerized diagnosis and problem list. *AMIA Annu Symp Proc* 2003;699–703.
 42. Enns RA, Barkun AN, Gerdes H. Electronic endoscopic information systems: What is out there? *Gastrointest Endosc Clin N Am* 2004;14:745–54.
 43. LinksSavides TJ, Chang K, Cotton P. Possible features of current electronic endoscopic information systems: What to look for. *Gastrointest Endosc Clin N Am* 2004;14:735–43.
 44. Soekhoe JK, Groenen MJ, van Ginneken AM, et al. Computerized endoscopic reporting is no more time-consuming than reporting with conventional methods. *Eur J Intern Med* 2007;18:321–5.
 45. Yousfi M, Gostout CJ, Baron TH, et al. Postpolypectomy lower gastrointestinal bleeding: Potential role of aspirin. *Am J Gastroenterol* 2004;99:1785–9.
 46. Kaushal R, Blumenthal D, Poon EG, et al. The costs of a national health information network. *Ann Intern Med* 2005;143:165–73.
 47. Middleton B, Hammond WE, Brennan PF, et al. Accelerating U.S. EHR adoption: How to get there from here. Recommendations based on the 2004 ACMI retreat. *J Am Med Inform Assoc* 2005;12:13–9.
 48. Melton LJ. The threat to medical records research. *N Engl J Med* 1997;337:1466–70.
 49. Win KT. A review of security of electronic health records. *HIMJ* 2005;34:13–8.

CONFLICT OF INTEREST

Guarantor of the article: Ashish Atreja.

Specific author contributions: Ashish Atreja: conception and design, literature review, and drafting and final approval of the article; Anil Jain: literature review and drafting of the article (section “conceptual overview of EHR”); Jean-Paul Achkar: drafting of the article (sections “GI and observational studies” and “GI and clinical trials”); C. Martin Harris: critical revision and final approval of the article; and Bret A. Lashner: conception and design, literature review, and final approval of the article.

Financial support: None.

Potential competing interests: None.
