Approaches to continuous improvement using large-scale data sets

Distributed Queries

DIGITAL DATA PRIORITIES FOR CONTINUOUS LEARNING IN HEALTH AND HEALTH CARE

INSTITUTE OF MEDICINE

Rich Elmore Coordinator, Query Health
• Distributed Queries - Strategic context
• Worked examples – Mini-Sentinel
• Query Health
Why a Distributed Database?

- Data Partners maintain HIPAA-mandated contractual control of their PHI
- Local content experts maintain a close relationship with the data
- Data Partners have the best understanding of their data and its uses; valid use and interpretation of findings requires input from the Data Partners.
- Easier to manage consent
- Lessens scale of breach / risk exposure / competitive exposure
- Accuracy, timeliness, flexibility, sustainability

Don’t wait for a pendulum swing here: Privacy is a healthcare constant
Distributed Queries

- Distributed Query Challenges
  - Absence of standards
  - Integrating each data source is a heavy lift
  - Cross-organizational governance

- Yet, path-breaking work is underway
  - ISDS Distribute
  - Primary Care Information Project
  - FDA’s Mini-sentinel
  - HMO Research Network
  - MDPHNet
  - i2b2 / SHRINE networks
  - DARTNet
  - OMOP
  - CDC’s BioSense 2.0

- Questions that return population measures (aggregate results) related to disease outbreaks, post-market surveillance, prevention, quality performance, etc.
Environmental scans identified data quality challenges

- Difficult to express a clinically intuitive, consistently computable query.
- Lack of semantic equivalency among systems and among users of systems.
- No commonly understood way to express clinical concepts such as Type 2 Diabetes and Asthma.
- Clinicians in the same practice, using the same clinical system are likely to code differently.
- Each organization establishes its own value sets – there are no starter sets that are maintained and usable.
- Other challenges exist such as missing data, the meaning of dates and many other interpretive questions related to disparate data sources.
Prospective look at proposed EHR standards’ impact on data quality

- Standardized set of vocabularies / code sets
- MAJOR improvement

<table>
<thead>
<tr>
<th>Data</th>
<th>2014 Edition (proposed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immunizations</td>
<td>CVX – Aug 15, 2011</td>
</tr>
<tr>
<td>Problems</td>
<td>IHTSDO SNOMED CT – Jan 2012</td>
</tr>
<tr>
<td>Procedures</td>
<td>ICD-10-PCS/HCPCS &amp; CPT-4</td>
</tr>
<tr>
<td>Lab Tests</td>
<td>LOINC 2.38</td>
</tr>
<tr>
<td>Medications</td>
<td>RxNorm – Feb 6, 2012</td>
</tr>
<tr>
<td>Race &amp; Ethnicity</td>
<td>OMB standards</td>
</tr>
<tr>
<td>Preferred Language</td>
<td>ISO 639-1:2002</td>
</tr>
<tr>
<td>Preliminary Determination of Cause of Death</td>
<td>ICD-10-CM</td>
</tr>
<tr>
<td>Smoking Status</td>
<td>Current every day; current some day; former; never; smoker, current status unknown; and unknown if ever smoked</td>
</tr>
<tr>
<td>Encounter Diagnoses</td>
<td>ICD-10-CM</td>
</tr>
</tbody>
</table>
• PopMedNet is proven across several distributed query networks, including Mini-Sentinel
• Uniquely supports the policy guidance from HIT Policy Committee
• Targeting full implementation of the Query Health proposed standards
FDA's Mini-Sentinel Program to Evaluate the Safety of Marketed Medical Products

A functioning distributed database and querying system.

And announcing a Query Health pilot.

Richard Platt / Jeffrey Brown
Harvard Pilgrim Health Care Institute
Harvard Medical School
for the Mini-Sentinel Investigators
March 23, 2012
FDA Amendment Act of 2007

- Mandates FDA establish capacity to use electronic health data to assess safety of marketed drugs
  - Data covering at least 100 million people required by mid-2012
- FDA is addressing drugs, biologics, and devices
Mini-Sentinel

• Develop scientific operations for active medical product safety surveillance
• Create a coordinating center with continuous access to automated healthcare data systems
Mini-Sentinel Partner Organizations
Mini-Sentinel Data Partners
Environmental scans identified data quality challenges

• Difficult to express a clinically intuitive, consistently computable query.
• Lack of semantic equivalency among systems and among users of systems.
• No commonly understood way to express clinical concepts such as Type 2 Diabetes and Asthma.
• Clinicians in the same practice, using the same clinical system are likely to code differently.
• Each organization establishes its own value sets – there are no starter sets that are maintained and usable.
• Other challenges exist such as missing data, the meaning of dates and many other interpretive questions related to disparate data sources.
The Mini-Sentinel Distributed Database

- Populations with well-defined person-time for which medically-attended events are known

- 126 million individuals*
  - 345 million person-years of observation time (2000-2011)
  - 44 million individuals currently enrolled, accumulating new data
  - 27 million individuals have over 3 years of data

*As of 12 December 2011. The potential for double-counting exists if individuals moved between data partner health plans.
The Mini-Sentinel Distributed Database

- 3 billion dispensings
  - Accumulating 37 million dispensings per month

- 2.4 billion unique encounters
  - 40 million acute inpatient stays
  - Accumulating 41 million encounters per month including over 400,000 hospitalizations

- 13 million people with ≥1 laboratory test result

*As of 12 December 2011*
Mini-Sentinel Distributed Analysis

1- User creates and submits query (a computer program)
2- Data partners retrieve query
3- Data partners review and run query against their local data
4- Data partners review results
5- Data partners return results via secure network
6- Results are aggregated
Example:
Rapid evaluation of drugs for smoking cessation and cardiac outcomes
Smoking Cessation Drugs and Cardiac Outcomes

Smoking Cessation Drugs and Cardiac Outcomes

FDA indicates intent to query Mini-Sentinel

Smoking Cessation Drugs and Cardiac Outcomes

- FDA indicates intent to query
- 4PM FDA provides final specs

Smoking Cessation Drugs and Cardiac Outcomes

- FDA indicates intent to query: 7/4/2011
- 4PM FDA provides final specs: 7/5/2011
- 6PM Programs distributed to 17 data partners: 7/6/2011
- 7/7/2011
- 7/8/2011
Smoking Cessation Drugs and Cardiac Outcomes

FDA indicates intent to query

4PM FDA provides final specs

6PM Programs distributed to 17 data partners

9AM Report delivered*


* High level summary with data from 13 data partners; complete report on 7/12
Query Specifications

- **Population:** New users of varenicline or bupropion (comparator)
  - First dispensing of bupropion or varenicline (180 day look back)
  - No cardiac outcome (below) or more general cardiac/atherosclerosis diagnosis (ICD-9 code 414.0x) in prior 180 days
  - Cohorts
    - All
    - Tobacco use disorder code (305.1), any setting, in prior 180 days

- **Exposure:** First treatment course
  - Bridge gaps ≤7 days to create treatment episode
  - Extend “treatment effect” for 7 days after presumed last exposure

- **Outcome:** Composite cardiac outcome codes
  - Diagnosis code in inpatient or ED setting during treatment course
    - Acute MI (410.xx) OR Intermediate coronary syndrome/unstable angina (411.1)
    - OR Acute coronary occlusion without MI (411.81)
## Results from 17 data partners

<table>
<thead>
<tr>
<th></th>
<th>New users</th>
<th>Person-time (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Varenicline</td>
<td>261,000*</td>
<td>32,000</td>
</tr>
<tr>
<td>Bupropion</td>
<td>746,000</td>
<td>210,000</td>
</tr>
<tr>
<td>With tobacco code</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Varenicline</td>
<td>90,000</td>
<td>11,000</td>
</tr>
<tr>
<td>Bupropion</td>
<td>113,000</td>
<td>23,000</td>
</tr>
</tbody>
</table>

* Nearest 1,000
Cardiac event rates, tobacco cohort

**Rate ratio = 0.97**

**Rate = 5.00**

**Rate = 5.14**

<table>
<thead>
<tr>
<th>Cardiac events</th>
<th>Varenicline + tobacco</th>
<th>Bupropion + tobacco</th>
</tr>
</thead>
<tbody>
<tr>
<td>Events/1,000 pers-years</td>
<td>56</td>
<td>118</td>
</tr>
<tr>
<td>Person time</td>
<td>11,197</td>
<td>22,942</td>
</tr>
</tbody>
</table>

* New users after ≥180 day washout
Cardiac events relative rates, tobacco cohort

Rate ratios and 95% confidence intervals

Incidence rate ratio

Adjusted for these factors

None  Age  Sex  Age/sex  Age/sex/health plan
Caveats

- Intended to be a quick look, not a final answer
- Result doesn’t exclude excess risk
- Exposures may be missing or have misclassified indication
  - Smoking cessation meds may not be covered
    - Potential missing exposures
    - Intentional misclassification of indication
- Cohort may be unrepresentative
  - Tobacco code identified a minority of smokers, presumably not typical
- Outcomes may be misclassified
  - No verification of coded diagnoses
- Potential for residual confounding
  - Smoking intensity
  - Comorbidities, including depression; other
Summary

- Demonstrated ability to rapidly query 300 million person years of experience
  - Defined population with complete eligibility and claims
  - Data quality checked in advance
  - Results evaluated for consistency by age, sex, year, site, dispensings, and amounts dispensed

- Distributed network approach required no transfer of Protected Health Information
Prasugrel and Prior Stroke/TIA

- Prasugrel indicated to prevent thrombotic cardiovascular events in selected patients with acute coronary syndrome who are to be managed with percutaneous coronary intervention.
- It is contraindicated in patients with a history of transient ischemic attack (TIA) or stroke
- Prasugrel and clopidogrel users’ prior history compared
Clopidogrel and Prasugrel: Prior Stroke or TIA

<table>
<thead>
<tr>
<th></th>
<th>Clopidogrel (153,191)*</th>
<th>Prasugrel (6,997)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior stroke</td>
<td>25,820</td>
<td>540</td>
</tr>
<tr>
<td>Prior TIA</td>
<td>11,815</td>
<td>134</td>
</tr>
</tbody>
</table>

* New users after ≥365 day washout
Conclusions / Limitations

- Some Prasugrel users have a prior diagnosis of TIA or stroke
  - Fewer than for clopidogrel users
- ICD-9 codes used for TIA and stroke not validated in Mini-Sentinel
- Longest look back for event was 1 year, patients that had an event >1 year prior would be missed
ARBs and celiac disease

- Potential signal identified in AERS database
- Review of cases inconclusive
### ARBs and celiac disease

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>New users</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOSARTAN</td>
<td>63</td>
<td>235,630</td>
</tr>
<tr>
<td>IRBESARTAN</td>
<td>10</td>
<td>40,071</td>
</tr>
<tr>
<td>OLMESARTAN</td>
<td>17</td>
<td>81,560</td>
</tr>
<tr>
<td>TELMISARTAN</td>
<td>5</td>
<td>24,596</td>
</tr>
<tr>
<td>VALSARTAN</td>
<td>50</td>
<td>153,159</td>
</tr>
</tbody>
</table>

**Notes:**
- ARBs: New users after ≥365 day washout; Celiac Disease: 1st dx code after >365 day without diagnosis.
Limitations

- Capture of relevant GI events may be incomplete
- Potential inclusion of irrelevant events
- Patients exposed to different agents may differ with respect to risk of GI symptoms
- Majority of exposures limited to a few months duration
- Observed risk doesn’t exclude excess
ARBs and celiac disease

Modular Program Type: MP 3 - Drug Use – Incident Outcomes
(See online specification for details: http://www.mini-sentinel.org/data_activities/details.aspx?ID=111)

Date Posted:

Medical product exposures of interest:

This Modular Program execution included 7 unique exposures, all in the Angiotensin II Receptor Blocker (ARB) drug category. The exposures were defined using National Drug Codes (NDCs identified by FirstDataBank), limited to the oral formulations, identified in the Mini-Sentinel outpatient dispensing file. The 7 drugs included were:

- Candesartan
- Eprosartan
- Irbesartan
- Losartan
- Olmesartan
- Telmisartan
- Valsartan
Welcome to Mini-Sentinel

Mini-Sentinel is a pilot project sponsored by the U.S. Food and Drug Administration (FDA) to inform and facilitate development of a national active surveillance system, the Sentinel System, to improve the safety of FDA-regulated medical products.

Mini-Sentinel is one piece of the Sentinel initiative, a multi-faceted effort by the FDA to develop a national electronic network that would complement existing methods of safety surveillance.

Mini-Sentinel Collaborators include data and academic partners that provide access to health care data and support ongoing scientific, technical, methodological, and organizational initiatives.
A Vision for Broader Use of Electronic Health Information in Evidence Development
Query Health

- An ONC-sponsored S&I Framework open government initiative
- Standards and specifications for distributed population queries.
- “Send questions to the data”
- Data sources including EHRs, HIEs, PHRs, payers’ clinical record or any other clinical record.
- Voluntary collaborative networks
- Declarative questions build on NQF / CMS work on population measures
- Aggregate responses
  - Patient level information secure
  - Support questions related to disease outbreak, quality, CER, post-market surveillance, performance, utilization, public health, prevention, resource optimization and many others.
- Dramatically cuts cycle time for deployment of population measures (e.g., quality measures) from years to days

NYC / NY State Pilot

Information Requestors

- NYC PCIP
- NYS DOH

Data Sources

- RHIO
- EHR

Flow: Sends Query to Data Sources

Flow: Distributes Query Results to Information Requestor
Distributed Queries
Relationship to Meaningful Use

• Bending the curve towards transformed health

• Distributed queries
  – Foundational to the digital infrastructure for a learning health system
  – Focus on the patient and patient populations
  – Ensuring privacy and trust

• For more information:
  – Mini-Sentinel.org
  – QueryHealth.org