May 8, 2013

B | ENGELBERG CENTER for Health Care Reform at BROOKINGS

Discussion Guide

Promoting the Effective Use of a Common Data Model

Background

Mandated by Congress to develop a system for postmarket risk identification and analysis using existing electronic health data,¹ the U.S. Food and Drug Administration (FDA) launched the Sentinel Initiative in 2008. Mini-Sentinel, Sentinel's pilot program, developed a distributed database of data partners with healthcare claims and electronic health records. The current data network is maintained by 18 partners and covers more than 130 million individuals. It captures approximately 4.1 billion unique encounters with the healthcare system and 3.7 billion product dispensings.²

FDA believes that the use of a common data model (CDM) by research and surveillance activities leveraging health care information could reduce the burden of data collection on data partners and could lead to the creation of future partnerships benefiting the public health.

Currently, there is a wide range of groups including health care payers, states, and federal agencies that make extensive use of data to evaluate the effectiveness of medical interventions and health care systems, to track patterns in population health, to identify and better understand potential safety issues associated with the use of medical products, and to assess cost-effectiveness, among other objectives. FDA utilizes electronic healthcare data, including health care claims, for postmarket risk identification and analysis of medical product safety through the pilot of the Sentinel Initiative, Mini-Sentinel.³ Centers for Medicare & Medicaid Services (CMS) is currently engaged in various activities focused on developing and implementing research to support the evaluation of quality in health care services, comparative- and cost-effectiveness, and care outcomes.⁴ The Centers for Diseases Control and Prevention (CDC) has developed a national electronic disease surveillance system that transfers data from the healthcare system to public health departments.⁵ Groups such as the National Committee for Quality Assurance (NCQA) collect and report performance measurement information for managed care organizations to inform purchasing decisions.⁶ Several states are developing all-payer claims databases (APCDs) that compile data from private and public payers to assess health care utilization and cost

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.

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 Poon, E.G., A. Wright, et al. (2010). Relationship between Use of Electronic Health Record Features and Health Care Quality: Results of Statewide Survey.

patterns.⁷ Efforts for this initiative have been working on the development of standards for data formatting and transmitting for secondary uses (e.g., National Council for Prescription Drug Programs and the Accredited Standards Committee [ASC X12]). The Patient-Centered Outcomes Research Institute (PCORI) recently announced plans to develop a national infrastructure to collect data and to support patient-centered comparative clinical effectiveness research.⁸ All of these efforts are critical to ensuring the safety of medical products, informing health care reform initiatives, improving the quality, efficiency and value of the nation's health care system.

Health care claims data are a rich source of information that are used for a variety of surveillance and research needs. The ongoing adoption and building out of electronic health and medical records provide additional opportunities to support large-scale national surveillance and research efforts. The groups using the data are faced with the daunting task of accessing data from a variety of different sources with very different data structures and systems. Health care organizations that hold these data are challenged by the need to leverage data from sources that were not designed for either surveillance or research purposes and to provide information from that data to multiple users in different formats. Both the data owners (e.g., payers and health care organizations) and those groups using the data (e.g., federal agencies, states, academia, researchers, etc.) face challenges in developing, implementing and maintaining the infrastructure, processes, and methods needed to make meaningful use of this valuable information.

In the current resource-constrained environment, it is important to identify potential opportunities to increase the efficiency and effectiveness of existing efforts, as well as potential new efforts to reduce duplication and the cost of conducting analyses.

Common Data Models

CDMs can provide a means to standardize data formats in order to streamline the exchange of information between data owners and groups requesting and utilizing data for secondary purposes. The structure and design of CDMs are driven by both the type of activity and the data sources available. There have been a number of projects that utilize CDMs such as the Vaccine Safety Datalink project,⁹ the HMO Research Network,^{10,11} the Meningococcal Vaccine Study,¹² the Observational Medical Outcomes Partnership (OMOP),¹³ the Post-Licensure Rapid Immunization Safety Monitoring (PRISM) project,¹⁴ Informatics for Integrating Biology and the Bedside (i2b2),¹⁵ Electronic Primary Care Research Network

⁷ Miller, P. (2012, August). All-Payer Claims Databases: State Progress and Future of APCDs. Retrieved April 29, 2013, from http://apcdcouncil.org/sites/apcdcouncil.org/files/2012-08-02%20APCD%20Overview.pdf.

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

⁹ Centers for Disease Control and Prevention. (2013). Vaccine Safety Datalink (VSD). Retrieved April 29, 2013, from http://www.cdc.gov/vaccinesafety/Activities/VSD.html.

¹⁰ HMO Research Network. (2013). Collaboration Toolkit: A guide to multicenter research in the HMO Research Network. Retrieved April 29, 2013, from

http://www.hmoresearchnetwork.org/resources/toolkit/HMORN_CollaborationToolkit.pdf.

¹¹ HMO Research Network. (2013). Retrieved April 29, 2013, from www.hmoresearchnetwork.org.

¹² Velentgas P, Bohn R, Brown JS, *et al.* (2008). A distributed research network model for post-marketing safety studies: the Meningococcal Vaccine Study. *Pharmacoepidemiology and Drug Safety* 17(12):1226-34.

¹³ Foundation for the National Institutes of Health. (2013). Observational Medical Outcomes Partnership. Retrieved April 29, 2013, from http://omop.fnih.org.

¹⁴ Coloma PM, Trifirò G, Patadia V, and Sturkenboom M. (2013) Postmarketing Safety Surveillance. *Drug Safety* 36 (3):183-197.

¹⁵ National Center for Biomedical Computing. (2013). Informatics for Integrating Biology and the Bedside (i2b2). Retrieved April 29, 2013, from www.i2b2.org.

(ePCRN),¹⁶ and Teradata.¹⁷ Several of these initiatives use a centralized analytic approach and common data model.

CDMs are often used for distributed research networks. Within a distributed network, the data partners maintain operational and physical control over their data, but transform and then store the data in the common format. A CDM makes it possible to centralize the development of queries and testing of methodologies. In the absence of a CDM, each data partner would be required to translate a study protocol for execution against its own data format.¹⁸

Mini-Sentinel Common Data Model

Mini-Sentinel offers FDA the ability to quantify the occurrence of many outcomes within a defined population for whom medical product exposure is known. As a result, a range of public health surveillance activities are possible, ranging from calculating crude incidence rates to conducting protocol-based assessments using data from a significant fraction of the United States population. Data partners execute queries within their own data repositories and then return de-identified results to the Mini-Sentinel Operations Center for aggregation and analysis. Notably, these activities can be conducted in much less time and at much lower cost than previously possible.

In order to make the queries efficient, Mini-Sentinel partners collaboratively developed a common data model (MSCDM) to standardize administrative and clinical information needed for postmarket safety surveillance. In developing the MSCDM, the researchers reviewed the CDMs of the other distributed systems listed above.¹⁹ The CDM allows for methodologies and queries to be developed centrally by the Mini-Sentinel Operations Center and for queries to be carried out consistently across all data contributors. The distributed database is populated primarily by payer claims data, and incorporates a number of different elements to characterize health care exposures and outcomes. The CDM makes use of parent tables that are linked by a patient pseudo-identifier and can be queried through Mini-Sentinel:

- 1. enrollment data (unique person identifier, starting and end dates of coverage),
- 2. demographic information (birth date, sex, and race and ethnicity data),
- 3. dispensing data (national drug code, dispensing date, days supply, amount dispense),
- 4. encounter codes (provider and encounter types, discharge status),
- 5. diagnosis codes (ICD-9 codes),
- 6. procedure codes (ICD-9, HCPCS codes),
- 7. death (death date, imputation method if unknown date),
- 8. cause of death (date, diagnosis code, source of information),
- 9. laboratory results (selected test captured by Data Partners), and
- 10. vital signs (height, weight, tobacco status, blood pressure).

The MSCDM was designed to allow for the addition of new tables. The MSCDM has undergone expansion from its initial design to include tables on vital sign information, laboratory tests, and state immunization records to support the work of the Post-Licensure Rapid Immunization Safety Monitoring

¹⁶ Electronic Primary Care Research Network (ePCRN). (2013). ePCRN Overview. Retrieved April 29, 2013, from www.epcrn.bham.ac.uk.

¹⁷ Teradata. (2013). Retrieved April 29, 2013, from www.teradata.com

¹⁸ 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.

¹⁹ 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.

(PRISM) project.²⁰ Additional laboratory test types were added in Year 3 of the Mini-Sentinel pilot, and efforts such as incorporating data from registries and to establish mother-infant linkage to monitor *in utero* exposures are being explored. Other potential refinements include indicating plan type, patient zip codes, service provider specialty, and unique device identifiers for medical devices.

Based on the data transformed into the MSCDM, summary tables are generated that can be stratified by some combination of age group, sex, year or year-quarter, care setting, medical coverage status, and drug coverage status. These summary tables also remain with the data partners. Enrollment summary tables can be used to calculate crude prevalence rates, and diagnosis and procedure of interest summary tables can provide counts of affected individuals stratified by age group, sex, year and care setting. Drug and generic drug summary tables provide counts of out-patient dispensing of a drug of interest, again stratified by age group, sex, and year or year-quarter.

The Mini-Sentinel pilot has primarily focused on retrospective studies of exposures and outcomes. Future work, however, will include the prospective evaluation of newly approved medical products. This expansion of surveillance capabilities will include an expansion of the MSCDM which will allow for semiautomated analysis on a list of predetermined exposure and outcome characteristics.

Workshop Discussion: Exploring the Utility of a Common Data Model

In cooperation with FDA, the Engelberg Center for Health Care Reform at Brookings is hosting a meeting, "Promoting the Effective Use of a Common Data Model" to explore three central topics: 1) identifying commonalities in different surveillance and research needs, 2) exploring the utility of a CDM to meet those needs, and 3) the possibility of utilizing MSCDM as a base for broader adoption and expansion.

Participants are asked to consider the following questions to identify opportunities that may exist for future partnership and collaboration:

- What are stakeholders' data and evidentiary needs?
- What efficiencies could be created through the use of a single or select few CDMs?
- What are the limitations of the MSCDM in its current form?
- How could the current MSCDM be enriched to support additional research initiatives?
- What are the next steps in moving towards the use of a CDM to support broader research?

²⁰ Curtis L, Vajani M, Lauer J, and Brown J. (2013). Mini-Sentinel Common Data Model Expansion Plan. Retrieved April 29, 2013, from http://www.mini-sentinel.org/work_products/Data_Activities/Mini-Sentinel_Common-Data-Model-Expansion-Plan.pdf.