Scraping if Intraoperative Textual Report for Acquisition and Storage of Clinical Data from the Epic Electronic Medical Record (EMR)

Presenting Author: David C Kramer, MD, Department of Anesthesiology, Montefiore Hospital and Medical Center and Albert Einstein College of Medicine. New York, NY 10467

Co-Authors: Avinash Malaviya BS MS, Albert Einstein College of Medicine, New York, NY 10461
Alban Cela, BS, Albert Einstein College of Medicine, New York, NY 10461
Steven Henick, BS, Albert Einstein College of Medicine, New York, NY 10461
Sheldon Goldstein, MD, Department of Anesthesiology, Montefiore Hospital and Medical Center and Albert Einstein College of Medicine, New York, NY 10467

Introduction: Since enactment of The American Recovery and Reinvestment Act of 2009, Electronic Medical Records (EMR's) have been broadly instituted in hospitals. Many anesthesia departments have implemented these systems, and others continue to do so. While automated and procedural-driven capture of intraoperative clinical information has expanded, much of this data is not archived for research and Quality Improvement (QI) efforts. Some institutions have excelled at this endeavor; however, many Anesthesia departments find the process of data retrieval and archiving into research and QI-friendly formats to be resource intensive and financially prohibitive. With these constraints in mind, our group has endeavored to access and digitize textual information from native clinical reports into a spreadsheet format that can then be easily imported into statistical programs.

Methods: After IRB approval, the charts of 1,066 patients, who underwent spine surgery at our Hospital, from January 1, 2016 to June 30, 2018, were reviewed. APACHE, MySQL server, and phpMyAdmin (PHP) computer programs were installed on a personal computer running Microsoft Windows 10. A set of clinical data to be extracted from these anesthesia records was enumerated by our team. In total, 54 parameters were identified for extraction. Each clinical report was hand copied using two sets of two keystroke commands and pasted into Microsoft Word 2010. A sample of 60 charts was chosen and examined for characters that demarcate the boundaries (loci) of clinical data to be extracted. A program was developed in PHP to identify these loci and to extract the clinical data between them. When the loci failed to identify the demarcation of data, the clinical report was re-examined, and conditional statements were added to the original PHP function. For each group of clinical datasets, a SQL statement was written in PHP to import these data points into 14 related tables in MySQL server, using SQL procedural language.

Results: Once a stable version of the PHP scraping program was finalized, the native clinical anesthesia reports of 1,066 patients were analyzed. The program proved to be very accurate in extracting data and placing it into the MySQL tables. In total, over 58,000 data points were analyzed with 97.1% accuracy. The inaccuracies were related to parsing of textural information without loss of data.

Conclusion: Our data scraping program has proven to be successful in extracting clinical information from a pre-existing EPIC report form and placing it in a retrievable database system, with greater than 97% accuracy. The advantages of this approach to data acquisition are that: 1) it is extremely inexpensive; 2) has low resource utilization; 3) data acquisition and storage are rapid; (4) it requires no utilization of hospital or departmental-based IT resources; 5) the data output can be readily accessed with simple SQL commands, or through a project driven dashboard;6) it can be completely isolated from the Internet or hospital networks to prevent breach of Health Care Information and; 7) the spreadsheet format can be easy importing into statistical packages for analysis . To improve on the 2.9% parsing oversight, we will in future versions of this application import a dictionary of terms to enhance the accuracy of the program. This technique of surface scraping pre-existing reports could be used to facilitate research and Quality Improvement efforts at small to medium-sized anesthesia departments that do not have resources of continuous IT support

Images:

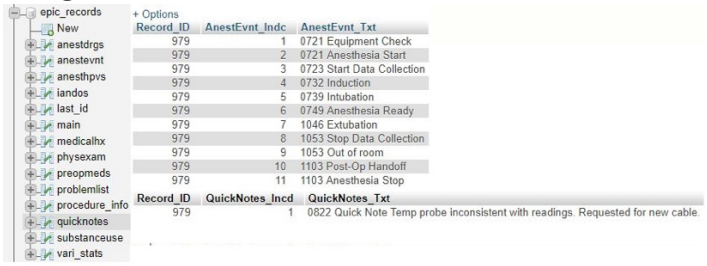

Figure 1 Image of database structure, sample capture of Anesthesia Events and QuickNote entries from one patient.