



# Tapping into the Value of Health Data Through Secondary Use

Daniel Garrett

hfm Magazine

February 2010

**As electronic health records (EHRs) proliferate across the nation, an important new opportunity awaits healthcare organizations that can find meaningful commercial uses for the data contained in their EHR systems.**

---

## At a Glance

- **Hospitals and health systems have an important opportunity, today, to use clinical, financial, administrative, and self-reported data to improve health outcomes, reduce medical errors, predict health trends, and demonstrate the comparative value of drugs and treatments.**
  - **Such secondary use of data requires that the data be standardized to be effective.**
  - **Danville, Pa.-based Geisinger Health System is an example of a larger healthcare organization that is pursuing a commercial venture to effectively use its vast data resources for research and other purposes in the larger healthcare marketplace.**
- 

Health care has always been data-intensive, even on paper. The U.S. healthcare system encompasses a vast amount of information, including patient history, claims data, diagnostic reports and images, financial records, clinical research findings, and now a growing body of genomic data, including all 3 billion pairs of DNA in the human genome. Moreover, the amount of health data being collected for electronic use will grow exponentially over the next several years as the nation's hospitals and health systems increasingly implement EHRs.

Yet these data today represent a largely untapped resource. Beyond the immediate uses for such data—which include improving continuity of care, avoiding adverse events, and billing for services—is a broad range of other uses that few healthcare provider organizations have even begun to realize. Indeed, if all that information could be aggregated and used to unlock the mysteries of medicine, the data resource would be a national treasure surpassing the value of even Fort Knox. But much like the gold ore trapped in the California hills that beckoned to prospectors in the 1850s, much of the nation's health data remain trapped, on paper and even in electronic files, as isolated veins of information within individual organizations or departments.

## The Elusive ROI

These veins of data may begin to open up with the growth of EHRs across the industry. And the government has taken steps to encourage that growth, with Congress allocating billions of dollars in the American Reinvestment and Recovery Act of 2009 to expand health IT, and promising incentives for the adoption and meaningful use of interoperable EHRs and penalties for not doing so. The pressure is on to invest in EHRs now, and the clock is ticking.

Even so, converting to EHRs is an expensive proposition, and financial leaders of physician practices and many hospitals are still struggling with the math. The upfront price tag for an integrated and mature electronic health system is steep, far outweighing the stimulus bill's incentives. A three-physician practice can expect to invest between \$173,750 and \$296,000, on average, over two years to purchase and maintain an EHR system.<sup>a</sup> The costs run into the tens of millions of dollars for large hospital systems,

Demonstrating an ROI for an EHR has been an elusive goal. A troubling concern for providers is the perception that the financial benefits of an EHR will accrue to public and private payers, but not to the provider's bottom line. Improving continuity of care, for example, clearly reduces costs for payers, but not necessarily for providers, and apart from the knowledge that they have contributed to improved outcomes for patients, providers do not see any financial return from their efforts. Thus, although the anticipated clinical payoff derived from eliminating redundancies and improving administrative efficiency is largely uncontested, many healthcare executives say that a financial return from investing in EHRs is a myth.

Even if the clinical benefits of EHRs will somehow always dwarf the financial benefits, however, providers should not disregard the financial value latent in their growing stores of health data. The reality is that it's never been simply about converting records from paper to electronic files. The real value of health data, and the elusive ROI from an EHR, lies not in the collection and storage of primary health data, but in the secondary use of the data.

### **The Potential for Secondary Use**

Some of the nation's larger healthcare provider organizations are beginning to realize the potential for secondary use of their health data. A relatively new practice for many providers, such use of data has the power to transform how health care is delivered in the future. Secondary use refers to the use of clinical, financial, administrative, and self-reported data—collected from electronic medical records (EMRs), personal health records, insurance claims, clinical trial information, billing information, and other sources—to improve health outcomes, reduce medical errors, predict health trends, and demonstrate the comparative value of drugs and treatments, among other benefits.<sup>b</sup> For this purpose, the data must be de-identified, aggregated, analyzed, and presented in a concise, actionable format.

The types of secondary data discussed here are possible only if data are standardized to eliminate information variation during the implementation process. Standardization is necessary to allow organizations to repurpose the collected health data, share them, and then put them to meaningful use. Once the requirement for standardization is met, aggregated health data can provide value to a broad range of research, quality, public health, and commercial applications.

For example, aggregated patient data could be mined for interesting correlations, say between an insect bite and pancreatitis or between emergency department throughput and payer mix. And armed with information about patients' care from other providers or health managers outside of their organization, physicians' would be in a better position to prevent chronic conditions from becoming catastrophic events or to intervene before patients present with diseases and claims are filed.

From a public health perspective, provider connectivity would make reporting of vital statistics and cases of certain diseases more efficient, potentially saving the nation millions of dollars. An even more important impact of public health interoperability would be biosurveillance and earlier recognition of emerging disease outbreaks as analysis of aggregated data from multiple sources makes it easier to identify warning signs and trends.

Recent advances in science and technology make it increasingly likely that human genomic data will be routinely available to clinicians as well as their patients. As patients gain the ability to access their own

health information, they will increasingly demand more personalized care, which will in turn create a greater demand for access to, and analysis of, health data from outside of clinical settings.

Already, commercial enterprises are beginning to collect healthcare data and develop new products and services aimed directly at consumers, as well as to third-party payers, researchers, public health officials, and marketing entities.

Last year, for example, the Centers for Disease Control and Prevention (CDC) selected GE Healthcare to provide extensive surveillance data for H1N1 and seasonal influenza activity throughout the United States. Every 24 hours, GE Healthcare reports secure and de-identified information from its proprietary Medical Quality Improvement Consortium (MQIC) database, containing more than 14 million patient records. The reports help the CDC monitor the spread of the H1N1 virus in near real-time (Businesswire: GE Healthcare news release, Oct. 28, 2009).

To gain a deeper understanding of healthcare costs, GE combined the Medical Expenditure Panel Survey (MEPS), from the Agency for Healthcare Research and Quality, with 500,000 MQIC records. By combining MEPS with GE's data, GE contends it can gain a more complete picture of the costs associated with chronic conditions. (An interactive tool that helps to visualize this data is available online for consumers at [www.ge.com/visualization/health\\_costs/index.html](http://www.ge.com/visualization/health_costs/index.html)).

It's difficult to know for sure what the potential market will be for secondary data, as stimulus-driven utilization of EHRs increases dramatically. Consider, however, that, in one instance, a single hospital cardiology department has been providing de-identified data to a pharmaceutical company, with no change required to source data, for \$400,000 annually.

The current annual market for clinical data purchased by life science companies and other entities exceeds \$6 billion. But these data, which are mostly claims-derived, do not fully meet the future-vision needs of pharmaceutical and life sciences companies, and the companies know that. The largest opportunity exists in outcomes studies that require a higher level of complexity and more advanced data analytics. The current market for data supporting outcomes studies exceeds \$900 million according to estimates by PricewaterhouseCoopers. Capitalizing on just a fraction of the current market represents millions if not tens of millions of dollars of opportunity to data sellers. Many data consumers are looking for a database of millions of patients, a geographically-dispersed patient mix, and standardized and consistent nomenclature.

### **Today's Environment**

Today, electronic exchange of clinical data between organizations is a nascent practice. The main reason is lack of access to EHRs. Relatively few providers have broad and mature clinical information systems. In a recent survey of nearly 500 providers, conducted in June 2009, fewer than half of providers said they have implemented any more than the most basic functions of EMRs, and only 23 percent are actually using EMRs for patient clinical documentation.<sup>c</sup>

These findings mirror the results of the EMR implementation analytics of the Healthcare Information and Management Systems Society (HIMSS), which found that only 1.2 percent of U.S. hospitals had implemented advanced clinical applications that improve patient safety and care delivery outcomes as of the end fall of 2009 (HIMSS Analytics EMR Adoption Model, as tracked through the HIMSS Analytics database of 5,172 hospitals, cumulative as of the end of the third quarter 2009, [www.himssanalytics.org/stagesGraph.html](http://www.himssanalytics.org/stagesGraph.html)).

The 2009 survey did find, however, that secondary use of health data is occurring in limited form, with the users for the most part being providers with robust EHRs and pharmaceutical companies, which are

primarily purchasing third-party information. Specifically, among the survey respondents, 65 percent of providers and 66 percent of pharmaceutical companies are currently using secondary data. By comparison, only 54 percent of the health plans surveyed are making secondary use of the data as described here, despite their access to a vast amount of claim information that would be useful to providers. Only 39 percent of payers offer personal health records, and few members use them.

Broadly, most organizations are currently using secondary data for their own quality monitoring and reporting, although they are interested in using health data for even greater benefits in the future ([see exhibit](#)).

Today, a handful of very forward-thinking healthcare organizations understand that the data they are now amassing in their various IT systems hold enormous potential outside of their enterprises. They are pioneering the secondary use of data for purposes such as disease management, clinical-studies validation, postmarket surveillance of drugs, and even to identify new revenue sources and service offerings.

However, there are unresolved issues around the secondary use of data, not the least of which is ensuring the privacy and security of patient information. The vast majority of health executives surveyed in 2009 (over 80 percent) cited privacy, legal implications, and public relations ramifications as concerns. It is interesting, however, that these issues were less likely to be mentioned as an actual barrier to secondary data use, indicating that privacy and security is an important concern but not an insurmountable barrier.

Although the assumption is that ‘secondary information’ will not identify individuals, the simple reality is that once an individual’s information is placed in a database, the possibility of individual identification is possible. Access to such information would be a powerful tool, and the utmost care and controls would be needed to protect it. Can privacy be guaranteed? Even the federal government has a visible disclaimer in all data collection arenas (Social Security, Medicare, Medi-Cal, workers compensation) that no guarantee of privacy exists nor is guaranteed.

Among survey responses, there is a clear consensus across all stakeholders that the patient should always be the focus of any data use, and nearly two-thirds executives surveyed agreed that reuse of individual and/or identifiable data is acceptable if it is in the best interest of the patient. But they also felt strongly that information should never be used without the individual’s explicit consent.

Unanimously, health executives agree that patients’ rights and privacy must always be protected, and that data must be transparent and overseen by honest brokers or stewards. Nearly all executives surveyed feel the industry needs better guidelines about how health information can be used and shared, and three-quarter feel that national stewardship over, or responsibility for, the use of the health data should be regulated.

### **Case Study: Geisinger Health System**

Geisinger Health System is an integrated delivery system based in Danville, Pa., that comprises a 700-plus multispecialty physician practice, three hospitals, 40 community practice sites, a health plan, three research centers, and an internal venture group. One of Geisinger Ventures’ portfolio companies, MedMining, is an excellent example of how one organization is applying its data for a secondary commercial purpose. MedMining provides customized, de-identified data extracts to promote health economics and other biopharmaceutical research. The data are derived from numerous Geisinger Health System databases, most notably its EHR system.

Geisinger was an early adopter when it implemented an EHR system in 1996. Through the years, it has invested more than \$100 million in the EHR system, clinical decision information system, and other

supporting resources to integrate its disparate data, including data in departmental and stand-alone databases and not readily reportable data in the EHR. As a result of a combination of Geisinger's financial investment, the nearly 100 percent adoption rate due to use of a physician-led staff model, and the early implementation, MedMining is now able to make use of more than 10 years of granular-level clinical, laboratory, and economic data.

Early on, Geisinger's senior leaders realized the clinical value of leveraging the organization's data and initiated several projects to acquire insight to improve patient care and provide clinical decision support for patients. They also saw an opportunity to "grow the research pie nationally" by de-identifying and licensing the organization's data. Pursuing this opportunity led them to leverage the commercialization engine of Geisinger's venture group to extend the project's focus beyond simply research.

Geisinger Ventures canvassed internal and external constituents and drew up a market-based business plan and a formal ROI analysis. The group went well above complying with HIPAA regulations to ensure that patient privacy remained sacred.

Geisinger Ventures interviewed prospective customers to ensure that this type of de-identified data could, in fact, help promote research that otherwise wouldn't be possible. The group also recognized the challenges and limitations of current research. In randomized, controlled trials, qualified patients often have just one disease state as opposed to the complex web of comorbidities that is more typical of the patients that hospitals and health systems see on a day-to-day basis.

One of MedMining's first clients was a company that had four important research questions that had gone unanswered for almost a year because the company could not identify a data set with the clinical detail across the care continuum needed to provide the desired insight. MedMining was able to provide granular data needed to answer two of these questions, and the company remains one of MedMining's largest returning customers.

Another pharmaceutical company was conducting a postmenopausal osteoporosis study that had languished for some time because the company was unable to collect data needed for the study. MedMining was able to extract data for fracture sites and bone mineral density scores, providing unique insight into strengths and weaknesses of different classes of therapies.

Had either company not gone to MedMining, the companies would have been left with only inefficient, ineffective, and possibly expensive alternatives: change the research questions, study something else that is equally or nearly as meaningful, or create a prospective trial. The third alternative would require millions of dollars and other resources to recruit an appropriately consented cohort of patients and several years to build up de-identified, retrospective data.

To date, MedMining's data partners include the top 10 pharmaceutical companies, as well as midscale and large biotech companies. MedMining's data support effectiveness research across the disease spectrum, including epilepsy, chronic obstructive pulmonary disease, hyperlipidemia, oncology, erosive esophagitis, gastroesophageal reflux disease, and postmenopausal osteoporosis.

MedMining leverages its "routine, open, and collaborative" dialogue with the leaders of Geisinger's three research centers—which are focused respectively on clinical trials, basic science and molecular biology, and clinical and applied outcomes. As a result, the company is at the forefront of Geisinger's efforts to attract nontraditional clients ranging from healthcare IT companies to genomics research firms.

Finally, a byproduct of MedMining's business of providing a granular view of patient data is the ability to present Geisinger's own physicians with a new, unique view of their patients' data that can provide insights to disease progression or therapeutic response that otherwise might have gone unnoticed.

### **The Value of Large Datasets**

Going forward, MedMining wants to have a greater impact on the industry, and on a much broader set of patients, by integrating its data with those of other like-minded health-system peers. To this end, Geisinger seeks relationships with peers that have three to four years of EHR history across the care continuum or nearly so. The challenge is that few health systems today have the technology in place and the requisite historical data.

This challenge may diminish as EHR adoption grows. Already, MedMining has been approached by a growing number of like-minded health systems with EHRs that would like to use the company as a vehicle for conducting better research. Instead of starting from scratch, these groups see the opportunity to leverage MedMining's validated technology platform, entrenched market position, and strong distribution channel relationships as a sensible, low-risk way to enter the market.

For Geisinger, this opportunity means removing geographic and ethnic diversity constraints, given the health system's regional clinical reach. As a result, the expanded data could positively influence research that examines, for example, the effect of therapies on certain ethnicities.

### **Deriving Maximum Benefits**

Clearly, the opportunity for secondary use of health data exists, and nearly two-thirds of health organizations, across all sectors, responding to the previously cited 2009 survey indicate they will increase their use of secondary data over the next two years ([see exhibit](#)). To drive greater benefits in the future, provider and payers should consider investing in health IT and prioritizing these investments around ensuring data accuracy and data mapping.

With national standards today neither completely defined nor adopted, it might be tempting for some health organizations to develop a nonstandardized system. Although such systems can indeed aggregate information from remote sources, they must reconcile diverse codes, data structures, and terminologies—an imprecise process that could limit the efficacy of clinical decision support and generate not only redundant information but also clinical errors. A more robust system, with seamless integration of local and remote records, would be far more likely to offer clinicians the integrated information they need to provide optimal care.

To derive maximum benefit, public and private sectors will need to work together to overcome barriers to the use of secondary data, address privacy concerns, and foster greater collaboration across sectors. As a first step, the industry should focus on establishing guidelines about how health information can be gathered, used, and shared. Among executives responding to the 2009 survey, 90 percent agree that this is a priority, and believe the healthcare market should be responsible for defining standards.

Some industry executives suggest the way to develop standards for secondary data use would be to form an industry consortium, comprising high-level decision-makers—including clinical and IT experts—who would represent provider, payer, pharmaceutical, and patient organizations. This type of consortium would be well positioned to determine the industry's needs and uses for collecting secondary data, to develop common goals for the types of data needed and how they can be used effectively, and to create a new data architecture that would enable interoperability among IT systems to facilitate the linking and analysis of the secondary health data.

Early adopters agree that the industry cannot go it alone. Incentives eventually will be needed to encourage the development of more regional exchanges of information, and if national standards are established, networks one day may be able to collaborate across a seamless, national health care information system. This result will likely not occur, however, without some initiative from the federal government. True interoperability will require strong policy incentives. Government has an important role to play in advancing secondary data use through the creation of incentives to motivate the industry to collect, report, and use the data that will, in turn, allow for quality, cost, and access to be measured meaningfully.

Working in unison, the industry, with the support of government, can begin to realize the benefits of secondary data usage and its potential to transform the healthcare system.

---

Daniel Garrett is a partner, PricewaterhouseCoopers, and leader of the company's Health Information Technology practice, Philadelphia ([daniel.garrett@us.pwc.com](mailto:daniel.garrett@us.pwc.com)).

---

### footnotes

a. Rock and a Hard Place: An Analysis of the \$36 Billion Impact from Health IT Stimulus Funding, PricewaterhouseCoopers Health Research Institute, April 16, 2009, [www.pwc.com/ehrs](http://www.pwc.com/ehrs).

b. b. Secondary use, in this instance, does not refer to the long-standing, and controversial practices attributed to health insurance companies of using medical data to screen policy applicants for rejection of coverage, for claims payment denials, for identification of preexisting conditions, and as a basis for retroactive denial of policy coverage.

c. c. Transforming Healthcare Through Secondary Use of Health Data, PricewaterhouseCoopers, Oct. 4, 2009, [www.pwc.com/](http://www.pwc.com/)

d. secondaryhealthdata.

---

### Sidebar

#### EHR Versus EMR

Is there a difference between an electronic health record (EHR) and an electronic medical record (EMR)? Or are these two, often-encountered two terms simply two ways of saying the same thing? In reality, the answer to these questions is probably yes in both cases. There have been efforts to make a clear distinction between the terms, but those efforts have not resulted in universal agreement, and given the lack of consensus terms, it is not surprising that the terms are used interchangeably in some quarters. Even so, a strong case can be made for making a distinction.

To this end, in 2008, the now-defunct National Alliance for Health Information Technology (NAHIT) proposed the following definitions for EHR and EMR.

*EMR:* An electronic record of health-related information on an individual that can be created, gathered, managed, and consulted by authorized clinicians and staff within one healthcare organization.

*EHR:* An electronic record of health-related information on an individual that conforms to nationally recognized interoperability standards and that can be created, managed, and consulted by authorized clinicians and staff across more than one healthcare organization.

“The principal difference between an EMR and an EHR is the ability to exchange information interoperably,” NAHIT says in its April 28, 2008, report, *Defining Key Health Information Technology Terms*. Because the industry is moving toward electronic records that meet national standards for interoperability, “The term EMR is on course for eventual retirement,” according to NAHIT.

Source: Amatayakul, M., “EHR Versus EMR: What’s in a Name?”, *hfm*, March 2009.