

Promoting the Effective Use of a Common Data Model

The U.S. Food and Drug Administration's (FDA) Sentinel Initiative pilot program, Mini-Sentinel, was launched in 2008 following a mandate from Congress to develop a system for postmarket risk identification and analysis using existing electronic health data.¹ Since its inception, the pilot has evolved into a distributed database, providing access to electronic health care claims records from 18 data partners covering more than 130 million individuals.² The Mini-Sentinel pilot has become a critical tool for FDA to monitor the safety of approved medical products in the postmarket setting. As the Mini-Sentinel pilot prepares to enter its fifth year, FDA is exploring opportunities to leverage existing resources and infrastructure to support ongoing collaboration and sustainability.

Under a cooperative agreement with FDA, the Engelberg Center for Health Care Reform (ECHCR) at Brookings convened a meeting in February 2013 entitled "Payer Involvement in the Sentinel Initiative." The purpose was to discuss the data partners' continued participation in Sentinel, and the possible expansion of Mini-Sentinel's capabilities to support safety surveillance and broader health care research, including outcomes and quality assessments. The data partners expressed enthusiasm about continuing their participation in Mini-Sentinel. However, they were concerned that the expansion of Sentinel beyond safety surveillance could contribute to the growing resource and administrative challenges imposed by data reporting requirements (e.g., accreditation, all payers claims databases, etc.) from other programs. In particular, data partners commented that the lack of a more broadly adapted common data model (CDM) has exacerbated the burden associated with formatting and aggregating data for the wide range of mandated health care research programs. CDMs are used to standardize data collection, and facilitate the exchange of information between data owners and groups requesting and utilizing data for secondary purposes. Data partners suggested that expanding the Mini-Sentinel common data model (MSCDM) to support other reporting requirements and evidence needs could reduce the burden of data collection if the model was broadly adopted.

To expand on the issues discussed in the February meeting, ECHCR convened a stakeholder meeting on May 8, 2013 entitled "Promoting the Effective Use of a Common Data Model (CDM)." Participants at the meeting included representatives from federal and state health research agencies, and several of the Mini-Sentinel data partners. Stakeholders were asked to consider the following questions:

- 1) What are the common data needs of the current programs using health care data?
- 2) Is it possible to develop a CDM or a series of compatible models to meet those needs?
- 3) Could the MSCDM serve as the basis for a more broadly adopted CDM?

These questions were intended to assess the feasibility of developing a shared CDM or set of compatible CDMs that could serve a wide range of research activities conducted at the state and federal level, as well as by private payers and research institutions. This type of tool, if possible, could reduce the burden of data collection and support the development of interoperable national health data resources.

¹ Food and Drug Administration Amendments Act of 2007. (2007). Public Law 110-85. Retrieved April 29, 2013, from <http://www.gpo.gov/fdsys/pkg/PLAW-110publ85/html/PLAW-110publ85.htm>.

² Mini-Sentinel Coordinating Center. (2012). Mini-Sentinel Distributed Database "At A Glance." Retrieved April 29, 2013 from http://mini-sentinel.org/about_us/MSDD_At-a-Glance.aspx.

Current State of Data Collection

While the primary use of health care claims and administrative data is to support reimbursement, these data are frequently used for secondary purposes by payers and external organizations. This information is currently collected by numerous groups to investigate a range of research and public health questions related to safety surveillance, program and outcomes evaluation, and effectiveness research. Federal agencies using these data include FDA,³ the Centers for Medicare & Medicaid Services (CMS),⁴ the Agency for Healthcare Research & Quality (AHRQ), and the Centers for Disease Control and Prevention (CDC).⁵ Additionally, a number of states are developing all-payer claims databases (APCDs) that collect health insurance claims data from all public and private payers.⁶

There are also groups working to establish broad data networks for conducting large-scale research in partnership with academic centers and health care delivery organizations. The National Institutes of Health's (NIH) Health Care Systems Research Collaboratory is seeking to improve how clinical trials are conducted by creating new infrastructure for collaborative research.⁷ The Patient-Centered Outcomes Research Institute (PCORI) is funding a National Patient-Centered Clinical Research Network Program that will conduct randomized, patient-centered comparative clinical effectiveness research.⁸ The ultimate aim of PCORI's program is to establish a rich national electronic data infrastructure that promotes comprehensive, longitudinal research centered on patient outcomes. Payers also have other reporting requirements for non-governmental organizations such as the National Committee for Quality Assurance (NCQA) to meet accreditation standards, and to report quality and performance measures.⁹

In most cases, these programs request similar data from payers but do not employ a common data format. The lack of standardization creates an administrative burden on the payers to curate and format a number of different datasets, and also potentially compromises the quality and comparability of the findings. The demand for these data will increase to meet the sophisticated evidence needs associated with health care reform (e.g., outcomes, cost-effectiveness, and quality research). Data owners and secondary users should collaborate to develop mechanisms which ensure that data collection, storage, and use employ standardized infrastructure, processes, and methods to support efficient and sustainable research.

Overview of the Mini-Sentinel CDM

A critical part of the Mini-Sentinel pilot's success has been the collaborative development and adoption of the MSCDM by all data partners. This allows the Mini-Sentinel Coordinating Center (MSCC) to centralize and streamline the development and refinement of methodologies, protocols, and other analytical tools, and to quickly aggregate and analyze results from across all partners.¹⁰ The structure and design of CDMs are typically driven by the type of activity, funding, objectives, and data sources available, differences in which have led to a

³ Mini-Sentinel Coordinating Center. (2012). Mini-Sentinel Distributed Database "At A Glance." Retrieved April 29, 2013, from http://mini-sentinel.org/about_us/MSDD_At-a-Glance.aspx.

⁴ Centers for Medicare & Medicaid Services Office of Information Products and Data Analytics. (2013). CMS Data Navigator. Retrieved April 29, 2013, from <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/ResearchGenInfo/OIPDA.html>.

⁵ Centers for Disease Control and Prevention. (2012). National Electronic Disease Surveillance System. Retrieved April 29, 2013, from <http://wwwn.cdc.gov/nndss/script/nedss.aspx>.

⁶ Miller, P. (2012, August). All-Payer Claims Databases: State Progress and Future of APCDs. Retrieved April 29, 2013, from <http://apcdouncil.org/sites/apcdouncil.org/files/2012-08-02%20APCD%20Overview.pdf>.

⁷ National Institutes of Health, the Health Care Systems Research Collaboratory. (2013). "About Us." Retrieved August 27, 2013 from <https://www.nihcollaboratory.org/about-us/Pages/default.aspx>.

⁸ Patient-Centered Outcomes Research Institute. (2013). Funding Announcements. Retrieved April 29, 2013, from <http://www.pcori.org/funding-opportunities/funding-announcements>.

⁹ National Committee for Quality Assurance. (2013). "About Us." Retrieved April 29, 2013 from <http://www.ncqa.org/AboutNCQA.aspx>.

¹⁰ Curtis LH, Weiner MG, Boudreau DM, Cooper WO, Daniel GW, Nair VP, Raebel MA, Beaulieu NU, Rosofsky R, Woodworth TS, Brown JS. (2012) Design considerations, architecture, and use of the Mini-Sentinel distributed data system. *Pharmacoepidemiology and Drug Safety* 21(S1):23-31.

diversity of CDMs employed in health research today. Many of these other models were referenced in the development of the MSCDM,¹¹ as were a set of core design principles developed by the MSCC; Data, Methods, and Protocol cores; and the Mini-Sentinel data partners.

One key principle was that the system be designed to protect patient privacy. The distributed data network enables participating data partners to maintain physical and operational control over their data, which allows the data partner to protect patients' information and privacy behind its firewall. Secondly, in order to engage data partners, the MSCDM had to consider the interests of the data owners to ensure the confidentiality of the data (including regulatory requirements and companies' internal clinical, economic, and risk and severity assessments). It was important that the data partners be able to use the MSCDM for internal research and that data collection not impose undue burdens on payers. The MSCDM was also designed to minimize the need for data transformation and variable derivation to reduce the work required by data partners and allow researchers to more easily reference source data.

Other guiding principles for the MSCDM include support for local, "site-specific" variables from data partners; ability to leverage existing and evolving data standards where they exist; and the ability to support a range of users through a transparent, intuitive, and well-documented design. Finally, the MSCDM was designed with the ability to evolve and incorporate new data standards, types and elements as they develop. The MSCDM and distributed network started small to address targeted surveillance questions. However, over time the MSCDM and its capabilities have expanded to include new data tables and types of data. The flexibility of the MSCDM may make it a good candidate for other kinds of health care research, including more sophisticated outcomes and quality assessments.

Opportunities and challenges of moving towards an expanded CDM

As discussed above, the principles and methodologies used to develop CDMs are shaped by the projects' objectives, scope, and available data sources. While a single (or set of compatible) CDMs could potentially support a number of stakeholder research efforts, the chosen data model(s) would need to be simple and flexible to support diverse data types and multiple use cases. Examples of use cases suggested include:

- Patient-level predictive modeling
- Health reform evaluations
- Electronic health record measure validation/group analysis
- Provider quality and resource use measures
- Episodic profiling and enhanced risk adjustment
- Comparative effectiveness
- Care quality improvement

Participants agreed on the importance of leveraging the lessons learned from the development of the MSCDM and other models in the creation of a more widely adopted CDM or set of compatible data models. While several participants supported the concept of using MSCDM as a base model for adoption and expansion, they noted that more detailed discussions would be required. In particular, additional work is needed to identify and address limitations of the current MSCDM. Participants cited several potential issues that may arise with an expanded CDM and data network, including concerns about increased complexity and data quality.

Data Complexity

Stakeholders highlighted the current MSCDM's flexibility as attractive features, and cautioned that these attributes should be preserved to maintain MSCDM's strengths. Thus far, the MSCDM has relied primarily on

¹¹ Examples of initiatives which use a centralized analytic approach and common data model include Vaccine Safety Datalink project, the HMO Research Network, the Meningococcal Vaccine Study, the Observational Medical Outcomes Partnership (OMOP), the Post-Licensure Rapid Immunization Safety Monitoring (PRISM) project.

standardized claims and administrative data, which require minimal variable manipulation and allows for a simpler data model. Researchers can then also easily refer to source data to examine any potential issues or anomalies resulting from CDM queries. Some noted that it might be difficult to maintain the MSCDM's simplicity when adding data from less standardized sources, like EHR, which generally requires more data transformation, ontology mapping, and variable derivation to be brought into a common format. Other CDMs, such as the Observational Medical Outcomes partnership (OMOP), include EHRs data and are therefore more complex than the MSCDM.

Expanding Mini-Sentinel's CDM through additional data tables may be fairly straightforward, but data partners must invest resources in populating them. This can be particularly resource intensive during the first major stages of transformation. Data partners argued that adding smaller data tables gradually as needed was more cost-effective than adding large amounts of data at one time.

Data Quality

Several data partners commented on the need to ensure that data quality is maintained and continually improved, particularly as data models become more complicated. Participants also cautioned against the proverbial "candy shop" of data and noted that care and intentioned design was necessary to cultivate the most informative data resource. This will require continued work on data collection and quality assurance. For instance, EHR data is typically collected by physicians, who may receive minimal, if any, training about the data capture process and their role in it. Some data partners started conducting outreach and education efforts to more effectively involve clinicians in research, simultaneously improving the data capture and collection process. They are also engaged in learning how data entry fits into physicians' clinical practice and work flow.

Likewise, data partners work to continually validate and assess data quality. The Mini-Sentinel system builds on the data partners existing data quality processes by including further checks in the initial data extraction, transformation, and loading (ETL) process and during data refreshes, and by conducting validation studies. Multiple stakeholders indicated that trust in data quality was a key factor for the continued engagement of data partners and researchers. Participants felt that concerns about data quality could be a potential impediment to expanding MSCDM from surveillance to more nuanced objectives, such as comparative effectiveness research, if researchers did not feel the data resource was as strong. Stakeholders agreed, however, that there needs to be a balance between validating data (often through resource-intensive processes like chart review) and keeping the system simple, efficient, and responsive.

CDM Management and Governance

An additional consideration in the development of a more broadly adopted CDM is the governance structure for its development, management, and maintenance. Mini-Sentinel, including the MSCDM, was designed with a governance structure that reflects FDA's Congressional mandate to conduct public health surveillance. FDA contracted with each of the data partners individually and the data partners' input was critical to the development of the MSCDM. The creation of a CDM that addresses broader research objectives or incorporates more diverse partners may potentially require a different management and governance structure.

Summary and Possible Next Steps

The data partners were very supportive of continued expansion of the MSCDM to enhance safety surveillance. Participants were divided on whether the MSCDM could support more nuanced research objectives such as clinical research, comparative-effectiveness studies, predictive modeling, or other types of research that would require incorporating demographics, systematic factors, and patient histories from EHRs. However, participants were generally very supportive of other groups which also utilize administrative and claims data adopting MSCDM.

Participants suggested that a possible pilot for an expanded version of the MSCDM might be the implementation of the states' APCDs. APCDs collect health insurance claims data to track eligibility, health care costs, utilization, and evaluate quality measures at the state level. Currently, ten states have created APCDs and six more states are in the implementation stage,¹² but the process has been piecemeal, with each state requiring slightly different collection of data in different formats. The current lack of standardization could create a substantial burden for payers and may have long-term implications for the comparability of findings at the inter-state, regional, and national levels. The non-profit APCD Council¹³ has been advocating for the adoption of a core set of data elements in a standard data structure for all APCDs to ease the burden on payers as well as to simplify the collection process.¹⁴ These data standards could be used to design a CDM used by APCDs nationally. While not every state might require data on every element, it would greatly simplify the submission process for the data owners.

Many of the organizations conducting health care research have already invested heavily in the development of their programs and data networks and might be hesitant to change directions. Achieving the objective of a more broadly adopted CDM will require stakeholders to collaborate in the development, implementation, and maintenance of a model that serves all their needs and provides a net benefit in terms of interoperability, simplicity, and flexibility. Meeting participants cautioned that strong leadership, substantial funding, and a collaborative effort will be critical to making this goal feasible.

¹² All-Payer Claims Database Council. (2013). Interactive State Report Map. Retrieved August 14, 2013, from <http://www.apcdouncil.org/state/map>.

¹³ All-Payer Claims Database Council. (2013). About APCD Council. Retrieved October 14, 2013, from <http://www.apcdouncil.org/about-apcd-council>.

¹⁴ All-Payer Claims Database Council. (2011). Standardization of Data Collection in All-Payer Claims Databases. Retrieved August 14, 2013, from http://www.apcdouncil.org/sites/apcdouncil.org/files/Standardization%20Fact%20Sheet_FINAL_for010711release_1.pdf.