

C.O.M.P.E.T.E. RDB V1.0 – Experiences with an EPR Database

J. P. Kerby¹, K. Keshavjee¹, A.M. Holbrook^{1,2}, D. VanderMolen¹

¹ Centre for the Evaluation of Medicines, St Joseph's Hospital, Hamilton, ON, CANADA

² McMaster University, Hamilton, ON, CANADA

KEYWORDS: EPR, EMR, CPR, Medical Informatics, Database, PMRI

INTRODUCTION

Functional research databases based on electronic patient records (EPR) used in primary care are rare. Attempts at standardizing Patient Medical Record Information (PMRI) are underway, but relatively undeveloped. A lack of standards for data entry creates concerns about data validity and completeness and allows for varying interpretations of the data. The C.O.M.P.E.T.E. Study involves aggregate analyses of quality care benchmarks, utilization, and efficiency; individualized feedback to physicians about their own utilization compared to peers; and patient-specific decision support¹. We present the development of our prototypical research database

BACKGROUND

The COMPETE study is a successful implementation of an EPR (EMR) network in a primary care setting². Based in a major metropolitan area (Hamilton, Ontario, CANADA), the COMPETE network consists of 32 physicians in 18 clinics. Each clinic's network hosts both a medical administrative system (York Med Medical Desktop v5.2) and an EPR System (Purkinje Dossier of Clinical Information v1.4). These two systems are loosely coupled by a patient identification number. Data from the sites are routinely collected and assembled in secure COMPETE central headquarters. It is from these collections of data that we have created the COMPETE research database (COMPETE RDB v1.0). Our experiences, interpretations and future goals are described.

METHODS

Due to the varying demands on this dataset, our design had to be sufficiently robust, yet flexible. The range of requirements of the database is broad. It was necessary that our database field queries at the encounter, patient, caregiver and aggregate level. It was found the combination of a sound design and standard relational model of the database provided us with a suitably powerful and flexible framework.

It was necessary to import data from disparate sources to create our research database. For the administrative system (York Med), the data migration was somewhat direct since the data are stored in a commercial (Dbase III) format. Alternatively, data from the EPR system are stored in an encrypted format. To extract data, a utility within the EPR was used. The data were then migrated to the database using utilities internal to the Access DBMS. Since the data from the EPR extraction utility are essentially "flat" (non-indexed), a data model had to be designed *a priori*. This gave us license to create a design that suited our needs. Data are extracted from the clinics via secure modem download. The database was implemented using Microsoft Access 2000.

Data from the two systems are organized into approximately 30 major tables. In total, there are over 500 fields. The administrative system records data relating to appointments and billings. The EPR data consist of fields related to the clinical encounter. Some of the EPR fields are structured (e.g. DIN and GPI coded medications and ICD9CM coded diagnoses). The two systems (billing and EPR), are linked by the patient identification number. This index allowed us to relate patient encounters across the two systems. Other non-direct methods of reconciliation, such as the comparison of date of birth, postal codes, and caregiver fields allowed us to corroborate the reconciliation. Ancillary tables, such as complete ICD9CM diagnostic codes, billing codes and medications lists (DIN and GPI coded) were introduced as lookup tables.

Our procedures were developed using structured query language (SQL). Routines and programs were implemented using Visual Basic. Using this approach we assembled a suite of tools.

RESULTS

With the data assembled as one relational database, we proceed to experiment with different data analyses. It was mandatory at this stage that clinical expert be closely involved. The interpretation of medical data requires both technical and medical expertise. We present some of our results.

Table 1 - Demographics

COMPETE RDB v1.0	York Med (billing)	Purkinje (EPR)
# of encounters	201,871	131,103
# of patients	45,617	26,511
# of diagnoses	177,897	155,207
% male	42 %	42 %
% female	58 %	58 %
% > 65 (age)	15 %	21%
% 18 - 64	60 %	57 %
% < 18	25 %	22 %

Reconciliation of Diagnoses – An Example

In Canada, most physicians are reimbursed for patient care on a fee-for-service basis through a Provincial insurance scheme. The Provincial government requires the submission of an invoice containing patient health card number, the visit data, a single diagnostic code (similar to ICD9), and a procedure code.

Charting and billing practices are non-standard and vary by physician. It is expected, for various reasons, that there will be a discrepancy between the EPR entry and the billing entry (e.g. for a given patient encounter). Interpretation of these two databases has proven to be challenging due to factors such as variability in physician recording behavior, and ambiguities posed by data entered as text. We aim to quantify this difference by assessing it on a per-encounter, and per patient basis. Due to the variability of explanations for this phenomenon, we do not expect to mitigate the problem. Our aim is to identify areas in which this association between the two systems can be improved. The product of this we expect to be a better quality of data, and a more valid association between the two systems.

In order to quantify the number, SQL queries were designed and implemented. These queries were used to collect from the database, those encounters that met our requirements. The condition for the clinical note to be created at the point of care is relaxed by the ability of the physician to record the encounter *ex post*. Through the use of a "visit date" field and a "Service Date" field in the billing systems, we were able to determine whether a clinical encounter note in the EPR was related to a billing transaction.

Using our relational database, we counted the number of times that an encounter in the EPR had a corresponding billing in the administrative billing. Criteria were that the entries on both sides matched in patient identification number, date of visit and diagnosis. The results are presented in Table 2.

Table 2 – Analysis of EPR Diagnoses

Diagnoses Group	% EPR Diagnoses with matching billing	% EPR Patients with matching billing
All Diagnoses	20 %	38%
Cardiovascular Diagnoses	41 %	57%
Musculoskeletal Diagnoses	32 %	35%
Diabetes Diagnoses	50 %	62%

Since the information in the EPR record supplements the billing record, we expect that this type of analyses can be used to improve and corroborate the data from the billing system. It is worth noting that there is room for such improvement given the extensive use of billing data in provincial policy making activities. A convergence towards a more accurate provincial billing database would enable more effective policy decisions. Although such endeavors may be beyond the domain of our study at this time, the experience is nonetheless, noteworthy and valuable.

Future Research

Given the ability to match diagnoses across systems, our research is now aimed towards further research in this area. Of particular interest is the identification of those diagnoses, which are either poorly or well recorded. Additionally, indices such as the co-morbidity of diseases, are both desirable and well suited to our database design.

Other Interesting Results

COMPETE RDB v 1.0 has proven to have other useful applications.

- 1) Routine generation of Physician feedback, showing quality of EPR entry as compared to their peers.
- 2) The generation of patient lists for the identification of groups for drug utilization quality of prescribing analysis including new federal medication harm warnings.
- 3) Creation of aggregate statistics, such as top ten prescribed medications, top ten diagnoses.
- 4) Aggregate analyses of quality care benchmarks, utilization, and efficiency.
- 5) Analyses to support Case Management (data quality control).
- 6) Quality of the recording of allergies.

DISCUSSION

Standard relational database technology offers vehicles for the representation of data collected from EPR and Medical Administration systems. Once the data are imported into the database, various interpretations are facilitated. Researchers of such a dataset are able to access patients, encounters, physicians and groups. Additionally, this very dataset can be used to identify areas of low data integrity, thereby allowing case management apparatus to intervene, and in turn improve the quality. The ability to perform both static and dynamic analysis was very beneficial. For example, our dataset supported static analysis, such as top ten medication lists, as well as dynamic assessments, for example, time sensitive identification of medication overlap.

Quality of data is paramount. Given the ability to interpret activities such as billing or prescribing practices, we were able to determine what needed to be improved. Additionally, a cross section of expertise is required when building and interpreting such a database. Our study enjoyed the luxury of on-site clinical, statistical and computer personnel.

It was noted that we will rapidly exceed the capacity of the Access database product, and have planned to migrate to Microsoft SQL server version 7.0. It is possible to estimate, given the current rate of data entry, the rate of growth, thereby allowing for estimates of database size. Such estimations should be made before committing to a database product.

The ability to interpret the EPR and administrative data allows us to concentrate on improving data entry in the system. We were able to determine where data quality was an issue, and assign commensurate resources. This enabled us to improve the quality of our database through case management techniques (e.g. extra training, user groups, comparison to peers, and other incentive providing methods).

As stated, we are now faced with the challenge of using the EPR database to provide encounter-specific support at the point of care. We feel that a thorough investigation of data entry quality and variability, and management to optimize data entry into the EPR will contribute to decision support success.

REFERENCES

1. Holbrook, A.M, Goldsmith C, Tucker J, Park L, Advancing the Measurement of Quality of Care using EMR, TEPR Proc., 1999, vol 1, pp 848 – 855.
2. Holbrook, A.M, Langton, KB, Keshavjee K, Troyan S, Ford PT, Sebalt R, Goldsmith C, Levine MAH, MacLeod SM, Is Electronic Medical Software ready for evidence-based therapeutics in Primary Care? Can J. Clin Pharmacol, 1998.
3. Purkinje Inc., Clinical Information Extractor, Administrator Guide, Version 1.4CA, 1999
4. Purkinje Inc., Dossier of Clinical Information User Guide, Version 1.4, Oct 30, 1998
5. YorkMed Medical Desktop v5.2 Users Guide., York-Med Systems Inc., 1999
6. Klander, L, Mercer, D, Access 2000 Developer's Black Book, The Coriolis Group, 1999.
7. Microsoft Corp., Microsoft Office 2000 Visual Basic Programmer's Guide, Microsoft Corp., 1999.

BIOGRAPHY

Jane Kerby is the Database Manager for the COMPETE Project, St Joseph's Hospital, Hamilton, ON, Canada. With a background in computer science and economics, Jane has spent 7 years as a physical scientist for the government of Canada, specializing in the analysis of disparate data sources in environmental health sciences. Prior to this, she was a software specialist for Digital Equipment Canada, and mainframe operations manager for a subsidiary of SunLife Canada.