

# Accelerating Biopharmaceutical Development in the Decade of Health Information Technology

## Applications of EHRs for outcomes research and clinical trials recruitment

Michael I. Lieberman<sup>1</sup>; Peter Embi, MD, MS<sup>2</sup>; Thomas N. Ricciardi, PhD<sup>1</sup>; Kevin Tabb, MD<sup>3</sup>

<sup>1</sup>GE Healthcare & Oregon Health and Science University, <sup>2</sup>University of Cincinnati Medical Center, <sup>3</sup>Stanford University Medical Center

The U.S. Department of Health and Human Services declared the start of a “decade of health information technology” in 2004. Under the leadership of the HHS Office of the National Coordinator of Health Information Technology (ONCHIT), a goal of this directive is for every patient in the United States to have an electronic health record (EHR) within 10 years. This goal is part of a promise to completely change how health information is managed. A key expectation of the National Health Information Network (NHIN) is to use the EHR to create streams of clinical information suitable for research, developing therapies, and improving population health. Clearly, there will be

value in the detailed clinical information that is gathered during the process of care; the proposed information technology overhaul in American healthcare could transform the way research is conducted.

There is vast potential to use clinical information derived from the EHR as a way to streamline patient recruitment for clinical trials. We also focus on the potential to use such information for prospective studies at any stage of development, as well as for postmarketing surveillance of therapies for safety, efficacy, and cost.

### TOOL TO FACILITATE RECRUITMENT

The growing pace of biomedical discoveries, their potential health benefits, and their high costs make the need to conduct clinical trials more pressing now than ever before (Rindfleisch 1998). Of all trial designs, prospective clinical trials yield the strongest evidence regarding the utility of the intervention being studied; in addition, they are essential to the biopharmaceutical development and approval process. Clinical trials often are difficult to conduct, how-

ever. Recruitment of eligible subjects represents a major bottleneck to the successful conduct of such studies, with only a small fraction of those eligible being referred for consideration. Recently enacted regulations, like the Health Insurance Portability and Accountability Act of 1996 (HIPAA) create additional obstacles and can delay this process further. Methods to improve recruitment efforts are sorely needed, and the EHR holds great promise as a tool to help in this regard.

### SOURCE OF RETROSPECTIVE REAL-WORLD CLINICAL DATA

The EHR plays two interrelated roles in facilitating clinical trial patient recruitment. First, existing data in an EHR can be used to expedite identification of prospective subjects. Second, the EHR can serve as a workflow engine to facilitate the evaluation and enrollment of patients. To understand how the EHR can play the first role, one must understand the characteristics of electronic clinical data.

The primary purpose of clinical data in an EHR is to assist a clinician in providing medical care. These

#### Author correspondence

**Thomas N. Ricciardi, PhD**  
GE Healthcare  
Information Technologies  
Manager, Clinical Data Services  
Medical Quality Improvement  
Consortium  
tom.ricciardi@med.ge.com  
www.gehealthcare.com

20540 NW Evergreen Parkway  
Hillsboro, OR 97124 USA

data can take a variety of forms to fulfill this function. The EHR can offer a significant benefit just by providing easily accessible, legible patient data at the point of care. Yet, to provide even this benefit, patient data must be entered into the system, and the clinician must perform this task. If data entry is too onerous, an EHR may not be adopted at all, or the data entry may be incomplete. In general, clinicians have been given fairly wide latitude in how they enter data into an EHR. While this practice has been necessary to enable the adoption of an EHR, the lack of constraints on data entry does make data analysis more difficult.

*Data quality.* EHR data can be classified into four groups: narrative free text, discrete text, discrete coded, and numeric. Narrative free text is the most common form and may be entered directly by a clinician, or it can be imported as a transcribed note. Documentation of an entire encounter, or such specific parts as "history of present illness," may be recorded in narrative free text, while other elements may be more structured. With discrete text, the user is limited to predefined choices. It might be entered from a pick list, or a series of check boxes, or buttons. Discrete coded data are similar to discrete text in that a user is limited to a predefined set of values. With coded data, however, each value is a member of a defined code set.

As EHR adoption increases, and the value of clinical information becomes more apparent, EHR users will become more aware of data quality. EHR data quality depends on three elements. First, the EHR must have the capacity to capture discrete data elements, and it should

have the ability to restrict the range of entries for certain elements. Second, implementation and administration of the EHR must be done carefully because they can have a significant effect on data quality. Third, the clinician must use the EHR appropriately.

Highly structured data can en-



**Michael I. Lieberman**



**Peter Embi, MD, MS**



**Thomas N. Ricciardi, PhD**



**Kevin Tabb, MD**

hance data quality, and, in general, it will. Nevertheless, structured data alone cannot guarantee data quality, and it is indeed possible that overly structured data could hurt data quality. For example, if a clinician is limited to a set of values for a finding, and none of the possible values accurately reflects the true clinical event, the clinician must decide either not to record any value or to choose an inaccurate representation of the finding.

*Strengths of EHR data.* For data analysis in general and clinical trial patient recruitment in particular,

discrete text, numeric, and coded data are utilized most easily. These types of data are found in vital signs and lab results (numeric), and diagnosis and medication lists. There can be significant variability in the structure of diagnosis and medication lists. These lists can be entered and recorded as free text, but more often they are recorded with some coding system. At present, most diagnoses in the United States are coded with ICD9-CM. Because of the lack of sufficient granularity in ICD9-CM, SNOMED-CT is gaining acceptance as a superior system for coding diagnoses.

If medications are coded, it is almost always with a proprietary coding system supplied by a drug-data vendor. Because they are proprietary, it is more difficult to make use of these data; though a small number of vendors (fewer than five) control most of the market, and, therefore, there are a limited number of code sets to account for.

To fully interpret a patient medication, in addition to understanding what medication is being taken, one must also understand how that medication is being taken, which is commonly referred to as the *sig*. The *sig* includes the dose (how many tablets, capsules, milliliters, etc.), route (orally, topically, etc.), frequency (once daily, before meals, etc.), and any other special instructions associated with a prescription. Each of these elements can be captured discretely, but often the entire *sig* is entered as free text, and therefore can be difficult to interpret.

*Making data useful for retrospective research.* To make clinical data available for widespread research, one must address privacy and security

concerns due to the sensitive nature of these data. Unlike clinical trial data, where subjects give explicit consent for researchers to collect and interpret their health data, explicit consent for using data retrospectively is not obtained routinely. Instead, data is de-identified in accordance with HIPAA. The privacy rule of this act stipulates that data can be disclosed without restriction, but only if certain criteria for de-identification are met. The privacy rule includes 18 specific elements, including but not limited to, names, telephone numbers, and dates that should be removed from data to be shared (NIH 2004).

Once data are available for retrospective analysis, then, depending on the quality of the data, additional processing may be required before the information can be analyzed. For numeric data, this cleanup may include stripping out text when doing so does not affect the meaning of the element; checking number formats (one would not expect numbers after a decimal point in blood pressures); and making sure values fall within a reasonable range.

For medications and diagnoses, data cleanup may include mapping the entry to a terminology such as ICD-9CM or SNOMED-CT, or verifying that codes are valid. If data are encoded in a proprietary system, that information may need to be converted to a common system if the data will be aggregated with other proprietary information.

*Issues with EHR data.* A major challenge in using clinical data for research is properly interpreting free text narrative data. Despite increasing awareness of the value of coded data, there continues to be an abundance of information locked up in

free text narrative. This information always will have a place in painting a clinical picture that cannot be fully communicated by coded elements. There is inherent value in allowing a clinician to describe a history and symptoms in a narrative form. Yet there is also value in extracting critical information from that narrative. Having a person read through these entries and assign appropriate codes is extremely resource intensive and not realistic on a large scale. An alternative is to use natural language processing (NLP) techniques to extract information in an automated fashion. Many studies have shown that NLP can be used with precision, but with relatively poor recall (Melton 2005, Huang 2005, Chapman 2005). Therefore, while data derived from NLP might not give a complete clinical picture, they could be useful for identifying patients meeting certain inclusion criteria for a clinical trial.

### FACILITATING RECRUITMENT AT POINT OF CARE

Over the past 20 years, much work has been done to address issues involved in clinical trials recruitment. Efforts to improve the awareness of clinical trials among physicians, patients, and the general public have been pursued, ranging from the distribution of paper and electronic fliers by trial centers and consumer advertising to the use of government and privately sponsored Web sites. In addition, several computer programs help to match patients with trials.\* Such efforts have shown promise and are important for some populations,

\*Breitfeld 1999, Ash 2001, Papaconstantinou 1998, Thompson 2003, Tu 1993, Ohno-Machado 1999, Fink 2004.

though few controlled studies demonstrate their effectiveness at improving trial-accrual rates.

Attempts to leverage the power of existing clinical data repositories to identify eligible subjects have been described as well. While approaches that employ data mining for mass screening hold promise, such efforts are limited somewhat by privacy regulations and have not yet been reported to enhance recruitment efforts. As researchers have stated for more than a decade, the value of such repositories for patient recruitment likely lies in linking them to integrated, computer-based medical record systems (Musen 1992, Carlson 1995). Recently, the alternative approach of linking clinical databases to alert systems to notify physicians or researchers of potentially eligible patients has shown promise.<sup>†</sup>

EHR systems allow for point-of-care processes that appear to be more effective than previous approaches. In fact, one of the first reports to document a significant benefit to trial recruitment rates was that of a new EHR-based Clinical Trial Alert (CTA) system (Embi 2005). By repurposing an existing EHR's clinical decision support and communication capabilities, the CTA system provided point-of-care physician alerting when a patient met designated trial eligibility criteria, and facilitated sending a secure message to the trial's coordinator during the course of a patient visit. The CTA worked much as described in the following workflow scenario:

A patient visits his primary care physician for a routine check-up. When the physician opens the patient's electronic

<sup>†</sup>Weiner 2003, Gennari 2001, Moore 2003, Afrin 2003.

chart, the EHR evaluates basic key patient attributes, such as diagnoses, to see if the patient may be eligible for any clinical trial entered into the EHR system. If the patient matches criteria for any of the studies, a popup screen will appear notifying the clinician which trials the patient may be eligible for. The clinician can decline further questioning and continue with the patient's visit, or can elect to proceed if both are interested. The EHR would prompt the clinician to collect more specific information to determine eligibility. If, after this information is collected, the patient is still eligible, an electronic notification will be sent to a study coordinator to follow up with the patient. When the coordinator receives the notification, he can use the EHR to do further chart review. If the patient is still a candidate, the coordinator can finalize eligibility and enroll the patient in the study.

In addition to increasing significantly the number of physicians who participated in trial recruitment and encouraging recruitment by primary care providers who had not previously recruited a patient, CTA use in the reported study led to a significant increase in referral and enrollment rates. Moreover, by placing the initial recruitment event at the physician-patient encounter stage, and requiring patient authorization prior to contacting the clinical trial's coordinator, the approach demonstrated its benefit in a HIPAA-compliant fashion (Embi 2005). While specific workflows may vary, depending on the capabilities of the particular EHR and how it is implemented, a CTA process should be possible to replicate in many available comprehensive EHR systems.

Given the growing implementation rate of comprehensive EHRs with such capabilities as computerized provider order entry and computerized decision support sys-

tems, which have demonstrated benefits to patient safety and healthcare quality, functionality such as that achieved by the CTA soon may be possible on a wider basis (Ash 2005, Bates 2003). As healthcare institutions engaged in clinical research activities begin to consider other uses for these systems, the ability to leverage them for clinical trial recruitment will undoubtedly

## Economic forces are pushing healthcare toward the “tipping point”—the widespread use of EHR that could create an effective platform for clinical trial recruitment.

grow and help to overcome the many obstacles facing the clinical research enterprise.

### TECHNICAL/POLICY ISSUES

In the United States, the ONCHIT Strategic Framework lists several goals related to deploying EHRs in the physician office, connecting systems across communities and the nation, wiring government health-related systems, and improving population health with research and quality-improvement efforts. Nevertheless, EHRs have been deployed in only 10 to 15 percent of physician offices, so most health information remains on paper rather than in a useful digital format. The Regional Health Information Organizations (RHIOs) and data exchanges envisioned by ONCHIT do not exist, except in pilot mode in selected communities. Needed data standards are in development — the Health Level 7 Clinical Document Architecture

(CDA), the Continuity of Care Record (CCR), and terminology standards such as SNOMED and RxNORM for clinical data are maturing for healthcare only now. Researchers in the pharma industry will require data from EHR systems in the format defined by the Clinical Data Interchange Standards Coalition (CDISC) format; these specifications are gaining wider acceptance.

For public or population health, Public Health Data Standards Consortium (PHDSC) are making progress to define the core data models and standards required to communicate information derived from the clinical domain.

Many stakeholders in both information technology and healthcare delivery are focused on improving the landscape to foster rapid adoption of EHR by providers. For instance, Integrating the Healthcare Enterprise (IHE) is focused on standards for interoperability; a newly formed EHR vendor association has, as its mission, the effective development and marketing of healthcare information technology. Many public-sector organizations and not-for-profit groups — such as the Markle Foundation, e-Health Initiative, and ONCHIT — foster dialog and collaboration among the stakeholders required to bring the right tools to clinical medicine and to help digitize healthcare.

Perhaps most importantly, the economic incentives required to accelerate physician adoption of EHRs are slowly taking shape. Taking its cue from payers and employers in Bridges to Excellence and the Leapfrog Group, the Center for

Medicare and Medicaid Services is sponsoring several pilot studies to implement and use EHR to improve and measure the quality of care in 2005. If successful, these pilots will inform the rollout of pay-for-performance programs in the near future.

These forces are helping health-care move toward the "tipping point," to allow the successful, widespread implementation of EHR that also would be an effective platform for clinical research and data capture.

## CONCLUSIONS

Clinical trial recruitment is becoming increasingly difficult for sponsors (Centerwatch 2002). It is estimated that nearly 50 percent of all delays in drug development are attributable to problems with recruitment, which also is growing ever more expensive. The EHR offers a novel, possible solution for recruitment bottlenecks.

Clearly, the EHR is not a panacea for the biopharmaceutical industry's problems with clinical trials. Adoption is not yet universal; the quality of data entered varies by system and individual user; national standards are in process but not yet implemented; and empirical evidence of the utility of the EHR in a clinical trial environment is still lacking. Additionally, most EHR systems now in place are designed for clinical documentation and not as a replacement for case report forms used in many clinical trials. It is generally accepted, however, that EHR adoption will be widespread in this country within a few years.

Marrying EHR tools already used by physicians at the point of care with trial recruitment tools on the

physician's desktop surely will have an advantage over tools posed extraneously to the physician workflow. In the increasingly data-rich environment of personalized medicine, technologies that allow users to match phenotype and genotype information in a single database will improve recruitment capabilities. The EHR is one such technology that promises to help. Those systems that can solve the issues raised above most quickly should be added to the arsenal of tools used by the biopharmaceutical industry in conducting trials.

## REFERENCES

- Afrin LB, Oates JC, Boyd CK, Daniels MS. Leveraging of open EMR architecture for clinical trial accrual. *AMIA Annu Symp Proc.* 2003;16–20.
- Ash JS, Bates DW. Factors and forces affecting EHR system adoption: report of a 2004 ACMI discussion. *J Am Med Inform Assoc.* 2005;12:8–12.
- Ash N, Ogunyemi O, Zeng Q, Ohno-Machado L. Finding appropriate clinical trials: evaluating encoded eligibility criteria with incomplete data. *Proc AMIA Symp.* 2001;27–31.
- Bates DW, Gawande AA. Improving safety with information technology. *N Engl J Med.* 2003;348:2526–2534.
- Breitfeld PP, Weisburd M, Overhage JM, et al. Pilot study of a point-of-use decision support tool for cancer clinical trials eligibility. *J Am Med Inform Assoc.* 1999;6:466–477.
- Carlson RW, Tu SW, Lane NM, et al. Computer-based screening of patients with HIV/AIDS for clinical-trial eligibility. *Online J Curr Clin Trials.* 1995 Mar 28; Doc. no. 179.
- CenterWatch. Sponsors take on patient recruitment. *CenterWatch Newsletter.* 2002;9(3):11.
- Chapman WW, Christensen LM, Wagner MM, et al. Classifying free-text triage chief complaints into syndromic categories with natural language processing. *Artif Intell Med.* 2005;33:31–40.
- Embi PJ, Jain A, Clark J, et al. Effect of a clinical trial alert system on internist participation in trial recruitment. *J Gen Intern Med.* 2005;20(suppl 1):96.
- Fink E, Kokku PK, Nikiforou S, et al. Selection of patients for clinical trials: an interactive web-based system. *Artif Intell Med.* 2004;31:241–254.
- Gennari JH, Sklar D, Silva J. Cross-tool communication: from protocol authoring to eligibility determination. *Proc AMIA Symp.* 2001:199–203.
- Huang Y, Lowe HJ, Klein D, Cucina RJ. Improved identification of noun phrases in clinical radiology reports using a high-performance statistical natural language parser augmented with the UMLS specialist lexicon. *J Am Med Inform Assoc.* 2005;12:275–285.
- Melton GB, Hripcsak G. Automated detection of adverse events using natural language processing of discharge summaries. *J Am Med Inform Assoc.* 2005 (in press); epub ahead of print March 31, 2005.
- Moore TD, Hotz K, Christensen R, et al. Integration of clinical trial decision rules in an electronic medical record (EMR) enhances patient accrual and facilitates data management, quality control, and analysis. *Proc Am Soc Clin Oncol.* 2003;22:557. Abstract 2242.
- Musen MA, Carlson RW, Fagan LM, et al. T-HELPER: automated support for community-based clinical research. *Proc Annu Symp Comput Appl Med Care.* 1992:719–723.
- NIH (National Institutes of Health). How Can Covered Entities Use and Disclose Protected Health Information for Research and Comply with the Privacy Rule? Bethesda, Md.: NIH. 2004. Available at: «[http://privacyruleandresearch.nih.gov/pr\\_08.asp#8a](http://privacyruleandresearch.nih.gov/pr_08.asp#8a)». Accessed May 15, 2005.
- Ohno-Machado L, Wang SJ, Mar P, Boxwala AA. Decision support for clinical trial eligibility determination in breast cancer. *Proc AMIA Symp.* 1999:340–344.
- Papaconstantinou C, Theocharous G, Mahadevan S. An expert system for assigning patients into clinical trials based on Bayesian networks. *J Med Syst.* 1998;22:189–202.
- Rindfleisch TC, Brutlag DL. Directions for clinical research and genomic research into the next decade: implications for informatics. *J Am Med Inform Assoc.* 1998;5:404–411.
- Thompson DS, Oberteuffer R, Dorman T. Sepsis alert and diagnostic system: integrating clinical systems to enhance study coordinator efficiency. *Comput Inform Nurs.* 2003;21:22–26.
- Tu SW, Kemper CA, Lane NM, et al. A methodology for determining patients' eligibility for clinical trials. *Methods Inf Med.* 1993;32:317–325.
- Weiner DL, Butte AJ, Hibberd PL, Fleisher GR. Computerized recruiting for clinical trials in real time. *Ann Emerg Med.* 2003;41:242–246.