

# *Executive Summary*

*Think Research  
Using Electronic Medical  
Records to Bridge Patient  
Care and Research*





# Introduction

## **Letter from Greg Simon, President of FasterCures**

*At FasterCures our mission is to save lives by saving time in the research and development of new cures for deadly and debilitating diseases. It is our role to ask “what if” and then see what it would take to realize those ideas that can accelerate the availability of new medical solutions.*

*That is why we have been working with leaders of organizations dedicated to realizing the healthcare and health research benefits of information technology through the work of the National Health Information Network (NHIN).*

*In order to understand where we need to go in using electronic medical records (EMRs) in research, it is important to know where we are. We commissioned this paper to understand the current landscape of EMRs.*

*For a long time, we have been promised that information technology could save lives and accelerate research. Now, the advent of the NHIN will make it possible to study large samples of medical records in order to understand disease progression and accelerate the development of treatment options.*

*At institutions such as the Regenstrief Institute at the Indiana University School of Medicine, EMRs have been useful for a variety of studies and for enhancing clinical trials data. The Mayo Clinic has completely converted to an EMR system, and research was one of their primary reasons for doing so. The Veterans Health Administration’s VistA system contains a variety of electronic databases predominantly accessed by Department of Veterans Affairs researchers to conduct health services research.*

*A sense of urgency and momentum is finally developing around the issue of creating a comprehensive, interoperable system of EMRs to improve patient care and reduce costs. EMRs hold just as much promise for speeding the discovery of new therapies by allowing researchers access to the precious data contained in them.*

*It is important that public- and private-sector leaders understand up front, as these systems are developed, the promise they hold for research so that appropriate steps will be taken to facilitate that use.*

*FasterCures believes this study can be the first step in ensuring that the special needs and challenges of the research community are an integral part of that conversation.*

A handwritten signature in black ink, appearing to read "Greg", located in the bottom right corner of the page.

*The advent of the NHIN will make it possible to study large samples of medical records in order to understand disease progression and accelerate the development of treatment options.*

The frontier of medical science has rarely been as exciting and as full of opportunity as it is today. From basic science through clinical research to health services research, the opportunities made available through the impressive advances of recent decades in the biomedical, physical, computational, and behavioral and social sciences have brought us to a place of unprecedented opportunity.

Successful completion of the Human Genome Project has launched scientists into the new world of genomics and proteomics, a journey that will lead to the eventual understanding of every human protein and its role in both health and disease. The next challenge will be to identify the functions of the genes that have been mapped, the proteins for which they code, and the functions of those proteins in the body.

To understand the relationships between this molecular information and human health, population-based and clinical studies are needed, entailing the generation, storage, and analysis of enormous quantities of epidemiologic, genotypic, and phenotypic data. To understand the connections between genes, proteins, and the environment, sophisticated comparisons must be conducted, and these comparisons cannot be done by hand or by eye or patient by patient. It is the collective observations of hundreds, even thousands, of patients that will shine a light on these associations.

## ***The EMR: An Electronic Bridge Between Medicine and Science***

The development of EMR systems presents a unique opportunity to support and further the nation's health research enterprise. To date, the utility of health information networks has been seen as related primarily to reducing healthcare costs, limiting medical errors, and generally improving the standard of care. While these benefits are important, there is another critical element in the healthcare continuum that could greatly benefit from the development of EMR systems: medical research.

Studying large samples of medical records or clinical datasets could be an essential step toward understanding the etiology and progression of disease, treatment methods, and outcomes across varied populations and disease groups. To understand the landscape of EMR system adoption better and to evaluate the challenges and opportunities involved in developing research uses of this vast resource, *FasterCures* commissioned this study to:

- 1** conduct a broad overview and characterization of current efforts to promote EMRs and their potential research use, and
- 2** assess what is needed to optimize the creation and use of such databases for research purposes.

This analysis rests on the premise that as the healthcare system addresses the challenges of widespread adoption of EMR systems, research capacity should be a part of the architecture.

## ***EMR Systems Are a Critical Research Resource***

Datasets generated from patient care have always been a research resource of enormous potential. Historically, paper medical records studies have been used to:

- monitor the health of the population and detect emerging health problems;
- identify populations at high risk for disease;

- determine the effectiveness of treatment(s);
- quantify prognoses;
- assess the usefulness of diagnostic tests and screening programs;
- influence policy through cost-effectiveness analysis;
- support administrative functions; and
- monitor the adequacy of care.

The trend toward the capture and storage of patient information in digital form provides researchers with a potentially more efficient and effective means of accessing these data to:

- form hypotheses about disease initiation and progression;
- look for patterns of health and illness in a given population;
- conduct post-marketing surveillance studies of new drugs to identify adverse events or improve prescribing and labeling practices; and
- most importantly, identify potential study participants for clinical research.

There are practical benefits to the research use of EMRs as well.

- Much of the patient information collected for clinical trials already exists in the patient record. If clinical researchers could quickly import such information (especially if it is standardized) from the existing practice record into the research record, both time and money could be saved.
- EMR systems could speed data acquisition and searching, allow mass computing and sampling, and provide the research community access to a broader and more diverse patient population.

## ***EMR Systems of Today: Pioneers and Innovators***

Although most healthcare practitioners and institutions in the United States are not yet ready to implement EMR systems, there are a few pioneers using EMR systems for research purposes. Not willing to wait for the NHIN to arrive, several institutions have overcome enormous obstacles and forged ahead to find ways to meld clinical practice data with research goals.

This goal is not easily accomplished, because most clinical EMR systems are designed to support clinical workflow, not research. Numerous academic medical centers are in the early stages of EMR adoption generally, and many are contemplating the value of EMRs as a research resource.

**Mayo Clinic.** The Mayo Clinic has converted completely to an EMR system as of July 2004 and has been using paper-based medical records in research for more than 80 years. Mayo conducts more than 4,000 clinical trials each year, and nearly every trial relies on information from medical records. In addition, researchers from IBM and Mayo are using supercomputing technology and applying customized algorithms, data mining, and pattern recognition to uncover correlations between particular proteins, genetic markers, patient outcomes, and other factors that could lead to new diagnostics and treatments.

**Regenstrief Institute.** Over the past three decades, Regenstrief has developed one of the nation's first EMR systems, along with the nation's only citywide EMR system, which allows doctors in ERs to view as a single record all previous care at any of 11 hospitals. The records have been useful for prospective, retrospective, epidemiological, longitudinal, and cohort studies, and for enhancing clinical trials datasets.

**Kaiser Permanente.** Kaiser has made a \$3 billion investment to automate records for its 8.4 million members. It has created both a transactional system for patient care and a data warehouse<sup>1</sup> with extracted data for use by researchers. Kaiser envisions using the data warehouse for planning studies; building complex physiologic or patient care models; identifying potential study participants; conducting post-marketing surveillance studies; combining clinical data with genomics data to identify genetic factors of disease; and providing insight into the effectiveness of various interventions.

**Indian Health Service.** IHS has a national data repository that accepts exports from other sites and is evolving into a true National Data Warehouse (NDW). On a national level, the NDW will offer a number of standard and customized “data marts” for appropriately qualified users. IHS facilities are active in many types of clinical research, often in collaboration with universities or government agencies.

**Veterans Health Administration.** VA’s pioneering health information system, VistA, is in the process of evolving into a new Web-based health data repository that would create a true longitudinal healthcare record including data from VA and non-VA sources, supporting research and population analyses, improving data quality and security, and facilitating patient access to data and health information.

**Collaborations in HIV Outcomes Research/United States (CHORUS).** CHORUS is a database constructed specifically for supporting both clinical care and research. It contains observational data on about 4,000 patients in four major U.S. HIV practices. The observational data collected through programs such as CHORUS are important because of the limitations of conducting randomized clinical trials. Because observational trials record real-world experience, they reflect prevailing medical practice and could be more likely to uncover treatment approaches that are realistic and feasible.

**Partners HealthCare Research Patient Data Registry (RPDR).** The RPDR is a data repository of information on two million Massachusetts General Hospital and Brigham and Women’s Hospital patients accumulated since the 1980s. Many believe it is likely to serve as a good first model for best practices in medical informatics and research data mining. The RPDR has a number of significant technical and procedural features that make it better than the more commercial offerings as a potential research tool. Its operations are tightly integrated with an Institutional Review Board (IRB), and the data available to researchers are filtered with two-way anonymization as they are included in the system, so that records can be searched with identifiable patient data when necessary.

**General Practice Research Database (GPRD).** The United Kingdom’s GPRD is the world’s largest computerized database of anonymized longitudinal medical records collected from primary care; currently data are being collected on more than three million active patients, and data are updated once every two weeks. The GPRD is a valuable tool for academic research in a broad range of areas, including clinical epidemiology, disease patterns, disease management, outcomes research, and drug utilization. More than 400 research papers have been published using the database.

**Directory of Clinical Databases (DoCDat).** Created by the London School of Hygiene and Tropical Medicine, the online DoCDat provides a directory of clinical databases in the U.K., along with up-to-date information on those databases, in order to enable greater access to and use of existing clinical databases and to improve their quality. DoCDat provides independent evidence regarding their uses and limitations.

<sup>1</sup> The term “data warehouse” refers to a set of technology tools and processes used to gather information from source information systems, cleanse and catalog this information, store it in retrievable format, and enable reporting. Important features, such as confidentiality of data, make data warehousing for disease management and research a unique application of the technology.

## **Reaching the “Tipping Point” for EMR System Adoption**

One of the greatest obstacles to the use of clinical data in research is the low level of adoption of EMR systems across the United States. Although many industries, such as banking and insurance, have embraced information technology and the benefits it offers, healthcare in many respects continues to lag behind in this critical area.

### **Growing Investment in EMR Systems**

Although statistics vary somewhat regarding the adoption rate of EMR systems, most point to low but growing rates on the part of physicians and hospitals, with one estimate indicating that 14 to 28 percent of doctors’ offices have EMR systems (with larger practices tending to be more likely to adopt one than smaller practices).

Rates of adoption appear to be growing, however, and many expect growth to increase significantly in the near future. For example, although only 18 percent of respondents in a Healthcare Information and Management Systems Society survey of 550 hospitals report a fully operational EMR in place, nearly two-thirds indicated they have plan to implement an EMR system or have begun to install one, and more than half of U.S. hospitals plan to add EMRs in the next two years.

**BOX 1**

### **What If the Nation’s Goal of Building a Nationwide EMR System Was Realized?**

*Building the National Health Information Network has clear benefits. These include:*

- *fewer medical errors, which are estimated to be responsible for 44,000 to 98,000 deaths a year<sup>2</sup>;*
- *the provision of more efficient healthcare services;*
- *reduced health services utilization;*
- *improved ability to manage chronic disease;*
- *improved health status;*
- *streamlined work processes; and*
- *cost savings.*

*The focus of most efforts to implement EMR systems has been primarily on healthcare delivery and specifically on connecting provider communities that are dedicated to treating patients. However, there are important gains that will come from the use of EMR systems for research.*

***The real savings, in terms of both reducing healthcare costs and, more importantly, in eliminating human suffering, will come from curing disease and from limiting its damage.***

<sup>2</sup> See the Institute of Medicine report, *To Err Is Human: Building a Safer Health System*, 1999, [www.nap.edu/openbook.php?record\\_id=9728&page=1#pagetop](http://www.nap.edu/openbook.php?record_id=9728&page=1#pagetop).

### **What Are the Barriers to EMR System Adoption?**

*For Some, the Investment Is Not Worth the Expense.* In one survey of healthcare information technology executives, 20 percent of respondents cited lack of adequate financial support as the most significant barrier to successfully implementing information technology at their organization.

*Physicians Are Skeptical of the Value of EMRs.* On the physician and provider side, there are concerns about how information technology may affect the care provider role and disrupt workflow. Other concerns include the increased time needed to enter data and the potential for a system to have limited usability at the point of care.

*Consumers Are Concerned About Security Breaches.* Ensuring the security and privacy of EMR systems remains a substantial challenge, one that will be central to continued system adoption, as there is little doubt that improved system security will help drive automation.

*Institutions Have to Make Big Changes in the Way They Do Business.* Transitioning to an EMR system will require extensive changes in a healthcare organization's business processes, as well as in the computer hardware and software needed to support and run the EMR.

*Technical Issues Abound.* The proprietary basis of many EMR products raises key issues regarding the interoperability of systems, including the need for standard underlying reference vocabularies and presentation formats for clinical data. And EMR adoption must be able to grow with the information technology and healthcare environment, including new technologies such as handheld devices, wireless communications, biometrics, continuous speech recognition, new imaging techniques, Web access, and personal health record support via the Internet.

### **How Are the Barriers Being Addressed?**

Many issues are being tackled through collaborative efforts. For example:

- The federal **Consolidated Health Informatics** initiative seeks to adopt uniform standards for electronic exchange of clinical information across the federal health enterprise.
- Eight large technology companies have formed the **Interoperability Consortium** to aid the federal government's development of a digital health network.
- The private-sector **Certification Commission for Healthcare Information Technology** was created last year with the goal of creating a mechanism for certifying health information technology products.
- The Markle Foundation established the public-private collaborative **Connecting for Health**, which has worked to build consensus on the adoption of an initial set of data standards, develop case studies on privacy and security, and define the electronic personal health record.

There is obviously a long way to go before a national network of EMR systems is in place. But it is clear that pressure points for EMR adoption are growing within physician practices and in larger healthcare institutions, as well as within the federal government and through state initiatives.

BOX 2

## **What If All of Our Medical Records Were Accessible to Researchers?**

*EMR systems could speed data acquisition and searching, allow mass computing and sampling, and provide the research community access to a broader and more diverse patient population.*

**Pointing the way to new treatments.** *Researchers from IBM and the Mayo Clinic are using Blue Gene supercomputing technology and applying customized algorithms, data mining, and pattern recognition to uncover correlations between particular proteins, genetic markers, patient outcomes, and other factors that could lead to new diagnostics and treatments. Experts envision a time when computer programs will be developed to note patterns and trends of interest on their own, without humans having to create a query.*

**Conducting post-marketing research.** *With an EMR system, one could look at entire populations and know the impact of a new drug or device on, for example, diabetics, asthmatics, African American women over 65, or whatever population subset was chosen. Impact could be ascertained on a weekly, monthly or yearly basis.*

*A recent study conducted with Kaiser Permanente records of almost 1.4 million patients into the effects of COX-2 selective agents on heart disease resulted in the manufacturers of these agents voluntarily withdrawing their products from the market pending further safety studies.*

**Making clinical trials cheaper and easier.** *The Mayo Clinic conducts more than 4,000 clinical trials each year, and many trials rely on information from medical records, primarily to identify potential research subjects. Mayo has 6.5 million patient records collected systematically and electronically indexed. More than 95% of its patients allow their records and samples to be used for research.*

## **Challenges of Adapting the EMR as a Research Tool**

Even if all health providers and systems adopt EMRs, there is no guarantee that these systems would be useful for research. The famous computer axiom “garbage in, garbage out” (or GIGO) is particularly relevant when considering the use of clinical practice data for research purposes. To serve research needs, EMR systems will have to meet different and somewhat higher standards.

Several systemic problems must first be addressed if EMRs are to be useful for research.

### **The Art and Practice of Medicine**

There is no one way to practice medicine. Clinicians learn to observe and record data in numerous ways, resulting in difference in language and style. Not all procedures are performed in the same way—for example, measuring blood pressure when the patient is sitting will give a different result than measuring it when the patient is standing.

The adoption of standard vocabularies such as LOINC and SNOMED and standard messaging formats such as HL7 will be very helpful in leveling the playing field. But physician variability is not something that can be eliminated from the healthcare system, nor would we want such an impersonal outcome.

### ***Reliability and Completeness of the Record***

A 2002 study published in the Journal of the American Medical Association found that clinical information is missing in the records of almost 14 percent of visits to primary care physicians. It is clear that relying on disconnected and incomplete clinical records will not be sufficient for clinical research.

Limited evidence suggests that when physicians use computer- versus paper-based records, they are more likely to complete the necessary documentation. And statistical methods can be developed to establish a minimum acceptable set of data to be gathered at baseline to ensure that the study has the necessary power to be meaningful. In other words, programs can be written to account and adjust for missing data.

### ***Limits of Administrative and Claims Databases***

Most of the EMR software systems in use today have an administrative/transactions orientation, and if they are linked as a database it is frequently for the purposes of billing, claims, scheduling, outcomes assessment, and resource management.

Insurance claims databases lack important diagnostic and prognostic information when compared to clinical databases. For example, one can code only a limited number of diagnoses, and the goal is to maximize payment, not to achieve perfect clinical accuracy. Better-paying diagnoses are often selected as primary over lesser paying ones when patients have more than one active condition.

### ***The Problem of Unstructured Text in a Medical Record***

Medicine is an observational science. Clinicians observe and record, and much of the patient record consists of physicians' notes and comments, which do not easily translate in automated systems.

Natural language processing (NLP) is an exciting new area that has the potential to create huge and clinically rich databases from narrative reports. In its anticipated form, NLP could put the information in millions of clinical reports at the fingertips of researchers and clinicians. Several studies have already demonstrated that NLP improves clinical care and facilitates clinical research in areas such as chest radiology, tuberculosis care and stroke research.

### ***Integrating Practice Databases for Data Mining***

Impressive examples of EMR systems are in place both in U.S. hospitals and in some ambulatory practices. But will all of these efforts be creating new silos of information that will mount technical and cultural barriers to mining data across institutions?

To facilitate data mining of EMR databases by investigators:

- all the systems to be searched need to be accessible somehow by the researcher (or, more specifically by the software the researcher is using) using HTTP connections, database connection tools, or a yet-to-be-developed standard;
- all of the translation issues between different systems and formats need to be addressed;
- the issue of patient consent to have their medical data used for research would need to be addressed;
- some mechanism would need to be developed and enforced that would appropriately anonymize the data to be searched.

These goals will be difficult to achieve unless there is a central authority providing guidance and a clear, detailed set of standards and instructions for the designers and users of EMRs, as well as for the developers and users of searching and interconnecting tools.

### **Research Regulations**

The human research enterprise is highly regulated, and medical records research is no exception. The major ethical concerns in research with human subjects focus on the protection of privacy and confidentiality and the minimization of risk.

*Common Rule.* According to the 40-year-old “Common Rule” (formally the Federal Policy for the Protection of Human Subjects in Research) and general practice, the twin protections of informed consent and independent review of research by an IRB provide the foundation for an ethical approach to human subjects research. Any large database of identifiable personal health information is subject to the Common Rule and similar Food and Drug Administration regulations.

*Health Insurance Portability and Accountability Act of 1996 (HIPAA).* The Privacy Rule embedded in HIPAA imposed a new set of requirements for research using personal health information. It generally requires authorization from individuals to use their protected health information in research, unless an exception applies. Many institutions and investigators say that determining compliance with HIPAA remains confusing and that the law is slowing clinical research because of the cost and time required to comply with it, along with a tendency of the provider community to misinterpret its implications for research.

### **Moving Forward**

The greatest challenge to using the EMR for research is the reliability and validity of the data in the record. At some point soon, however, a critical mass may be reached in EMR adoption that will substantially accelerate progress in standardization. This must occur in tandem with efforts to integrate multiple databases for data mining. Standards for security and confidentiality also are needed, as is the consistent use of messaging standards.

*A high priority must be placed on understanding and integrating the needs of medical researchers into EMR systems generally and the NHIN specifically.*

## **Summary: Encouraging Institutions to “Think Research”**

The application of information technology to patient records offers the promise of new knowledge that can be obtained only by integrating and analyzing data extracted from hundreds if not thousands of patient records, including clinical information, medical images, environmental profiles, and genetic analyses, combined with new findings from molecular and genomics research. As institutions struggle with the adoption and implementation of EMR systems, it is crucial that they consider the needs and seek the advice of the research community.

Importantly, improvements made in EMR systems in response to research needs will ultimately serve clinical care needs as well. For example, in trying to achieve consistency and standardization, patient record systems will not only become more useful for research but will also contribute to improved quality of patient care. Likewise, the development of customized algorithms and pattern recognition systems will aid researchers while simultaneously providing physicians with smart clinical decision-support tools.

The clinical research community sees enormous potential in the ability of researchers to access and analyze the clinical information contained in millions of medical and personal health records. With appropriate privacy and human subjects protection safeguards in place, this capability could speed the discovery of new therapies beyond anything imaginable today.

**BOX 3**

## **EMRs and Research—Challenges and Solutions**

### **Challenges to Widespread EMR System Adoption**

- *Cost*
- *Infrastructure requirements*
- *Acceptance by and training of healthcare providers*
- *Interoperability of systems (hardware, software)*
- *Standard vocabularies and presentation formats*
- *Security and privacy*
- *Ownership issues*
- *Organizational and cultural factors*

### **Challenges to Research Uses of EMRs**

- *Reliability and completeness of the record*
- *Limits of legacy databases (administrative and transactional)*
- *Variability in medical practice*
- *Pervasiveness of unstructured text*
- *Lack of specificity of patient data*
- *Ability to look across many records and many databases*
- *Privacy and human subjects protections regulations*

### **Emerging Solutions**

- *Integration of practice databases for data mining*
- *More sophisticated abstraction and encryption systems*
- *Development of database connection tools*
- *Creation of translational systems*
- *Online informed consent procedures*
- *Evolving data mining and pattern recognition systems*
- *Interactive patient query programs*
- *Creation of patient databases/warehouses/registries*
- *Directories of clinical databases*

## **Recommendations**

*A high priority must be placed on understanding and integrating the needs of medical researchers into EMR systems generally and the NHIN specifically. Addressing these issues will be critical to the success of the NHIN and create new opportunities for the improvement of care.*

- 1** *All government and private-sector working groups focusing on the NHIN and EMRs should include representatives of the research community to ensure that research uses of EMRs are integral to all their statements of policy and principle.*
- 2** *Any Congressional legislation or federal regulation to promote EMRs should include provisions that will facilitate medical research uses of those records.*
- 3** *Organizations representing the research community should develop and advocate for standards and practices that would enable research use of EMRs.*
- 4** *Health information technology companies and organizations should work together with the medical research community in order to incorporate research needs into the architecture of EMR systems and the NHIN.*
- 5** *Institutions and agencies with existing medical records databases should make it a priority to make those databases open, accessible, and useful to the research community.*
- 6** *A Web-based resource should be created to allow researchers as well as other healthcare professionals to easily search and navigate the universe of already existing databases of patient records and medical information.*
- 7** *Employers and payers should consider the role of health research in lowering the cost of healthcare when planning their investments in health information technology.*
- 8** *Requirements of HIPAA that affect medical research with patient data must be clarified and, if necessary, revised.*

***Given the number of people who die from medical errors every year and from slow progress in finding medical solutions for deadly and debilitating diseases, it is clear that the transition to an EMR system with research capacity is not merely a desirable goal but a critical one.***



# **FasterCures**

*The Center for Accelerating Medical Solutions*

509 Seventh Street, NW Washington, DC 20004 Telephone [202] 654.7090

**[www.FasterCures.org](http://www.FasterCures.org)**

*To catch breaking news in the medical solutions arena,  
sign up for the FasterCures SmartBrief – a free twice-weekly e-mail – at [www.fastercures.org/smartbrief](http://www.fastercures.org/smartbrief)*