

Mini-Sentinel Common Data Model

Guiding Principles

November 2010

Version 1.0

Introduction

The primary goal of the Mini-Sentinel pilot is to build and operate a national public health surveillance system to improve the safety of FDA-regulated medical products, including drugs, biologics, and devices. Mini-Sentinel is a major element of the Sentinel Initiative, the FDA's response to a Congressional mandate to create an active surveillance system using electronic health data for 25 million people by 2010 and 100 million people by 2012.

The Mini-Sentinel pilot will undertake three major types of activities: (1) prospective evaluation of accumulating experience about specific medical products and specific suspected safety problems; (2) evaluation of the impact of FDA actions (e.g., labeling changes) on medical practice and health outcomes; and (3) rapid assessment of past experience in response to FDA questions about specific medical product exposures and health outcomes.

A wide range of Collaborating Institutions will provide access to data environments and other resources, including expertise, as needed to meet the epidemiologic requirements of Mini-Sentinel. In addition, representatives of the Collaborating Institutions will provide ongoing scientific, technical, and methodological expertise by participating in Mini-Sentinel in various capacities, including as members of the Planning Board, the Safety Science Committee, the three Mini-Sentinel Coordinating Center Cores (Data, Methods, and Protocol), and various Mini-Sentinel workgroups.

Mini-Sentinel uses a distributed data model¹⁻³ that gives Data Partners complete autonomy over access to and use of data in their possession. The distributed model requires development and implementation of a common data model to allow a single analytic program to be distributed and run identically in each data environment.

The Mini-Sentinel Coordinating Center (MSCC) Data Core coordinates the network of Data Partners and leads development and utilization of the Mini-Sentinel Common Data Model (MSCDM), a standard data structure that allows Data Partners to quickly execute programs against their local data. In addition, the MSCC Data Core facilitates creation of the individual Mini-Sentinel Distributed Databases (MSDD) at Data Partner sites using the MSCDM. The Data Core also works closely with the MSCC Methods and Protocol Cores. The MSDD refers to the data held and maintained by the Data Partners in the MSCDM format.

This document describes the Guiding Principles of the MSCC Data Core as well as the initial priorities and approach to the MSCDM.

Guiding Principles

The MSCC Data Core coordinates the network of data partners who actively participate in the creation, implementation, updating, maintenance, enhancement, and use of the MSCDM and their MSDDs. The

design and implementation of the MSCDM strives for a high level of cross-institutional and longitudinal consistency and requires that data comparable in format and meaning are stored at all sites.

The following principles guide the development and maintenance of the MSCDM:

1. The MSCDM accommodates all requirements of Mini-Sentinel activities and may change to meet FDA objectives.
2. The MSCDM is able to incorporate new data types and data elements as needs indicate.
3. Development of the initial MSCDM and all enhancements requires input and acceptance from the Mini-Sentinel Data Partners.
4. Documentation of Data Partner specific issues and qualifiers that may impact use and interpretation of the data is crucial for the effective operation of Mini-Sentinel activities.
5. The MSCDM design is transparent, intuitive, well-documented, and easily understood by analysts, investigators, and stakeholders. It is easy for experienced analysts and investigators to use; special skills or knowledge beyond those commonly found among pharmacoepidemiologists and professional analytic staff is not necessary.
6. The MSCDM leverages evolving healthcare coding standards.
7. The MSCDM captures values found in the source data. When necessary, mapping to standard vocabularies is transparent. Validated mappings should be used whenever available.
8. Calculated variables should not be included in the MSCDM.
9. Distributed programs should be executed with minimal to no site-specific modification.
10. Data Partners have the best understanding of their data and its uses; valid use and interpretation of findings requires input from the Data Partners.
11. Only the minimum necessary information should be used and shared with authorized staff of the MSCC.
12. Data Partners may include "site-specific" information in their implementation of the MSCDM.

Initial Priorities and Approach to the Mini-Sentinel Common Data Model (v1.0)

The overall goal of Version 1.0 of the MSCDM is to build the foundation for Mini-Sentinel to begin active surveillance activities and to have the capability to quickly generate information in response to urgent public health needs. Initial functionality will rely on claims and administrative data with additional functionality to be added in subsequent years.

In order to achieve this goal, Version 1.0 of the MSCDM will be implemented by Data Partners representing at least 25 million lives. It was agreed that Version 1.0 should:

- i. Reflect the guiding principles
- ii. Focus on claims and administrative data elements
- iii. Leverage the cumulative experience of the data partners
- iv. Rely on existing and standardized coding schemas (e.g., ICD-9-CM, HCPCS/CPT, and NDC)
- v. Be compatible with claims-based components of existing CDMs (e.g., Observational Medical Outcomes Partnership, HMO Research Network Virtual Data Warehouse)
- vi. Include all of the data elements necessary to achieve the goals for Year 1 of the Mini-Sentinel pilot

Revisions and enhancements to the MSCDM are expected in subsequent years, including the addition of clinical information, incorporation of other data types and sources, and revisions based on lessons learned from use of the MSDD and other programs' CDMs. This may include adopting variables and formats developed by other programs.

References

1. Maro JC, Platt R, Holmes JH, *et al.* [Design of a National Distributed Health Data Network](#). Ann Intern Med 2009;151:341-344.
2. Brown JS, Lane K, Moore K, *et al.* [Defining and Evaluating Possible Database Models to Implement the FDA Sentinel Initiative](#); U.S. Food and Drug Administration: FDA-2009-N-0192-0005.2009.
3. Velentgas P, Bohn R, Brown JS, *et al.* [A distributed research network model for post-marketing safety studies: the Meningococcal Vaccine Study](#). Pharmacoepidemiology and Drug Safety 2008;17:1226-1234.