

MODERNIZING THE CHART REVIEW PROCESS USING TEXT RECOGNITION TO MINE ELECTRONIC MEDICAL RECORDS

Ryan H Rochat¹, Gail J Demmler-Harrison²

¹ Baylor College of Medicine, Department of Pediatrics, Infectious Diseases

² Texas Children's Hospital, Pediatrics, Infectious Diseases

Background: The Electronic Medical Record (EMR) has become a modern compendium of health information, from broad clinical assessments down to an individual's discrete heart rate. The wealth of information in these EMRs holds promise for clinical discovery and hypothesis generation. Unfortunately, as these systems have become more robust, mining them for clinical information is hindered by the overall data architecture, and often requires the expertise of a clinical informatician to extract data. Consequently, many research efforts rely on manual chart review for extraction of clinical information, a process that is both time consuming and prone to transcription error. We propose that a modernized chart review process, which capitalizes on the structure of information presented to the clinician through the digital workspace, can expedite this arduous task.

Materials/Methods: Here we present a program which can parse output from Epic Hyperspace®, generating a relational database of clinical information. The ICD-10 codes, laboratory values, and antiviral administration records for 190 patients seen in our Congenital Cytomegalovirus Clinic at Texas Children's Hospital in Houston were obtained using this approach.

Results: Manual extraction of this data took 1.5 hours to complete, and the scripts to parse the data took less than 1 second of real-CPU time to complete. Data from these patients included: 3800 ICD-10 codes, 33,000 individual laboratory values, and 800 antiviral administration records. Compared to conventional chart review, this approach was approximately 100-fold faster and free of any transcription errors.

Conclusions: High-throughput data mining tools have the potential to improve the feasibility of studies dependent upon information stored in the EMR. When coupled with specific content knowledge, this approach can consolidate months of data collection into a day's task.