Sentinel

International Comparison of Approaches to Common Data Models for Comparative Effectiveness Research

International Population Data Linkage Network Conference 2018
Banff Centre, Banff, Alberta, Canada
September 12, 2018

Jeffrey Brown, PhD

DEPARTMENT OF POPULATION MEDICINE
HARVARD MEDICAL SCHOOL
Harvard Pilgrim Health Care Institute

Disclosures

- I am an employee of Harvard Pilgrim Health Care Institute.
- I have research funding from FDA, NIH, PCORI, BBCIC, IMEDS, and Pfizer.
- I am the inventor of PopMedNet, an open source software application used to support distributed health data networks.
- I consult for IBM, Roche, and Bayer on issues related to the use of observational data for real-world evidence generation.

Data holders, lives covered, data access model, and network governance

- 18 data partners: 17 health insurers and 1 national hospital system
  - 9 integrated delivery systems with EHRs
  - All can request access to full medical charts
- ~300m member identifiers; ~67m current members
- >14 billion dispensings and > 13 billion medical visits
- Data access: Distributed data model - send questions to the data, securely return *information* via PopMedNet
- Governance: FDA directs all querying, data model changes, and tool development; Sentinel Operations Center distributes all network queries

https://www.sentinelinitiative.org/sentinel/data/snapshot-database-statistics; www.popmednet.org

Process for transforming a repository's data into the common data model

- Common data model created based on guiding principles developed in partnership with collaborators and based on FDA's specific needs
- Partners transform local data to the common data model based on the data dictionary, documentation, and ongoing guidance
- Initially, weekly calls to share knowledge and lessons learned; now monthly
- Rigorous data checking and characterization process after every data refresh; data must be approved by operations center before use
- Data model changes implemented using standard progress that includes distinct Discovery, Planning, Testing, and Implementation phases

www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model/sentinel-common-data-model
www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model/data-quality-review-and-characterization
www.sentinelinitiative.org/rss/sentinel-data-quality-assurance-practices

Target audience(s), process of identifying queries and knowledge dissemination plan

- Sentinel resources used at FDA's direction and managed by the Sentinel Operations Center (SOC)
- All queries are initiated by FDA
- Data model and tools publicly available
- Query results posted on public website, with complete details
  - Query specifications, dates, code sets, and analytic tools used (version number)
- By policy and guiding principles, the Sentinel Distributed Dataset held by partners can be used for other purposed at their discretion
  - Designed for FDA with other uses considered
  - IMEDS, BBCIC, NIH Collaboratory, multiple industry sponsored studies

Two key challenges faced by the network and the lessons learned

Sentinel

Data Standardization and Change Management

- Data curation requires substantial ongoing resources and trust
- Data model changes have important change management ramifications for partners, data curation, and tools
- Small change can be complicated

Managing Expectations for Speed and Data Completeness

- Highly curated data and standard tools enable rapid response
- But all queries are subject to validation processes and verification of appropriateness
- Programming “tweaks” are risky in a distributed network