Implementing and Maintaining a Researchable Database from Electronic Medical Records: A Perspective from an Academic Family Medicine Department

Mise en place et maintien d’une base de données à partir des dossiers médicaux informatisés : le point de vue d’un service universitaire de médecine familiale

by MOIRA STEWART, PHD
Professor, Canada Research Chair in Primary Care
Department of Family Medicine
Schulich School of Medicine and Dentistry
University of Western Ontario
London, ON

AMARDEEP THIND, MD, PHD
Associate Professor, Canada Research Chair in Health Services Research
Department of Epidemiology and Biostatistics
Department of Family Medicine
Schulich School of Medicine and Dentistry
University of Western Ontario
London, ON
Implementing and Maintaining a Researchable Database from Electronic Medical Records

AMANDA L. TERRY, PHD
Post-Doctoral Fellow, Department of Family Medicine
Schulich School of Medicine and Dentistry
University of Western Ontario
London, ON

VIJAYA CHEVENDRA, BED, MSC
Systems Analyst, DELPHI Project
Department of Family Medicine
Schulich School of Medicine and Dentistry
University of Western Ontario
London, ON

J. NEIL MARSHALL, MB, BCH
Liaison Physician, Department of Family Medicine
Schulich School of Medicine and Dentistry
University of Western Ontario
London, ON

Abstract

Electronic medical records (EMRs) are posited as a tool for improving practice, policy and research in primary healthcare. This paper describes the Deliver Primary Healthcare Information (DELPHI) Project at the Department of Family Medicine at the University of Western Ontario, focusing on its development, current status and research potential in order to share experiences with researchers in similar contexts. The project progressed through four stages: (a) participant recruitment, (b) EMR software modification and implementation, (c) database creation and (d) data quality assessment. Currently, the DELPHI database holds more than two years of high-quality, de-identified data from 10 practices, with 30,000 patients and nearly a quarter of a million encounters.

Résumé

Les dossiers médicaux informatisés (DMI) se veulent un outil pour améliorer la pratique, les politiques et la recherche en matière de soins de santé primaires. Cet article décrit le projet DELPHI (Deliver Primary Healthcare Information) du service de médecine familiale à l’Université Western Ontario, en mettant l’accent sur sa mise en place, son statut actuel et son potentiel de recherche, et ce, afin de partager l’expérience avec les chercheurs qui travaillent dans un contexte semblable. Le projet s’est déroulé en quatre étapes : (a) le recrutement des participants, (b) la modifica-
tion et la mise en place du logiciel de DMI, (c) la création de la base de données et (d) l’évaluation de la qualité des données. Actuellement, la base de données du projet DELPHI conserve plus de deux années de données anonymes de grande qualité, qui proviennent de 10 cliniques, comptent 30 000 patients et représentent près d’un quart de million de visites.

Moira Stewart et al.

There is currently keen interest in electronic medical records (EMRs) as a tool for improving practice, policy and research in family medicine and interdisciplinary primary healthcare (PHC). Evidence from the literature suggests that EMRs can improve practice by providing point-of-care information to assist clinical decision-making (Bates et al. 1999; Garg et al. 2005) and by giving feedback on standards of care leading to improved patient management (Mitchell et al. 2005; Toth-Pal et al. 2004; Vogt et al. 2007). EMRs can help policy making by providing evidence about primary care workload community needs, which are expressed as health services utilization (Okkes et al. 2002).

Unfortunately Canada lags behind other countries in harnessing the full potential of EMRs, both for patient care and research (Protti 2007; Schoen et al. 2006). Data from a recent study show that only 12.3% of Canadian primary care physicians were using electronic charts instead of paper charts in 2007 (College of Family Physicians of Canada et al. 2007). The Centre for Studies in Family Medicine (CSFM) at the Department of Family Medicine, University of Western Ontario (UWO), embarked on the DELPHI Project in 2003 with the aim of creating a researchable database from the EMRs of community family physicians in southwestern Ontario. Our paper describes (a) general issues of definition and research potential worldwide and (b) specific issues of the development, current status and research potential of the DELPHI Database in an effort to share our experiences with researchers in similar contexts.

Background

The terminology used to describe electronic charts and patient records varies depending on the agency and the purpose. In general, the term electronic health record (EHR) has been used to describe patient records that are accessible from many sites or by many different providers. The term electronic medical record (EMR) has more commonly been used to describe electronic patient records that are kept in one location and are accessible on only one provider’s site. Iakovidis (1998) described the EMR as a stand-alone system, whereas the EHR is defined as digitally stored healthcare information that accumulates over a person’s lifetime to support continuity of care. Likewise, Canada Health Infoway described EHRs as interoperable records that
follow patients as they move through the system, providing complete information to all providers as needed (Booz Allen Hamilton 2005). In this paper, we use the term EMR to describe the clinical records held in electronic form within primary healthcare practices (possibly connected to laboratory and some hospital data) and used in the course of everyday care of patients. These records typically contain such elements as procedures and investigations, immunization lists, referrals, laboratory results, clinical notes, examination results, medications lists and a problem list.

Research uses of EMR data are fairly well established in countries such as the United Kingdom, where large practice-based data collections, such as the General Practice Research Database, exist. More recently, researchers in North America (Gill et al. 2006; Ornstein et al. 2007) have begun to use these data for research. Worldwide, specific research uses of EMR data include helping to improve primary healthcare for patients with chronic conditions such as diabetes (Gill et al. 2006; Ornstein et al. 2007; Kupersmith et al. 2007) and hypertension (Mitchell et al. 2005), as well as to enhance preventive care (Toth-Pal et al. 2004; Vogt et al. 2007), to examine relationships between symptoms and ensuing disease in patients (Jones et al. 2007) and to support family concordance studies (Hippisley-Cox et al. 2002). However, significant challenges remain in using EMR data for research (Lobach and Detmer 2007).

The DELPHI (Deliver Primary Healthcare Information) Project

The DELPHI Project began in 2003 with a Canadian Foundation for Innovation (CFI) grant, which was supplemented by a substantial grant from the Ontario Primary Health Care Transition Fund (PHCTF) in March 2004. The overarching goal was (a) to facilitate the development of an EMR system for interdisciplinary PHC for the purpose of improving information-sharing in an interdisciplinary care setting and (b) to describe, assess and improve the quality of PHC delivery. Although housed at the Department of Family Medicine at UWO, partnerships with the Institute for Clinical Evaluative Sciences (ICES), Healthscreen Solutions (the EMR software company) and the Department of Epidemiology and Biostatistics at UWO were instrumental in several components of the project.

The project progressed through four stages: (a) participant recruitment, (b) EMR software modification and implementation, (c) database creation and (d) data quality assessment. Each of these four stages is briefly described below.

Participant recruitment

Family practices were recruited through a variety of complementary approaches. First, using a strategy similar to that of Borgiel and colleagues (1989), the researchers enlisted key community leaders in family medicine in identifying suitable EMR
software, locating practices that were already using EMR software candidates or that might be interested in its future use. Second, a notice soliciting interest was sent from the Centre for Studies in Family Medicine to all 160 family physicians on the Family Medicine Education and Research Network (FERN) e-mail discussion group. Third, the software company suggested family physicians who had expressed an interest in their product. All family physicians who responded with an expression of interest were personally visited by the principal investigator to discuss project details. Approximately half the group practices that were invited by the principal investigator opted in to the EMR implementation and the research. Those who opted in were (a) very interested in the specific EMR chosen, (b) appreciative of and even excited about the usefulness of the EMR data they would be providing, in comparison to the usual billing data (the only data available in a structured form in Ontario up to that time) and to manual chart audits (the time-consuming, labour-intensive alternative) and (c) content with the degree of interoperability with hospitals and diagnostic tests, which varied from community to community but was a deal-breaker in several instances. The recruitment strategy resulted in a final sample of 25 family physicians in the DELPHI Database, as well as 25 family practice nurses, one nurse practitioner and one chiropodist. These primary care practitioners are located in 10 group practices.

Although the strategy did not result in a strict random sample of family physicians, the project covers a wide geographic area of southwestern Ontario, stretching from near Windsor in the south to Kincardine in the north and Brantford in the east, as well as the London area (Figure 1). The distribution of family physicians is broadly representative by age and gender, although the sample is slightly less urban than Ontario family physicians (Table 1).

**FIGURE 1.** Locations of DELPHI practice sites in southwestern Ontario
Implementing and Maintaining a Researchable Database from Electronic Medical Records

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44 years and under</td>
<td>24%</td>
<td>35.3%</td>
<td>29.5%</td>
</tr>
<tr>
<td>45–54 years</td>
<td>24%</td>
<td>31.3%</td>
<td>31.7%</td>
</tr>
<tr>
<td>55–64 years</td>
<td>36%</td>
<td>22%</td>
<td>26.3%</td>
</tr>
<tr>
<td>65+ years</td>
<td>12%</td>
<td>8.6%</td>
<td>11.6%</td>
</tr>
<tr>
<td>Unknown</td>
<td>–</td>
<td>2.7%</td>
<td>1%</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>64%</td>
<td>68%</td>
<td>61%</td>
</tr>
<tr>
<td>Female</td>
<td>36%</td>
<td>32%</td>
<td>39%</td>
</tr>
<tr>
<td><strong>Practice Location</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inner city</td>
<td>–</td>
<td>6%</td>
<td>12.4%</td>
</tr>
<tr>
<td>Urban/suburban</td>
<td>20%</td>
<td>38.4%</td>
<td>58.4%</td>
</tr>
<tr>
<td>Small town</td>
<td>48%</td>
<td>24.5%</td>
<td>14.4%</td>
</tr>
<tr>
<td>Rural</td>
<td>32%</td>
<td>15.7%</td>
<td>8.3%</td>
</tr>
<tr>
<td>Other/no response</td>
<td>–</td>
<td>20%</td>
<td>7.5%</td>
</tr>
</tbody>
</table>

* We do not have data on age of DELPHI physicians. Year of graduation was used as a proxy for age, with the assumption that most graduates would be approximately 28 years old at the time of graduation.

The 25 family physicians originally committed to a three-year period. Currently, at the end of year 2, all 25 continue to be involved in all facets of the project. The views of the participants in the DELPHI Project toward EMR implementation in their practices are described elsewhere (Terry et al. 2009).

**EMR modification and implementation**

In order for the database to be fully researchable, the EMR software had to accommodate research-oriented data input technologies and data extract possibilities. These two components were absolute requirements for software selection. After spending considerable time and resources, the research team identified Healthscreen as the most suitable software for its purposes. The company was willing (and eager) to develop these modifications, and a close relationship was established and maintained between the researchers and the EMR software company, Healthscreen Solutions Inc.
The two research-oriented data input technologies were the incorporation of the International Classification of Primary Care (Verbeke et al. 2006; Soler et al. 2008; O’Halloran et al. 2004) and a diabetes flow-sheet, which were developed over the course of a year. After successful beta testing, the revised software was installed in all participating practices. In addition, practice-specific hardware was purchased and installed. While remote training was offered by the software vendor, participants were also offered individual intensive training sessions with the DELPHI staff. This included customized site-specific training, and trouble-shooting during the implementation process. The DELPHI team worked to build relationships between local information technology service providers and the family practices, thus helping to maintain a supportive presence throughout the project.

Decisions about the two new data input strategies were based on a number of considerations. First was our choice of the International Classification of Primary Care (ICPC). From the perspective of family medicine and interdisciplinary PHC, the usefulness and feasibility of a detailed dictionary of 300,000 terms such as SNOMED-CT is questionable. In contrast, ICPC has two advantages: (a) it is a hierarchical classification and therefore groups problems into chapters that are relevant to clinical medicine, as does ICD-9 and ICD-10, and (b) its terms include undifferentiated problems (such as back pain), which comprise 50% of a family doctor’s workload (Crombie 1963; Blacklock 1977; Jerritt 1981).

Second was our decision to create a diabetes flow-sheet to computerize the common and popular paper-based flow-sheets. This decision was a response to both practitioners’ interest and the interest of policy makers who were beginning, in 2004, to become concerned about the high prevalence of diabetes in the population (Ontario Ministry of Health and Long-Term Care 2004).

Overall, the need for these data input strategies existed because of the lack of data on the realities of PHC and to provide data on the types of problems presented to PHC, such as symptoms, psycho-social problems, tentative diagnoses and relevant interventions.

Database creation

Concurrent with EMR modification, the DELPHI team worked closely with the Office of the Information and Privacy Commissioner of Ontario, the Chief Privacy Officer at ICES and the Ethics Review Board of UWO to formulate a privacy policy. Once signed consent from the physicians was obtained indicating their willingness to participate in the project, the DELPHI team put up posters (which were prominently displayed) in the practices’ waiting and examining rooms. Patients who did not wish to participate were able to refuse by informing the named project coordinator or their primary care practitioner directly; their EHR records were not taken during the data
extraction process. To date, several families (not more than 10 individual patients) have opted out through the project coordinator.

Data extraction occurs on a quarterly basis. The data extracted for all patients include the billing code, problem lists, family history, medications, allergies, immunizations, physical examinations, investigations, laboratory tests, interventions and referrals. On the random subset of patients for whom physicians are doing coding in ICPC, additional data elements extracted include (a) up to five reasons for encounter (RFE) per visit (these are recorded in the patient’s own words), (b) up to five diagnoses per visit and (c) tracking of the non-chronic diagnoses during episodes of care.

The repeated extracts are conducted in a manner that does not require patients’ identification. Each extract contains a longer time period than the previous extract, the longer period including the time of the previous extract as well as the new time period, as Figure 2 shows.

**FIGURE 2.** Successive cumulative extracts of electronic health records recreates the DELPHI Database each quarter

---

Each patient record is assigned a unique number for the study. The patient’s name, address and telephone number are not taken from the doctor’s office. The extracted records are taken to UWO, where they are pooled with information from the other primary care practices. This pooled database is known as the DELPHI Database. Since the database does not contain any identifying data, it is not possible to identify a patient or physician in it. Moreover, the database is accessible only to DELPHI Project staff, who have signed strict confidentiality agreements.

To facilitate comparison between the EMR and health administrative data, it was
necessary to link these data sets. This linkage between DELPHI and the Institute for Clinical Evaluative Sciences (ICES) in Toronto followed a precise process to ensure data security. ICES has been designated a prescribed entity (s. 45 and s. 18 of Reg. 329/04) in the Personal Health Information Protection Act (PHIPA) of 2004, which allows it to receive personal health information from the healthcare practitioners (termed “health information custodians” under PHIPA). This transfer must be for the intention “of analysis or compiling statistical information with respect to the management of, evaluation or monitoring of, the allocation of resources to or planning for all or part of the health system.” The purpose of the data transfer for this study fits within this definition.

To link DELPHI data with ICES data, a DELPHI key was created. The key was constructed by the project coordinator during the visit to the physicians’ offices. The following pieces of information constituted the key: (a) the patient’s unique DELPHI study number, (b) patient’s OHIP number, (c) patient’s postal code, (d) patient’s date of birth, (e) OHIP billing number of the physician providing care to that patient and (f) the physician’s unique DELPHI study number. It is important to note that this key did not contain any information about the patient’s medical care. It was password protected and encrypted, and the project coordinator transported it directly from the primary care practice to ICES in Toronto. At no time was this key in the same location as the DELPHI Database. At ICES, the DELPHI key was used to link the DELPHI Database with the health administrative databases at ICES. The key was destroyed immediately after this linkage. The linked data sets use only anonymous data for purposes of analysis as per ICES policies. All individuals who were given access to the DELPHI–ICES Linked Databases signed a confidentiality agreement to ensure that they did not disclose individual patient information to any other person, as per ICES privacy policies.

Data quality assessment

As almost half the participants were novice users, the DELPHI team developed a proactive approach to ensure data quality. To assist these users in moving to the advanced level, the team provided a variety of supports, including one-on-one training, continuous trouble-shooting, flexible project timelines and general facilitation of the use of the EMR software by maintaining a supportive presence. Keeping in view that these were extremely busy family practices, the team adopted a user-centred strategy so that the implementation of EMRs could proceed without disrupting the daily patient workflow in the clinics. Once the database was populated with data from the extracts, an ongoing quality monitoring system was put in place to ensure data completeness and standardization across the sites. The DELPHI team provided additional training to the participants and have continually emphasized the importance of consistent data entry.
Current Status of the Database

To date, the DELPHI Database holds more than two years of high-quality, de-identified data from the 10 practices, with 30,000 patients and nearly a quarter of a million encounters. Two key linkages are depicted in Figure 3. A linkage with ICES administrative data has been conducted to create and test patient-level indicators of primary healthcare provision. Using the postal code, the database has been linked to Statistics Canada Census data to provide a wealth of socio-economic data. Figure 3 also depicts the possible studies that are being (or can be) conducted using the researchable database. A sample of work currently underway is described below.

![Figure 3: Components of the DELPHI Database (March 1, 2006 – February 29, 2008)](image)

Characterizing primary care practice

The DELPHI team conceptualized five indicators of primary care and delineated their domains by conducting an exhaustive literature review and discussions with practitioners. Through an iterative process, indicators related to volume, diagnoses, referral patterns, quality of diabetes care and interdisciplinary care were created. For example, volume was defined as the number of patients per provider per day as well as the number of encounters per provider per day; the indicator of referrals describes the breadth of specialties. The interdisciplinary care measure described the team members’ activities in each practice. These indicators are being compared to health administrative data. For example, the DELPHI diabetes indicator has been validated using health administrative data (Hux et al. 2002).
Wait times and equity

Wait times are a major policy issue today, and provincial and federal governments are focusing attention on reducing wait times for cancer surgeries, cataracts, hip and knee replacements and CT/MRI scans. However, there is a paucity of data characterizing wait times in primary care, which is the stage at which the wait times “clock” really begins. A grant from the Canadian Institutes of Health Research (CIHR) has allowed us to use the researchable database to study this component of wait times that has hitherto not been examined in Canada. Using a referral as the unit of analysis, the date of family physician referral and the date of the specialist visit are abstracted, allowing us to construct a detailed picture of such wait times across southwestern Ontario.

Developing algorithms for case ascertainment of patients with chronic disease in EMRs

As EMRs become ubiquitous in the future, accurate identification of patients with a specific condition will become necessary. The DELPHI team is working on developing and testing an algorithm for accurate identification of patients with diabetes, using data elements readily available in an EMR. For example, there could be as many as four options for defining a person as diabetic: (a) if the patient’s active medication list includes a plasma glucose-lowering agent, (b) if the patient’s problem list contains diabetes mellitus (DM), (c) if the laboratory list contains more than one result for HbA1c, (d) if the patient has an ICPC diagnosis of either T89 or T90 (Type 1 DM or Type 2 DM) or (e) any combination of these. Future work will expand to study other chronic conditions.

Strengths and Limitations

The central limitation of using EMRs for research is that data are collected during patient encounters using a system that is designed for individual patient care, not research. EMRs provide users with many options for entering and storing data. First, users may enter data in free text form or by picking information from a list of structured terms. Second, the same type of information may be stored in multiple places in the EMR. Third, a variety of terms may be used for the same thing. In addition, information that is not important to clinical care may not be found in the record. Finally, digitized reports (which are not readily analyzable) may be stored in the EMR. These EMR features create challenges for researchers trying to extract and analyze the data. For example, to find a particular type of information, all possible locations in an EMR must be searched. Collaboration with information technology professionals is required to create a researchable database. Further, rigorous data quality assessment is also necessary to ensure that the data are suitable for research purposes. These are resource-inten-
Implementing and Maintaining a Researchable Database from Electronic Medical Records

Sive steps required simply to arrive at the data analysis stage of the research process. However, EMR data also have several advantages for research in primary healthcare. First, EMR data contain a great many variables on multiple aspects of PHC. This includes both clinical measures such as blood pressure, medications and laboratory test results, as well as health services variables such as referral types and wait times. Second, these data are longitudinal, allowing researchers to explore the natural history of conditions treated in PHC as well as care patterns over time. Third, assuming issues of access to data have been addressed, data may be collected relatively quickly for large numbers of patients. EMRs represent a unique source of data for answering questions about PHC.

Future Prospects and Projects

Projects that are in the conception or early analysis stage include work to improve understanding of clinical inertia in treating diabetics, focusing on time to treatment change and its determinants; characteristics of low back pain patients, their treatment and prognosis; a study of symptom progression to identify symptom clusters in primary care; and the development of metrics to quantify data quality in EMR-derived databases. Future studies will attempt to identify red-flag symptoms of rare and serious diseases, such as colon cancer, through case control and cohort studies.

The DELPHI Canadian experience is similar to the General Practice Research Database in the United Kingdom in that symptoms, diseases and interventions are coded (Jones et al. 2007), albeit using different classifications; however, the DELPHI Database, being regional, contains a smaller number of patients than the larger UK databases. As well, unless and until Canada requires these three types of structured data, they will be available only in smaller, well-resourced, purpose-built research databases. DELPHI’s usefulness in monitoring chronic disease management and preventive care is similar to the US studies (Vogt et al. 2007; Ornstein et al. 2007). Somewhat unique to the DELPHI Database are the health services research questions that are being answered, such as workload, wait times and the degree of interdisciplinary care.

In conclusion, EMRs are well suited to study both morbidity and workload of primary care providers in a thorough manner, measures that are not available through surveys (Okkes et al. 2002). In fact, only EMRs can assist researchers in their efforts to better characterize the content and quality of family practice and interdisciplinary primary healthcare. The advantage of EMR data is that they are comprehensive and longitudinal, covering all visits and clinically relevant interventions. The Centre for Studies in Family Medicine is committed to the long-term development of the researchable database, and is actively building an innovative EMR-based program of research.
ACKNOWLEDGEMENTS

Dr. Moira Stewart is funded by the Dr. Brian W. Gilbert Canada Research Chair. Dr. Amardeep Thind is Canada Research Chair in Health Services Research. The DELPHI Project was funded by the Canada Foundation for Innovation and Primary Health Care Transition Fund. Dr. Terry holds a post-doctoral fellowship funded by the Canadian Health Services Research Foundation and the Canadian Institutes of Health Research.

Correspondence may be directed to: Amardeep Thind, MD, PhD, Centre for Studies in Family Medicine, 245 – 100 Collip Circle, London, ON N6G 4X8; tel.: 519-858-5028; fax: 519-858-5029; e-mail: athind2@uwo.ca.

REFERENCES


Implementing and Maintaining a Researchable Database from Electronic Medical Records


