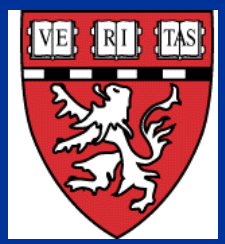
Defining and Evaluating Possible Database Models to Implement the FDA Sentinel Initiative

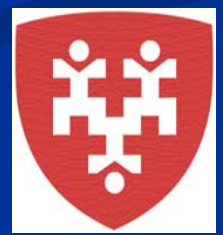
Second Annual Sentinel Initiative Public Workshop
The Marriott Metro Center
Washington, DC

Jeffrey Brown, PhD

January 11, 2010



Department of Population Medicine
Harvard Medical School and Harvard Pilgrim Health Care Institute



Contract and Contributors

Contract: Defining and Evaluating Possible Database Models to Implement the Sentinel Initiative. US Food and Drug Administration; Contact No. HHSF223200831315P

Submitted: May 2009

Report Authors:

Jeffrey Brown
Kimberly Lane
Kristen Moore
Richard Platt

Project Objectives

- Describe Sentinel System user needs
- Define potential database models (ie, how to organize data)
- Compare & contrast data models with user needs
- Identify environments & systems operating under various models
- Recommend characteristics & capabilities for Sentinel database models

Key Database Model-related Questions

- What does the system need to do?
- What data are needed to meet system needs?
- Where will the data be stored?
- How will the data be analyzed?
- Is a common data model needed
  - What should the model look like?

Selected Primary Uses of Sentinel

- Adverse event surveillance & signal strengthening
- Confirmatory safety studies
- Monitor adoption & diffusion
- Augment registry information
- Calculation of background incidence rates
- Better information regarding appropriate use of electronic data/ validity of coding
- Many other potential uses

What Data Are Needed?

- Minimum data needs for primary uses
  - Exposures (with dates)
  - Outcomes and comorbidities
  - Enrollment and demographic information
- Additional desirable data capabilities
  - Linkage across datasets (cross-sectional; longitudinal)
  - Clinical details (lab results, radiology, full-text records)
- Very few data elements are needed for most intended uses of Sentinel
- Needs are distinct from those of payors to pay claims and providers to support care delivery

What Data Are Needed?

- Systems with administrative & claims coupled with EMR data are most useful, *but represent a small minority of the U.S. population*

Recommendation: Initially focus on health insurer administrative & claims databases covering defined populations

Where are Data Stored?

- Centralized: data holders send data to a central location; data stored outside the physical control of data holder
- Distributed: data holders maintain physical control of their data behind their firewalls, protected by their security procedures and rules

Recommendation: Distributed Model

- Allows data holders to maintain physical control over their data and its uses: ➡ acceptable to data holders.
- Can accomplish about as much as centralized model; in principle it can do all the same things.

How are Data Analyzed in a Distributed Model?

1. Single analytic program distributed without local modification; requires local datasets to have same format (*common data model*)
2. Data holders individually implement a single study protocol

<u>Recommendation</u>: Single Analytic Program

- Ensures that complex analytic approaches are implemented identically and results comparable across institutions

Implementation of a Common Data Model?

- Common data model: set of definitions for the structure of databases and data elements
- Can be virtual or physical
  - Physical: extract, transform, and load process
  - Virtual: software layer to map local concepts to system concepts; requires real-time data or query transformations

Recommendation: Initial implementation should require a physical transformation for all or portion of a population

Types of Common Data Models

- Encounter-based person-level model
- Drug and condition era model (person-level)
- Person-level summary model
- Summary data model

- Many variants to these
- Many ways to enhance functionality through standardization of terminology

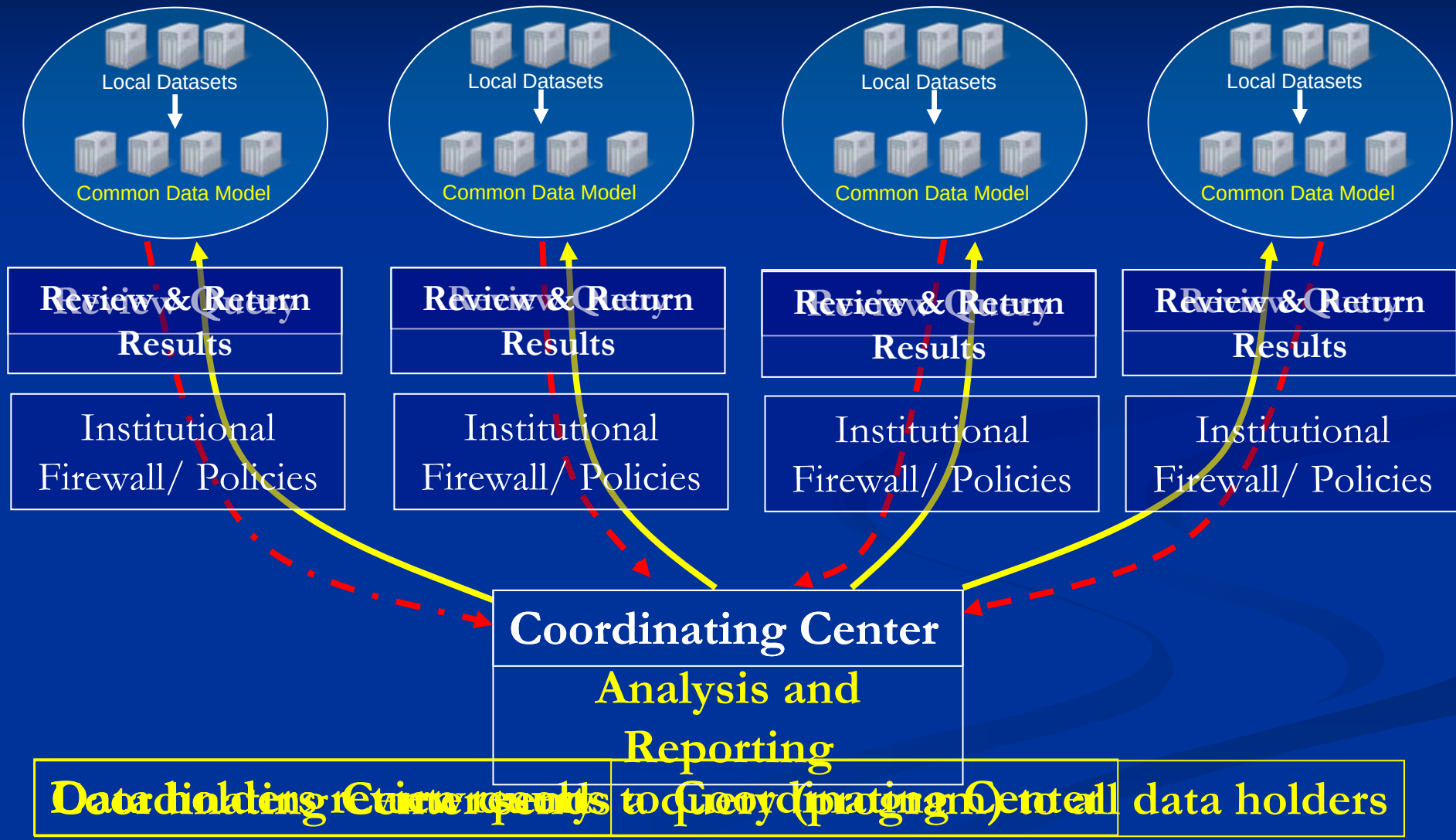
Additional Data Model Characteristics of Relevance to Sentinel

- Linkage: Medical charts
  - Access to detailed information from the full text records is essential for primary users
    - Allows validation of outcomes & some exposures
    - Provides information on co-existing conditions, indications, and other data to elucidate findings
- Linkage: External data sources
  - Can provide data beyond that found in administrative & claims databases or EMRs

Additional Data Model Characteristics of Relevance to Sentinel

- Linkage: Between institutions
  - Identify individuals across different care settings
  - Longitudinally identify individuals across data holders
- Timeliness
  - Interval until data becomes available for analysis
  - Varies by data source and system

Illustration of Proposed System

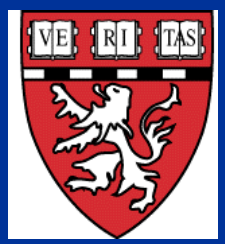


Defining and Evaluating Possible Database Models to Implement the FDA Sentinel Initiative

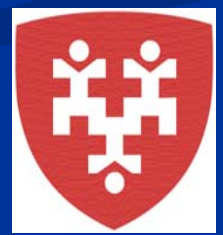
Second Annual Sentinel Initiative Public Workshop
The Marriott Metro Center
Washington, DC

Jeffrey Brown, PhD

January 11, 2010



Department of Population Medicine
Harvard Medical School and Harvard Pilgrim Health Care Institute



ENGELBERG CENTER for Health Care Reform at BROOKINGS

Sentinel Initiative Public Workshop

Monday, January 11, 2010