

Introduction

Background

Electronic Health Records (EHRs) are now the standard means of documenting patient encounters for urologists. The raw data contained in these EHRs do not integrate well with national registries or quality improvement databases for two reasons:

- The EHRs for patient encounters are not standardized, inviting ambiguities into the data collection process.
- There is no software to extract these data *en masse*, and thus manual data entry into registries is still necessary.

These roadblocks make data entry into quality improvement registries both error-prone and costly.

Registry for Stones of the Kidney and Ureter (ReSKU™)

The ReSKU registry is a longitudinal, prospective database for patients with urinary stone disease. Paired with this registry are EHR templates that standardize each of the patients' clinical encounters. Thus, each datum in a clinical note corresponds to a variable in the registry.

This link between EHRs and the registry eliminates ambiguities in data entry, but not the need for manual data entry into the registry. **In this analysis, we evaluate the efficacy of software designed to automate the transfer of data from ReSKU clinical notes to the ReSKU registry.** This software would eliminate the need for manual data entry.

Objectives

- To determine the degree of discrepancy between manually entered data and data entered automatically.
- To determine which features of the registry increase the likelihood of discrepancies between these data.
- To investigate the cause of these discrepancies.

Methods

Patient Population

- Records from all patients presenting to the urology clinic with clinical evidence of nephrolithiasis who consented to be a part of the ReSKU program were included.
- Records from patients without both manually entered and automatically entered data were excluded.

Data Format

- Patient encounters were recorded into one of four clinical encounter templates: New Patient, Operative, Post-Operative, and Follow-up encounters.
- Each template contains three types of data: single answer multiple choice questions, many answer multiple choice questions, and free text.

Methods (Con't)

Statistical Analyses

- Ordinal regression was used to determine which features were associated with inconsistencies between manual and automated entries.

Software

- Data from clinical encounters were manually entered into ReSKU clinical encounters the Epic (Madison, WI) EHR system.
- MySQL, version 8.0.11, was used to extract records, and R, version 3.4.0, was used to transform extracted records.
- Data were stored in a Research Electronic Data Capture (REDCap) database to ensure HIPAA compliance.²

Results

- 391 manually entered and automatically transferred records from 149 patients were compared.
- Matching manual entries with automatically entered data revealed 2,859 discrepancies on a total of 67,441 data points, for a **discrepancy rate of 4%**.
- **Free text data had 2.1 times the odds of being discrepant** when compared to single answer multiple choice data (95% CI 1.97 – 2.5). When free text was dropped, matching rates increased from 96% to 97%.
- Compared to data from New Patient Encounters, **data from Post-Operative Encounters had 3.6 times the odds of being discrepant** (95% CI 3.0 – 4.3).
- Only three root causes were identified in 100 discrepancies:
 - Mistakes in manual data entry into the registry.
 - Improperly parsed free text by the software.
 - Clinicians and scribes replacing EHR templates with free text, preventing the ETL from recognizing the data.

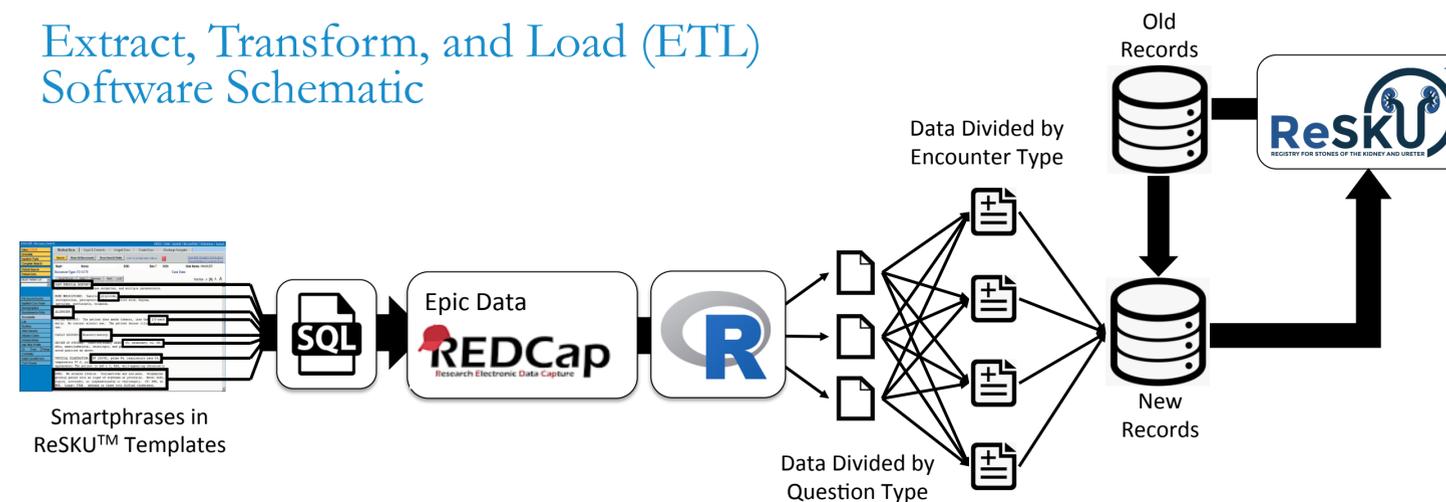
Discussion

- Automated extraction from EHRs into a registry is possible with a high accuracy that avoids errors in manual entry.
- Using these systems requires training, as evidenced by the identified root causes of discrepancies.
- Reducing free text entries should be considered when designing integrated database systems.

References

- 1 Chang HC, Tzou DT, Usawachintachit M, et al. Rationale and Design of the Registry for Stones of the Kidney and Ureter (ReSKU): A Prospective Observational Registry to Study the Natural History of Urolithiasis Patients. *Journal of Endourology*. 2016;30(12):1332-1338.
- 2 Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of biomedical informatics*. 2009;42(2):377-381.

Extract, Transform, and Load (ETL) Software Schematic



Extract from Epic's Clarity Database

ReSKU patient data is extracted from Epic with MySQL. These data are then loaded into REDCap using REDCap's API.

Transform Data

Data are loaded onto an R dataframe and split into separate frames by question type (single answer, multiple answer, and free text). These newly formatted frames are then organized according to encounter template, then organized by patient, and formatted for ReSKU.

Reconcile with Old Records and Load

Old records are pulled from the ReSKU Database and overwritten by new records when applicable. The new records are pushed back into ReSKU.