Monitoring and Managing Data Quality in the INPC

Wanzhu Tu
Tim Imler
Dick Roscam
Shaun Grannis
Goal

To develop tools and processes to enable corrective feedback mechanisms that better identify and address data quality issues for data captured in the INPC.
Benefits / Impact

• Improved data quality
  – Increased confidence in the data
  – Improved decision making
  – Greater assurance that data is being received, data is present
  – Ensuring that quality attributes are curated, flagged where curation is needed
Target Users

• HIE operators (IHIE, Regenstrief)
• Customers of the system
  – Clinical organizations
  – Researchers
Four Phase Approach
Phase 1

Ensuring processes are operational: “Am I receiving data at an expected rate?”
Clinical Data Monitoring: An Example

Frequency of Data in Appointment Table over Time
Phase 2

Ensuring global data characteristics are consistent/expected: “Have the {patient demographics/chief complaint} characteristics, chief complaint changed?”
Figure 3: Monthly entropy (bits) for chief complaints classified into specific syndrome categories, stratified by source system for 5 high-volume emergency departments. Note the shift in values for one source that changed registration software.
Phases 3 & 4

- Phase 3: Implementing specific validation checks to flag suspected erroneous data – data we don’t want to commit to the database. ("Is BMI = 3?")
- Phase 4: Implementing trending rules to determine whether newly received data is consistent with prior patient data, or erroneous. ("Does the current weight represent > 60% decrease from previous?")
Resources Needed

• People, Time, Money
• People
  – Analysis / Development
    • Data managers
    • Data scientists
    • Statistical Support
  – HIE Operations
    • IHIE Data managers
    • IHIE Systems Engineers
  – HIE Customers
    • ????