Digital Transformation of Healthcare

Assessing Data Quality

Michoel Snow, MD PhD, Glen Ferguson, PhD

Center for Health Data Innovations

Data Quality

Overview

Analysis is only ever as good as the data its built upon.

Overview

Analysis is only ever as good as the data its built upon.

- What is data quality? What makes data high quality vs low quality?

Overview

Analysis is only ever as good as the data its built upon.

- What is data quality? What makes data high quality vs low quality?
- Where along the process can you affect data quality?

Overview

Analysis is only ever as good as the data its built upon.

- What is data quality? What makes data high quality vs low quality?
- Where along the process can you affect data quality?
- How can you design a study to collect high quality data (Quality assurance)?

Overview

Analysis is only ever as good as the data its built upon.

- What is data quality? What makes data high quality vs low quality?
- Where along the process can you affect data quality?
- How can you design a study to collect high quality data (Quality assurance)?
- How can you identify and correct errors during and after data collection (Quality control)?

Overview

Analysis is only ever as good as the data its built upon.

- What is data quality? What makes data high quality vs low quality?
- Where along the process can you affect data quality?
- How can you design a study to collect high quality data (Quality assurance)?
- How can you identify and correct errors during and after data collection (Quality control)?

- Data quality consists of the objective [accuracy, validity (not outside range of possibilities, all data is for the same pt, formatting requirements, DICOM dates), reliability (dx matches problem list matches coding), legibility (units, shorthand)] and subjective [completeness]

- Steps
  - Definition/Design - lack of clear definitions for data items/collection, incompatible units, precision, scope, depth
  - Collection - not enough documentation (drug given/dosage altered but no start and end date), non-adherence to data definitions (collecting data outside of protocol time), human variance/error (bp cuff, RR, incorrect units), Orders are placed (procedures, medications) which are not connected to a rationale or sufficient reason
  - Processing - interpretation ('initial' lab, diagnosis date), coding error (mis-entering information such as order of birthdate, or height as 9 cm instead of 90 cm), random (mistyping, illegible handwriting), software errors, Assigning codes to problems treated vs problems tested and ruled out, which complaints do you code/document (doctors as coders)

Overview

Analysis is only ever as good as the data its built upon.

- What is data quality? What makes data high quality vs low quality?
- Where along the process can you affect data quality?
- How can you design a study to collect high quality data (Quality assurance)?
- How can you identify and correct errors during and after data collection (Quality control)?

- quality assurance - training of personnel (mock exams and reporting), site visits, reduce open-ended questions
- quality control - data monitoring (compared to independent source), hand verification, entering data in twice (by different sources), consistency checks

Extract, Transform and Load (ETL)

- Extract
  - Extract data from source(s)
- Transform
  - Modify the data for the purposes of analysis or further querying
- Load
  - Load the data into the target database

EDW and Clarity

Walkthrough EDW

Sources

- WHO data quality
- Healthcare Data Warehousing and Quality Assurance
- (2002). Defining and improving data quality in medical registries JAMIA, 9(6), 600-611.