MINI-SENTINEL COORDINATING CENTER DATA CORE

YEAR THREE COMMON DATA MODEL (CDM) REPORT

REPORT OF DATA CORE ACTIVITIES, OCTOBER 2011 TO SEPTEMBER 2012

Prepared by: Mini-Sentinel Data Core and Data Partners

July 2013

Mini-Sentinel Coordinating Center Data Core
Year Three Common Data Model (CDM) Report
Report of Data Core Activities, October 2011 to September 2012

TABLE OF CONTENTS

I.   INTRODUCTION

This report describes the activities of the Mini-Sentinel Data Core during Year 3 - October 2011 through September 2012 - of the Mini-Sentinel pilot project. Ongoing activities are included in the report, as well as one-time activities that were undertaken during the project year.

A. OVERVIEW OF THE MINI-SENTINEL PROGRAM

Mini-Sentinel is a pilot program sponsored by the U.S. Food and Drug Administration (FDA) as a part of its Sentinel Initiative to inform and facilitate development of a fully operational active surveillance system for monitoring the safety of FDA-regulated medical products, i.e., the Sentinel System. Mini-Sentinel is a major element of the Sentinel Initiative, FDA's response to Section 905 of the Food and Drugs Administration Amendment Act (FDAAA) of 2007 to create an active surveillance system using electronic health data for 100 million people by 2012.

The Mini-Sentinel program currently focuses on three major activities:

- Assessments - Medical product exposures, health outcomes, and links between them
- Methods - Techniques for identifying, validating, and linking medical product exposures and health outcomes
- Data - Mini-Sentinel Distributed Dataset and tools used to access the data

Collaborating Institutions enable access to data environments and provide other resources to support meeting the requirements of Mini-Sentinel. In addition, representatives of the Collaborating Institutions provide ongoing scientific, technical, and methodological expertise by participating in the Planning Board, the Safety Science Committee, the three Mini-Sentinel Coordinating Center Cores (Data, Methods, and Protocol), project-specific workgroups, and other developmental activities. For additional information, please see www.mini-sentinel.org.

This report describes the Mini-Sentinel Data Core activities undertaken during Year 3 of the Mini-Sentinel program. The report covers the period October 1 2011 through July 31, 2012.

B. MINI-SENTINEL SCIENTIFIC OPERATIONS CENTER

The Mini-Sentinel Operations Center (MSOC) leads Mini-Sentinel's scientific and management operations, via the Scientific and Management Operations Centers. The Scientific Operations Center oversees the data infrastructure and overall operation of the program. It supports the scientific work of the Methods, Protocol, and Data Cores and all Mini-Sentinel project workgroups. The Scientific Operations Center is the central point of contact for the FDA and all Collaborating Institutions regarding scientific aspects of Mini-Sentinel (see **Figure 1**).

The Data Infrastructure Division oversees data development and data source documentation, as well as evaluation implementation activities of Mini-Sentinel. Individuals working within this Division possess expertise in database design, implementation, and analysis. Data Infrastructure Division staff are members of the Mini-Sentinel Data Core and support and work closely with the FDA, the Data Core, and Data Partners on these Mini-Sentinel activities.

1. Responsibilities of the Data Infrastructure Division

- Coordinate and support the activities of the Data Core
- Coordinate and oversee development and implementation of the Mini-Sentinel distributed data approach and common data model
- Document data sources and characteristics
- Assess data quality and characteristics
- Develop reusable analytic tools (e.g. Modular Programs)
- Develop standard operating procedures for writing distributed programs
- Coordinate Mini-Sentinel data activities and projects to ensure use of available tools and adherence to programming standards
- Provide programming and analysis support to workgroups, as necessary
- Develop and manage Mini-Sentinel public website and private secure communications systems

Figure 1. Mini-Sentinel Coordinating Center
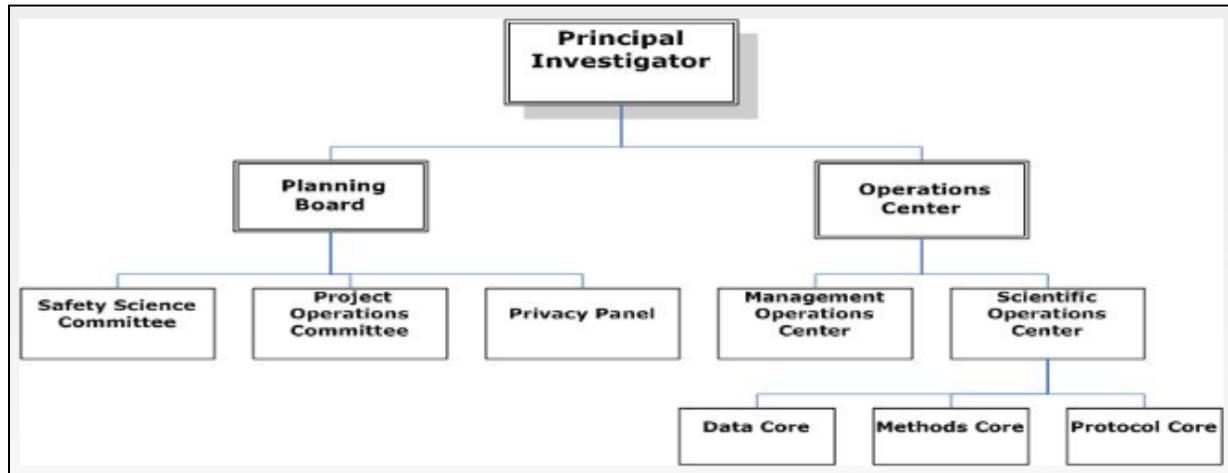
C. MINI-SENTINEL DATA CORE

1. Overview

The Mini-Sentinel Data Core directs the development and implementation of the Mini-Sentinel Common Data Model (MSCDM), distributed data approach, and related data standards and quality measures. The Data Core establishes additional workgroups as needed and interacts regularly with the Methods and Protocol Cores. A key responsibility of the Data Core is to facilitate communication across the Data Partners and manage the maintenance of the Mini-Sentinel Distributed Database, the data held by Data Partners in the MSCDM format. The Data Core also serves as the main conduit for communication among Data and Academic Partners, project workgroups, and other parties interested in data-related aspects of Mini-Sentinel activities.

2. Roles and Responsibilities

- Develop, implement, and manage a scalable and extensible common data model to meet the needs of Mini-Sentinel

- Incorporate national data standards, as appropriate, into development of the MSCDM and data analysis
- Create and update Mini-Sentinel distributed datasets that conform to the MSCDM
- Establish and implement data quality measures
- Lead strategic planning of data development
- Establish ad hoc data workgroups to investigate specific topics of interest
- Oversee and review data workgroup activities
- Develop, coordinate, and conduct data-related reviews and training for the FDA and Mini-Sentinel affiliate organizations
- Collaborate with Methods Core, Protocol Core, Operations Center, and FDA staff
- Communicate with external stakeholders as directed by FDA

3. **Members of the Data Core**

- Data Core Leaders
- Mini-Sentinel Operations Center staff
- Representatives from each Data Partner
- Representatives from FDA
- Additional analytical and technical staff as needed

4. **Members' Terms and Selection**

Member terms are one year and renewable. Data Core Leaders are selected by the Mini-Sentinel Principal Investigator and approved by the Planning Board. Data Partners and FDA representatives are chosen by their respective institutions.

5. **Data Partners**

Mini-Sentinel Data Partners with health plan administrative claims data in the MSCDM format include Aetna, HealthCore, Inc. (working with WellPoint data), the HMO Research Network, Humana, Kaiser Permanente Center for Effectiveness and Safety Research, OptumInsight and Vanderbilt University (working with Tennessee Medicaid data). The Mini-Sentinel includes other Collaborating Institutions that have access to additional data sources of interest for medical product safety surveillance, including laboratory data, electronic health record (EHR) data, inpatient systems, and disease and device registries. Efforts to incorporate these data areas into the MSCDM are ongoing and will continue to be the focus of activities in subsequent years.

D. DISTRIBUTED DATA APPROACH

Mini-Sentinel uses a distributed data approach in which Data Partners maintain physical and operational control over electronic data in their existing environments.[1-7]The Mini-Sentinel Common Data Model standardizes administrative claims and clinical information across Data Partners. Data Partners execute standardized programs provided by the Operations Center or project workgroups and typically share the output of these programs in summary form with the Operations Center and project workgroups. By

allowing Data Partners to maintain control of their data and its uses, the distributed model avoids or reduces many of the security, proprietary, legal, and privacy concerns of Data Partners, including those related to the Health Insurance Portability and Accountability Act (HIPAA)[i]. This approach also incorporates the need to have local content experts maintain a close relationship with the data. For example, only a local expert can easily and effectively trouble-shoot an unexpected finding or anomaly. In addition, the distributed model allows Data Partners to accurately assess, track, and authorize query requests, or categories of requests, on a case-by-case basis, and ensure that only the minimum data necessary are shared with the MSOC or FDA.

A mixed model is used on a case-by-case basis when evaluations require person-level intermediate analytic datasets, for example, when performing multivariate analyses.[1,3] A mixed model uses a distributed approach for analyses that can be conducted in a distributed manner (e.g., incidence rates, safety surveillance, identification of specific cohorts) and only transfers person-level data for combined analysis (e.g., case-control or cohort approach) if necessary. Only the minimum necessary data are transferred, which typically include 1 row per person with highly summarized aggregate information such as age in an age range, number of prior hospitalizations, and total days exposed to a treatment.

II.   OVERVIEW OF COMMON DATA MODEL

The MSCDM v3.0 includes 10 tables that represent specific data domains and the table structures for the Mini-Sentinel Summary Tables.[ii] This section describes the 10 data domain tables. The Mini-Sentinel Summary Tables are described in Section VII below. Each of the data domain table serves a specific purpose and the overall structure is designed to facilitate data access while preserving the granularity and nature of the source data. The data tables keep similar clinical concepts together and whenever possible keep the source "data streams" separate so that tables can be updated individually at different intervals if necessary. For example, outpatient pharmacy dispensings are kept separate from other claims sources so that the pharmacy table can be updated without affecting other tables in the data model. Details of the tables and each individual variable are available at www.mini-sentinel.org: Overview and Description of the Mini-Sentinel Common Data Model v2.1.

A unique person identifier is included in all tables to allow linkage across the tables and comprehensive view of patient care during an enrollment period. The unique person identifier is not a true identifier (e.g., Social Security Number), but rather a health-plan generated, alpha-numeric string that is unique to each person in the data files. Each health plan maintains a link between the unique person identifier and the true identifier, which is retained by the Data Partner. The person identifier is unique within a health plan and is not shared outside the health plan with either the MSOC or the FDA.

Each table is briefly described below.

Enrollment. The ability to ascertain who is eligible to receive specific kinds of care at any particular time is required for most Mini-Sentinel investigations. In many medical product safety evaluations, it is

---

[i] http://www.hhs.gov/ocr/privacy/

[ii] MSCDM v3.0 is the current version. As the MSCDM is revised, newer versions will replace the older documents. MSCDM v2.1 is available at http://minisentinel.org/data_activities/distributed_db_and_data/details.aspx?ID=105.

important to know the period of time during which an event of interest would be observed if it occurred. That is, confidence in the absence of care is often as important as the observation of a medical event.

The enrollment table contains records for all individuals who were health plan members during the period included in the data extract. The table includes the unique person identifier, the starting and ending dates of coverage, and flags for medical and pharmacy coverage. Patients can have multiple periods of coverage that are continuous or disjointed. Continuous periods of coverage are joined together into one period. For example, if a coverage period that ends on December 31 is followed by another that begins on January 1, the two periods are joined. A change in any variable, such as the drug coverage flag, in the enrollment table generates a new record even if the coverage is continuous. Disjointed periods of coverage—those that are separated by more than 1 day—are listed as separate records. Data Partners are not required to "bridge" gaps of more than 1 day in coverage; when appropriate, bridging will be incorporated into analysis programs based on the specific needs of the evaluation.

Most Mini-Sentinel evaluations use the enrollment table to verify the specific dates during which medical utilization identified in other tables (e.g., exposed to a specific medication) are eligible to contribute to an evaluation. The table structure is a simplification of the HMO Research Network's Virtual Data Warehouse (VDW)[8] enrollment table structure and similar in structure to the other common data models evaluated.

Demographic. The demographic table includes the unique person identifier, sex, birth date, race, and an ethnicity marker. However, only a subset of the Data Partners collects meaningful race and ethnicity information. The demographic table includes everyone found in the Data Partner database and is not limited to members included in the enrollment table. For example, everyone in the enrollment and dispensing tables must be in the demographic table, but the reverse is not true.

Dispensing. The dispensing table represents outpatient pharmacy dispensing captured by the Data Partners. Each outpatient dispensing to a patient is captured in the table. The table includes a unique record that lists the unique person identifier, dispensed date, dispensed NDC (in 11 digit format), and the days supplied and amount dispensed as listed on the dispensing record. Data Partners are instructed to process source transactions to remove rollback transactions and other adjustments before populating the dispensing table. This typically requires summation of dispensing information by unique person identifier, dispensing date, and dispensed NDC. No negative days supplied or amounts dispensed appear in the table and no corrections are made for values that are "out of range," such as 900 days supplied.

Individual dispensings can be linked to create treatment episodes based on any algorithm or specification necessary for the evaluation. For example, dispensings with out-of-range values can be cleaned or removed, and treatment episodes can be created on a case-by-case basis depending on the specific drug dispensed, patient cohort, or any other criteria as specified by the evaluation team.

Medications dispensed at discount pharmacies (e.g., Walmart, Target) may or may not be included in the table, depending on whether or not the pharmacy submits the claim to the health plan and whether the drug benefit includes dispensings at pharmacies external to the health plan. Similarly, the purchase of over-the-counter medications is only included in the dispensing table if the transaction is submitted via the pharmacy to the health plan (which is rarely the case). An analysis of pharmacy dispensing data

for 11 HMORN health plans found that OTC medications accounts for 2% to 9% of all outpatient dispensings between 2000 and 2007, although this rate of capture is likely to be a small portion of all OTC use.[9] Infused medications, vaccinations, and other medications (e.g., injections) provided directly by medical providers are captured in the separate **procedures table**, because those administrations are considered "procedures" within the existing medical coding nomenclature and are captured by the Data Partners in a separate data stream. A very small percentage (less than 0.1%) of outpatient dispensings represent NDCs for procedures.[9] Similarly, medications dispensed in the inpatient setting are captured in a separate data stream and are not included in the Dispensing Table.

Encounter: Each time a patient sees a provider in an ambulatory setting (including emergency department care) or is hospitalized, a record is entered into the encounter table. Each record within the table is a unique combination of person, admission/encounter date, provider, and care setting. For example, if a patient sees a primary care physician who sends the patient to the emergency department and the patient is later admitted to a hospital, the encounter table contains three records. Additional information in this table includes discharge date of the hospitalization, provider code, facility code, 3-digit provider zip code for the facility, Diagnosis Related Group assigned to the admission, the admitting source, the discharge status, and the discharge disposition.

Diagnosis: Each encounter, whether inpatient or ambulatory/outpatient, is associated with at least one diagnosis. Therefore, the diagnosis table is linked to the encounter table in a one-to-many relationship so that all the associated diagnoses are recorded in the diagnosis table. The diagnosis table includes one row for each unique diagnosis recorded during an encounter. The table also includes a flag for whether the diagnosis was recorded in the primary diagnosis field for the encounter (applies only to care in the inpatient setting), an indicator for the care setting in which the diagnosis was recorded, and an indicator for the type of diagnosis code. This "long and thin" table structure facilitates searching for specific diagnosis codes in large tables.

The diagnosis table can be used to identify disease cohorts or health outcomes of interest. The structure makes it easy to apply cohort algorithms, such as identifying patients with at least one inpatient diagnosis or two outpatient diagnoses of bipolar disease, or those with a primary inpatient diagnosis of stroke.

Procedure: Similar to diagnoses, each inpatient and ambulatory/outpatient encounter is associated with one or more procedures. Therefore, the procedure table is linked to the encounter table in a one-to-many relationship so that all the associated procedures are recorded in the procedure table. The procedure table includes one row for each unique procedure recorded during an encounter. The table includes the unique person identifier, the procedure code, an indicator for the care setting in which the procedure was recorded, and the specific type of procedure recorded (e.g., ICD-9 CM, CPT-4, HCPCS). Currently many coding standards are used to record procedures, including ICD-9 CM procedure codes, CPT-4 codes, and HCPCS codes; the table allows capture of any existing or future coding standards. This "long and thin" table structure facilitates searching for specific procedure codes in large tables.

The procedure table can be used to identify patients who have undergone specific surgical procedures (e.g., hip replacement surgery), received certain outpatient infusions, or received specific vaccinations.

Death: The Data Partners have various mechanisms for acquiring information about an enrollee's death. If a patient dies while in the hospital, the death is recorded in association with a related discharge

disposition. However, many patients die outside the clinical setting and the only clue to the death is the cessation of health utilization activity. Therefore, to confirm the death, many of the Data Partners link to local (state) death registries to update the death status of their members. This update is performed relatively infrequently—about once a year for most Data Partners. As a result, a two-year lag in death data is not uncommon. Within the death table, the death date is recorded, along with imputation method if the exact date is not known.

Cause of Death: Since each death can be associated with one or more contributing conditions, the death table is linked to a separate cause of death table that records diagnosis codes reflecting the underlying condition, along with coding dictionary used, type of contribution to the death, and the source of the information.

Laboratory: The laboratory table represents results and information from selected laboratory tests captured by select Data Partners. Because laboratory results can have different interpretations based on type of test or method of test administration, the model also includes variables for test subcategory, specimen source, patient location, result location, and result unit.

Aetna, HealthCore, Humana, Kaiser Permanente, and selected HMORN sites have implemented the following laboratory results: alkaline phosphatase (ALP), alanine aminotransferase (SGPT), total bilirubin, glucose, glycosylated hemoglobin (HbA1c), creatinine, hemoglobin, International Normalized Ratio (INR), fibrin d-dimer, absolute neutrophil count (ANC), lipase, troponin I, troponin T, platelets, creatine kinase total, creatine kinase MB fraction, pregnancy, and influenza.

Vital Signs: Nine sites are currently contributing information on height, weight, systolic and diastolic blood pressure, and tobacco status for this table.

Detailed information on the addition of the Laboratory and Vital Signs tables are included in the following section.

Summary Tables: There are nine prevalent Summary Tables that are used to enable rapid querying through the Mini-Sentinel Query Tool. These tables were added to the MSCDM last year for completeness and transparency. During Year Three, MSOC team began enforcing this new aspect of the model with all Data Partners by sending summary table test queries during the data refresh approval process.

III.   EXPANSION OF THE MINI-SENTINEL COMMON DATA MODEL

A. CLINICAL DATA ELEMENTS

1. Overview

In Year Two, the MSCDM was expanded to include Clinical Data Elements consisting of selected vital signs and laboratory results. The Mini-Sentinel Clinical data Elements workgroup led development and implementation of the addition of these variables to the MSCDM. Year Three activities included the addition of two new Data Partners to contribute laboratory results data, expansion of the number of laboratory tests contained in the model, update to the data model, and new data checks to provide

more detailed data characterizations. Additionally, more in-depth characterization of the vital sign data was conducted.

2. Selection of Additional Data Elements

The initial set of laboratory tests included in the MSCDM during Year Two were:

- Glucose
- Hemoglobin
- Hemoglobin A1c
- Creatinine
- Alanine Aminotransferase
- Alkaline Phosphatase
- Total Bilirubin
- International Normalized Ratio
- D-dimer
- Lipase
- Absolute Neutrophil Count (HealthCore only)

The initial set of vital signs included during Year Two was **height, weight, systolic blood pressure, diastolic blood pressure and tobacco status**. No additional vital signs data elements were added during Year Three. Data Partners engaged in the first round of laboratory additions included Kaiser Permanente, HMORN, and HealthCore. In Year Three, eight new laboratory results are being incorporated:

- Troponin-T
- Troponin-I
- Platelets
- Creatine Kinase total
- Creatine-Kinase MB fraction
- Pregnancy
- Influenza testing
- Absolute Neutrophil Count (KP, HMORN, Humana)

Also in Year Three, Aetna and Humana began contributing clinical data to the Mini-Sentinel Distributed Database (MSDD). As Aetna and Humana were new participants in the clinical data activities, they added the 10 lab results from Year Two as well as a subset of the Year Three labs. Furthermore, some Partners that were involved in the Year Two clinical data element activities will continue to update the Year Two laboratory tests, but will not be including the Year Three laboratory test results.

Further information regarding building the clinical components' data model can be found in the Year 2 MSCDM report which is available online.[10]

3. Revisions of the Data Model for Clinical Data/Standards and Terminologies

The addition of new Data Partners contributing labs created additional data mapping challenges related to variation in the ways laboratory tests and results are recorded across the sites. Developing a uniform

standard for labeling results, units and ranges were priorities that were necessary to simplify and standardize the data to enable distributed queries. To enable more transparency of the mapping processes implemented at each Data Partner, the data model was expanded to include source and transformed values of several data elements including result values and result units. Through these changes, we will standardize different results that have the same meaning (such as "+", "POSITIVE", "POS"). As an example of the issues faced with standardization within and across sites, **Table 1** shows a sample of the approximately 70 different units that describe numeric platelet count results across all of the Data Partners.

Table 1. Sample of Different Units Describing Numeric Platelet Count Results

Numeric Platelet Count Units
10E3/uL
10e3/uL
10e9/L
E9/L
BIL/L
bil/L
K/CMM
k/cmm
K/CU MM
K/CUMM
K/MCL

In order to enable distributed queries this type of variation must be normalized. The entries with the 'k' or '10e3' numerators logically translates to thousands, and the 'uL', 'mcl', and 'cumm' are variations of a microliter which Mini-Sentinel has abbreviated as 'UL'. Therefore, all of the above for units should be standardized to k/UL. The entries with '10e9','E9', and 'bil' in the numerator and L in the denominator refer to billions per liter. One liter is equal to one million microliters, and one billion is equal to one million thousands, a result with units bil/L is mathematically the same as k/UL, so these entries for units can also be standardized to k/UL.

Work to convert units where the corresponding result may need to be mathematically transformed has not been performed. Transformation may also involve changes to the normal range values for the test, however normal range values are not always present. Furthermore, while some transformations may be obvious, such as mg/ml into mg/L, others may not be so obvious, such as the potential difference between U/L and IU/ml. Preliminary analyses that examine the distribution of results of differing units has been conducted on a test by test basis so that sensible mappings to standard set of result units can be determined.

The revised data model for the clinical data elements is included in V3.0 of the MSCDM available on the Mini-Sentinel website.

4. Implementation

Implementation of the new lab results presented several challenges. While the troponin, CK and platelet tests were relatively straightforward to identify, pregnancy and influenza testing were more challenging due to the variety of test types and specimen sources, and multiple ways of specifying results. Influenza testing could include different subtypes, and be processed as an antigen, antibody or DNA/RNA test, with qualitative or quantitative results. Initial analyses also revealed that influenza testing was not performed routinely or uniformly across all participating Data Partner sites. We were only able to capture influenza testing that was performed by central laboratories that reported results to the Data Partners. We did not capture rapid influenza testing that was performed in the office setting. While there were over 300 different influenza test codes recognized in the Logical Observation Identifiers Names and Codes (LOINC) dictionary, most of the available results were represented by a few dozen test types. The focus was on results for Influenza A or Influenza B antigen tests only, since this represents acute infection. Influenza antibody tests were excluded because the value of assessing prior exposure was less certain.

Similarly, pregnancy testing could be reported as qualitative or quantitative tests from blood specimens or qualitative testing of the urine. Furthermore, there was the traditional pregnancy test which is the beta subunit of Human Choriogonadotropin and an unspecified, heterodimeric Human Choriogonadotropin which is often tested in suspected cases of abnormal pregnancy. At the present time we have elected to include qualitative and quantitative beta-HCG from serum and urine, and qualitative and quantitative HCG from the serum.

While it is anticipated that the MSDD will contain several million results for tests like Creatine Kinase (CK) and Troponin, one important caveat is that tests from HealthCore, Humana and Aetna, which represent a significant proportion of the universe of tests, contain only a portion of outpatient lab results. Given the severity of a potential myocardial infarction for which these tests may be markers, these tests are likely only ordered in the ambulatory setting when the suspicion is low. If the suspicion is high, the patient may skip the ambulatory setting entirely, or the patient will be sent to an acute care setting right from the ambulatory clinic before labs are drawn. Therefore, the troponin and CK tests available from HealthCore, Humana and Aetna may be skewed toward the normal, compared with results from KP where inpatient results are available. Comparison of result ranges from the different Data Partners may demonstrate the presence and degree of this bias.

5. Data Checking Approach

In Year Two, the focus of the clinical data elements data checking was to assess counts of lab tests overall and per year, number of patients having any lab test, and number of patients having each particular lab test. In Year Three, a set of checking algorithms was developed to ascertain changes in trajectory of laboratory results that may be a signal for an adverse drug event. While tests such as troponins are only ordered in the setting of a presentation concerning for myocardial infarction and are expected to be ordered only a few times in selected patients, other tests such as hemoglobin, ALT, Creatinine, and Hemoglobin A1c are ordered more routinely. A single abnormal test result may be discovered in the setting of a clinical suspicion of a problem; however a more convincing scenario for adverse drug events would be the presence of a track record of several normal results, followed by an inflection in the trajectory of results occurring after an exposure to a medication. At the data checking stage, it is infeasible to anchor every patient's series of test results for every possible drug exposure.

However, the number of patients with multiple test results overall, and within one year of an index date in each calendar year will be reported. In this way, the data can inform the FDA about the number of patients having enough tests where a change in trajectory of results may be noted.

Vital signs data, including blood pressure, weight, height, and tobacco status has been incorporated into the MSCDM from all six KP sites and three HMORN sites. Blood pressure and weight are routinely measured at adult visits and weight and height are routinely measured at pediatric visits. It is feasible to track changes in values over time as all values are included in the MSCDM, leaving the decision to include or exclude a value based on the needs of the specific analysis.

6. Potential Next Steps for Clinical Additions

There are multiple potential areas of future work related to the characterization of and additions to the clinical data elements included in the MSCDM. Some of these include:

1. Additional characterization of the data elements to gain greater understanding of patterns, frequency, usefulness, and logical inconsistencies within the existing data (e.g., implausible changes in height),
2. Feasibility for additional Data Partners to contribute clinical data
3. Feasibility of capturing test results from additional data sources such as inpatient facilities and specialty clinical centers,
4. Refining the laboratory and vital signs data tables' structures to enable more efficient use
5. Developing laboratory and vital signs summary tables to enable rapid querying,
6. Incorporating laboratory results and vital signs data into modular programs.

The clinical data included could be further explored based on more extensive data characterization and ongoing lessons learned to identify laboratory tests that are high and low "value added" (e.g., pregnancy screening lab tests, troponins in outpatient setting) within the context of medical product safety surveillance. Finally, a need exists to more thoroughly document and educate other Mini-Sentinel teams and workgroups about the MSCDM Clinical Data Elements availability, the status of these data tables, and the capabilities for their use.

7. Summary

Investigations can now consider incorporation of the MSCDM clinical data elements in Mini-Sentinel activities. It is important for Mini-Sentinel investigators to remain mindful of the caveats for use related to how the clinical data are collected, captured, standardized, and stored and the sub-populations that have clinical data available for analysis.

B. OTHER REVISIONS TO THE MSCDM

During Year Three, the following minor modifications were made to the MSCDM v2.0: 1) clarifications to MSCDM descriptions, and 2) addition of new variables, and 3) addition of Summary Tables to the MSCDM. All revisions are described below and included in the updated MSCDM available on the Mini-Sentinel website. MSOC developed a data refresh Standard Operating Procedure (SOP) that details the steps required to update data in the MSDD.

1. **MSCDM Tables: Text Revisions**

The Data Core team at MSOC continued to revise and clarify the MSCDM based on lessons learned and feedback from Data Partners and others. As Mini-Sentinel investigators used the model more frequently and as new collaborators and programmers began to work with it, more descriptions and definitions needed to be clarified or fixed. Minor typos were edited, and the structure of the document was slightly revised for ease of use. Definition of some of the fields and tables (e.g., meaning of uniqueness of a row in each table) as well as the examples provided were refined to improve understanding of the model.

2. **MSCDM Variables and Tables: Additions**

Even though no formal revision to the existing core tables of the MSCDM were implemented during Year Three, one Data Partner created an additional table unique to their site to enable exclusion of a subset of members who cannot due to contractual obligations contribute to activities that involve chart reviews. The creation of that additional table was implemented in collaboration with the Data Core team. The new "exclusion" table, which contains patient identifiers that should be excluded for certain activities based on the membership agreements with those members, was successfully used to exclude members in various Mini-Sentinel activities. This type of exclusion is not uncommon in the secondary use of observational data. The exclusions can be related to State laws or membership agreements that limit certain activities. Finally, the State Vaccine table was added to the MSCDM document for those sites involved in the Post-Licensure Rapid Immunization Safety Monitoring (PRISM) activities.

C. EXPANSION OF THE MSCDM

The MSCDM can be expanded by adding new variables or allowable values to existing tables, or by adding new tables to the MSCDM. The addition of the laboratory results and vital signs table is an example of adding new tables to the MSCDM that are linked by unique individual identifier. Several expansion activities have been pursued by the Mini-Sentinel PRISM team that focuses on the safety of vaccines. The PRISM team is in various phases of development of several addition tables to support vaccine safety efforts but that could be beneficial for other Mini-Sentinel activities. These additional tables are include an immunization table based on linkages to State immunization registries, a birth certificate table based on linkage to birth certificate registries, and a mother-baby linkage table to enable assessment of birth outcomes.

To continue consideration of additional data needed for surveillance, FDA tasked the Data Core with constructing and implementing a three-year plan for expansion of the MSCDM. The expansion workgroup is led by the Data Core co-Leads and includes MSOC Data Core staff and several FDA representatives.   The workgroup solicited specific priorities for exposures and health outcomes of interest to the Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER) and gathered information about exposure types of interest, timing of outcomes with respect to exposures of interest, and capture in the current MSCDM. The workgroup activities were coordinated with a separate activity supported by CBER focusing specifically on the ability to identify exposures to blood products. With FDA and the Protocol Core, the workgroup assessed the need for additional clinical and laboratory data for health outcomes of interest.

Several themes emerged. First, of the 260 exposures of interest identified by CDER and CBER, 153 (58.4%) are administered by injection or intravenous infusion and, of those, the majority (n=105, 68.6%)

are administered in both outpatient and inpatient settings. Second, with respect to inpatient exposures, health outcomes of interest are likely to occur during the same inpatient stay as some exposures (e.g., blood products), but are likely to occur post-discharge for others (e.g., medications, vaccinations). Third, although many of the laboratory data needs identified by the Protocol Core are being addressed by the Clinical Data Elements workgroup, current expansion efforts focus on outpatient labs whereas inpatient labs may be necessary for better ascertainment of some acute outcomes.

Next, the workgroup engaged Data Partners in structured discussions about (1) the capture of infused therapies in the current MSCDM and how the MSCDM might be expanded to enhance capture, (2) completeness of inpatient administrations of exposures of interest, and (3) completeness of medical utilization data for members with multiple sources of medical coverage. The final report summarizes the results of these data model expansion discussions and the workgroup's recommendations for expansion of the MSCDM.[11] The recommendations included these areas: 1) Access to Inpatient Data Streams; 2) Cause of Death Data; 3) Mother-infant Linkage; 4) State Vaccine Table; 5) Linkage to Registries and Other Networks; and 6) Refinements to Current MSCDM Tables. Details of the recommendations are available in the report.

IV. MINI-SENTINEL DISTRIBUTED DATABASE

A. DATA QUALITY ASSURANCE AND CHARACTERIZATION

1. Overview

All data transformed by the Data Partners into the MSCDM were checked through the use of standard programs/data characterization code developed by the Mini-Sentinel Operations Center and refined through feedback from the Data Partners. Data Partners ran the data characterization programs on their local implementation of the MSCDM after each data "refresh". Each data refresh requires the Data Partner to perform an Extract-Transform-Load (ETL) process to update their implementation of the MSDD.). The ETL process is described in detail in our Year 1 report (http://minisentinel.org/data_activities/details.aspx?ID=128). The data review process includes MSOC review of the data checking output, documentation of the findings from the data checking output, identification of data "issues" that require discussion or documentation, and agreement with the Data Partner on next steps. The next steps could include acceptance of the refresh, acceptance of the refresh with specification for corrections to be made during the next refresh, or rejection of the refresh which would require a revised ETL and compete review. The specific steps included in the refresh process are described in the Mini-Sentinel Data Quality Checking and Profiling SOP and include the following high-level steps:

1) Data Partner implementation of their local ETL process
2) Data Partner execution of data characterization code
3) Data Partner review of data characterization output, revision of ETL as necessary, re-run of data characterization code
4) MSOC review of data characterization output, within and across sites and within and across ETLs
5) MSOC data characterization report provided to Data Partner for review and comment
6) Review and discussion of data characterization report by MSOC and Data Partner, agree to any necessary changes and their timeline
7) Acceptance of the ETL

Once the ETL is accepted, the Data Partner executes the Summary Tables program and updates their Summary Tables to enable use by the Mini-Sentinel Query Tool. The Data Partner is required to run the "update dates" query on the Query Tool to inform the MSOC that the data are ready for querying.

2. Data Characterization Specifications

The Mini-Sentinel program relies on the comprehensiveness and quality of the data available in the MSDD. MSOC works closely with each Data Partner to assess the quality and completeness of their MSDD data and to identify any caveats for use. To ensure MSDD data meet quality expectations, MSOC developed a series of measures to check data quality and to characterize the breadth and depth of the data available for querying. The specifications and report address areas such as missing data, invalid values, invalid date ranges, and internal inconsistencies. Issues identified in the report are discussed with Data Partners and resolved on a case-by-case basis. The design and the scope of the data characterization programs take into account:

- The way Mini-Sentinel Data Partners access administrative and claims data and electronic health record information can vary, possibly leading to variation in data capture and completeness.

- It is vital that the tables created match the defined Mini-Sentinel requirements.

The data characterization programs are run after each data refresh. The data quality activities are organized into three levels of data characterization, based on the type of checks being performed. A description of the data characterization approach and the findings accompanies this report and can be found under the Data tab of the Mini-Sentinel website in a separate document titled "*Data Quality and Characterization Procedures and Findings.*"

a. Level 1 Data Characterization

The Level 1 assessments review completeness and content of each variable in each file to ensure that the required variables contain data and conform to the formats specified by the MSCDM data dictionary. For each MSCDM variable, data characterization verified that data types, variable lengths, and SAS formats are correct and reported values are within the specified range. For example, in the demographic table, the date of birth must be a SAS numeric data type, with a length of 4 bytes. Additionally, the date of birth must be in the range of January 1, 1885, through the date in which the demographic table was created. Categorical variables must include only the values specified in the data dictionary. **Table 2** illustrates several of the Level 1 data characterization items for the dispensing table.

Table 2. Level 1 Data Characterization: Example for the Dispensing Table

	Variable Name	Rule	Error Code
1	PatID	Must be character data type	DIS1.1.1
	PatID	Must be non-missing	DIS1.1.2
2	RxDate	Must be a SAS date value of numeric data type	DIS1.2.1
	RxDate	Must be of SAS length 4	DIS1.2.2
	RxDate	Must be non-missing	DIS1.2.3
3	NDC	Must be character data type	DIS1.3.1
	NDC	Must be exactly 11 characters in length	DIS1.3.2
	NDC	Must be non-missing	DIS1.3.3
	NDC	Must only contain digits from 0-9 (i.e., no space or other characters)	DIS1.3.4
4	RxSup	Must be a SAS date value of numeric data type	DIS1.4.1
	RxSup	Must be of SAS length 4	DIS1.4.2
	RxSup	Must be non-negative	DIS1.4.3
5	RxAmt	Must be a SAS date value of numeric data type	DIS1.5.1
	RxAmt	Must be of SAS length 4	DIS1.5.2
	RxAmt	Must be non-negative	DIS1.5.3

b. Level 2 Data Characterization

Level 2 characterizations assess the logical relationship and integrity of data values within a variable or between two or more variables within and between tables. For example, the unique person identifier can occur more than once in the enrollment table, as there can be more than one span of enrollment for an individual. However, in the demographic table, the person identifier should occur only once. Further, the person identifier in the enrollment table must have a corresponding value in the demographic table. This ensures that, for all patients for whom enrollment spans are created, corresponding demographic information exists. **Table 3** illustrates several of the Level 2 data characterization items for the enrollment table.

Table 3. Level 2 Data Characterization: Example for the Enrollment Table

	Variable Name	Rule	Error Code
1		The combination of PatID, Enr_Start, Enr_End, and DrugCov must occur only once in the table	ENR2.0.0
2	PatID	Must have a corresponding value in the Demographic table	ENR_DEM2.1.1
3	Enr_Start	Must be earlier than or equal to Enr_End	ENR2.2.1
	Enr_Start	In combination with PatID, MedCov, and DrugCov, must occur only once in the file	ENR2.2.3
4	Enr_End	In combination with PatID, MedCov, and DrugCov, must occur only once in the file (implemented in Year Two)	ENR2.3.4

After each data refresh, Level 1 and 2 data characterization reports are sent to MSOC for review. MSOC will inspect the Level 1 and Level 2 reports, identify and report data anomalies, and discuss next steps with the Data Partner. All anomalies are reported to the Data Partners to determine whether the issue can be fixed or is part of the underlying data. If necessary, a plan for remedying the anomalies is developed—this typically entails a correction in the subsequent data extract—or the anomaly is documented so it will not signal an alert in the next data checking process.

c.   Level 3 Data Characterization

In contrast to the Level 1 and Level 2 data checks, the Level 3 data assessments "profile" the data, focusing on characterizations that do not have an expected outcome or True/False finding. Rather, the expectation is for some level of consistency across partners and over time for some assessments and some level of inconsistency for other assessments. For example, trends in the number of outpatient dispensings per person or the rate of hospitalizations should follow similar patterns across Partners, and any obvious divergence from the general trend requires investigation. Periods of sharp increases or decreases are also unexpected. These characterizations generate counts and proportions and show the spread of values within each relevant field across Data Partners and time. This profiling characterizes specific data fields for each Data Partner and aggregates information for cross-institutional comparisons. The Level 3 data characterizations also evaluate trends to help identify data gaps and unusual patterns both within an ETL and across Data Partners' ETLs. Examples of trends within a single ETL include:

- Outpatient pharmacy dispensing per member per month
- Hospital admissions per member per month
- Total dispensing per month
- Total encounters by encounter type per month

Examples of trends across ETLs, include number of members and number of records—both of which are expected to always increase with each ETL and with the addition of new data. Other Level 3 data characterization topics include counts of procedures per encounter by encounter type and year and diagnoses per encounter by encounter type and year. This approach has been used successfully by the

HMO Research Network, the Vaccine Safety Datalink, and other distributed networks to identify issues within their distributed databases.

As an example, several Level 3 data characterizations for the dispensing table are:

- Overall table statistics
    - Number of records in the table
    - Number of unique PatIDs (includes number/percent with missing, if any)
- Distribution of dispensing date (RxDate)
    - Dispensings by month and year
- Average number of prescriptions per PatID
    - By year
- Distribution of days supplied (RxSup)
    - All years
    - Overall
- Distribution of dispensed amount (RxAmt)
    - All years
    - Overall

By examining the counts and proportions, both Data Partners and the Operations Center are able to ensure that the data are reasonable within Data Partners and consistent across Data Partners. For example, age in years is profiled in the following ranges: 0-1, 2-4, 5-9, 10-14, 15-18, 19-21, 22-44, 45-64, 65-74, 75+. If a Data Partner's Level 3 data showed an unusually large proportion of any one age range, this would indicate that there may be an issue with how the MSCDM was populated. Or, if the age proportions at one Data Partner are substantially different from the other Partners, it may indicate a difference in the underlying populations. The Level 3 data characterizations are designed to identify areas where variation within and across sites represents a potential concern to be further evaluated. Active participation from the Data Partners is essential to addressing unexplained variability. We note that this level of data check is not intended to find all data anomalies, but rather to assess metrics that can be readily checked and flagged for explanation. Detailed, topic-specific data checking is required for every Mini-Sentinel query as review of specific data areas or patient cohorts may uncover anomalies not identified in the initial data checking activities.

3. Reporting

Results of the data characterization activities are shared with the Data Partners. Two companion documents—the *Data Quality and Characterization Procedures and Findings Report* and the *Mini-Sentinel Distributed Database Year Three Summary Report*—provide details of the data checking and characterization activities and results across all Data Partners. These reports accompany this report and can be found on the Mini-Sentinel website (www.mini-sentinel.org/data_activities).

B. INCORPORATION OF NATIONAL DATA STANDARDS AND CONTROLLED TERMINOLOGIES

MSOC is committed to adoption and use of relevant national terminology standards related to electronic health care data. The two primary activities under this task are incorporation of standards into the MSCDM and engagement with standards bodies, as directed by FDA.

1. Incorporation of Standards into the MSCDM

Incorporation of national electronic health data standards into the MSCDM entails three key components: 1) identification of relevant standards based on the operational characteristics of the Mini-Sentinel distributed data system; 2) identification of the electronic health data standards used by the Mini-Sentinel Data Partners, and 3) incorporation of relevant and available standards into the MSCDM.

As a distributed health data network, the Mini-Sentinel approach requires all Data Partners to conform to a single data model that can accommodate longitudinal health data going back as far as the year 2000. The common data model enables a fully distributed analytic approach that allows a single analytic program to execute identically at each Data Partner site. The distributed analytic requirement also requires adoption of a transparent and easily-understood data model that all Data Partners can implement within their existing electronic data capture systems. Currently, the Mini-Sentinel Data Partners use a limited yet comprehensive set of controlled terminologies to capture medical encounter, pharmacy dispensing, demographic, laboratory results, and health plan enrollment information. The information in MSDD represents the values found in the source files and does not include complex clinical mappings between coding standards or terminologies.

To facilitate adoption and use of the MSCDM, the MSCDM was developed as a simplified version of data models used in similar distributed networks such as the HMO Research Network. As described in the Mini-Sentinel Year 1 Common Data Model report (http://www.mini-sentinel.org/data_activities/details.aspx?ID=128), the common data model was developed over several months of iterative discussion with the Mini-Sentinel Data Partners and informed by the Mini-Sentinel Common Data Model Guiding Principles (http://www.mini-sentinel.org/work_products/Data_Activities/Mini-Sentinel_CommonDataModel_GuidingPrinciples_v1.0.pdf). The current version of the MSCDM is available online (http://www.mini-sentinel.org/data_activities/details.aspx?ID=105). The MSCDM was designed to accommodate other coding terminologies such as ICD-10. The key data areas included in the MSCDM are listed below, with the national standards used within each data area.

Diagnoses. Diagnoses are captured using International Classification of Diseases, 9th Revision (ICD-9-CM)[iii] codes recorded during inpatient and outpatient medical encounters. Depending on the Data Partner, diagnoses are recorded on health insurance claims submitted for reimbursement and/or in electronic health record systems for Mini-Sentinel Partners that operate as integrated delivery systems. Each of our Data Partners uses this standard terminology. The data model allows capture of ICD-10, or any other controlled terminology.

Procedures. Medical procedures are captured using ICD-9 procedure codes and Healthcare Common Procedure Coding System (HCPCS)[iv] codes, including Current Procedural Terminology-4 (CPT-4)[v] codes, recorded during inpatient and outpatient medical encounters. Procedures captured using these terminologies include a wide range of medical interventions, ranging from well-child visits to

---

[iii] http://www.cdc.gov/nchs/icd/icd9cm.htm

[iv] http://www.cms.gov/Medicare/Coding/MedHCPCSGenInfo/index.html

[v] http://www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/cpt/about-cpt.page?

immunizations, drug infusions, and inpatient surgical procedures. In addition, both CVX (Health Level 7 Table 0292, Vaccine Administered) and MVX (Health Level 7 Table 0227, Manufacturers of Vaccines) codes describing vaccine administration and manufacture have been adopted for vaccine-specific work involving immunization registries. Each of our Data Partners uses ICD-9 procedure and HCPCS codes.

Outpatient Pharmacy Dispensings. Pharmacy dispensings are identified using National Drug Codes (NDCs) that are recorded by pharmacies at the point of distribution. Each of our Data Partners uses this standard pharmacy dispensing terminology.

Death and Cause of Death. The death and cause of death tables use ICD-9 and ICD-10[vi] diagnoses codes. These are the codes available through the source of the information, typically State death registries.

Laboratory Results. Our Data Partners use a mixture of LOINC and local codes to identify laboratory test result types such as Influenza A, Influenza B, creatinine, and pregnancy. The local LOINC and local codes are mapped to the Mini-Sentinel laboratory result test type nomenclature. To the extent possible, LOINC codes are used to identify laboratory result types. Laboratory test result units also must be standardized to a set of uniform unit types. Laboratory test results can be numeric or text. For example, '+', '++', 'POS', and 'positive' are all potential pregnancy result units found in the source data. To enable distributed querying those results units must be standardized. In addition, numeric results could be measured in different units such as per liter or per microliter, and those units could be represented in a variety of ways (e.g., 'k', 'K', and '10e3' refer to thousands and 'uL', 'UL' U L' 'mcl', and 'cumm' are variations of a microliter). The MSCDM uses a standard abbreviation of 'UL' for microliter to enable distributed querying.

Some commonly referenced controlled terminologies such as RxNorm and the Systematized Nomenclature of Medicine--Clinical Terms (SNOMED CT) are not currently included in the MSCDM. Although these and several other potential relevant controlled terminologies are increasingly being adopted by electronic health record systems and some health plans providers, the Mini-Sentinel Data Partners do not uniformly capture information using those terminologies. The MSOC will continue to work with FDA and the Data Partners to assess inclusion of these and other standards as possible.

2.   Engagement with National Standards Bodies

There are a wide range of health data standards initiatives supported by public and private partnerships in the US and abroad. These activities and the growing adoption of electronic health record systems have the potential to improve semantic and syntactic interoperability and expand the range of potential Data Partners for Mini-Sentinel. For instance, the Meaningful Use standards[vii] related to data capture and transmission promulgated by the Office of National Coordinator for Health Information Technology (ONC) have the potential to standardized data content and vocabularies, thereby enabling distributed querying of a broad range of medical practices and health facilities.

[vi] International Classification of Diseases, 10th Revision; http://www.cdc.gov/nchs/icd/icd10.htm

[vii] http://www.healthit.gov/policy-researchers-implementers/meaningful-use

Not all health data standards are relevant to Mini-Sentinel, especially within the context of the Mini-Sentinel Data Partners and the Mini-Sentinel distributed querying approach. All uses of Mini-Sentinel are "secondary uses" of electronic health data and are therefore not directly related to approaches and standards targeting point-of-care transmission of health information. So although initiatives such as health information exchanges have potential application to the MSCDM, all standards are assessed within the context of the needs of the Mini-Sentinel distributed data approach and the needs of the FDA within the system.

FDA has identified the ONC Standards & Interoperability (S&I) Framework[viii] as a key binding point for engagement related to Mini-Sentinel data standards, specifically the **ONC Query Health Initiative**. Several members of the MSOC staff, and associated vendors, are actively engaged with the S&I Framework activities, especially the S&I Frame Query Health Technical and Clinical Workgroups, and will remain engaged with those activities. In addition, MSOC and FDA are participating in a Query Health pilot project to investigate the potential for incorporating inpatient and ambulatory electronic health record data querying within the Mini-Sentinel framework. The pilot will focus on a widely-used standardized clinical data model – Informatics for Integrating Biology and the Bedside (i2b2) - and a newly-developed clinical querying approach called the Health Quality Measure Format (HQMF). The goals of the pilot are to 1) assess, adopt and implement the ONC Query Health meta-data standards for the Mini-Sentinel Query Envelop used by the Mini-Sentinel Distributed Query Tool, 2) beta-test an upgrade of the Mini-Sentinel Distributed Query Tool with PopMedNet Version 3.0 which is consistent with current Query Health standards for distributed querying, 3) incorporate the i2b2 HQMF query adapter into the PopMedNet architecture, and 4) work with Beth Israel Deaconess Medical Center (BIDMC) in Boston to pilot end-to-end querying using the i2b2 HQMF adapter with the existing BIDMC i2b2 installation. In addition to the Query Health pilot project, MSOC involvement has included face-to-face meetings with S&I Framework staff, webinars, and participation on several working groups. These activities will continue in Year Four.

C. LESSONS LEARNED

Increasing use of the MSDD brought with it a series of challenges and lessons. MSOC has successfully increased the scale, effectiveness, and timing of our data core activities. We successfully incorporated a new, large Data Partner to the network and managed over 75 data refreshes, including detailed data characterization and review of every refresh. Our data checking process was updated to improve efficiency and reporting to Data Partners, and to continue to identify and address data anomalies identified during the review process, and has worked effectively with Data Partners to manage the data refresh and review process. Selected specific lessons learned are listed below.

Revision and Examination of the ETL Process. During Year Three several Data Partner sites upgraded their data platforms, switched internal data warehouses, or began using alternate streams of data to populate the MSCDM. Three sites revised their ETL process based on a change in either their internal data systems or system platform in Year Three. Some of these revisions led to delays in data refreshes. Ongoing open communication between the Data Partners and MSOC was vital in managing these changes to minimize the impact of Mini-Sentinel activities. MSOC and Data Partners together examined

---

[viii] http://www.siframework.org/

the potential effect that of changes in the ETL process and/or data refresh timing. This enabled MSOC to inform FDA and other users about the planned changes so they could effectively plan their data request activities. For example, some projects chose to delay data requests until after a planned refresh. Effective communication between FDA, MSOC, the Data Partners, and the other Mini-Sentinel data requesters allowed for a smooth transition for these data changes.

In addition to revisions to the ETL process, the data characterization programs revealed several idiosyncrasies that required investigation. As a result, three Data Partners were asked to investigate these potential ETL anomalies by comparing their source data to the data in their transformed MSDD. These investigations led to both changes in the ETL process and confirmation of the validity of the process and documentation of the results of the investigation.

Local changes in data sources and data warehouse upgrades will continue to occur within our Data Partners. MSOC and Data Partner communication about the timing and impact of these changes will help MSOC and FDA manage the impact of these changes.

Adherence to MSCDM Specifications. In Year Three, several data model issues arose after Mini-Sentinel distributed programs failed to successfully execute. The issues were the result of a combination of programs that did not fully incorporate "defensive coding" approaches and sites that were not fully compliant with the MSCDM specifications. Issues arose in situation in which the format of a variable did not meet MSCDM specifications (e.g., a SAS date format of numeric 8 instead of numeric 4), the variable was not well populated, or the values within the variable included unrecognized values. MSOC is working with Data Partners on each new data refresh to ensure full compliance with the MSCDM, is enhancing verbal and written communication with programmers to improve understanding of the data, and improving distributed program testing and validation to identify potential issues before programs are distributed to Data Partners.

V.   MINI-SENTINEL ANALYTIC TOOLS

A. OVERVIEW OF MODULAR PROGRAMS

As part of the Data Core's Year Two activities, the MSOC developed seven modular programs to facilitate rapid response to common queries by each Data Partner. Each of these programs has several required input parameters (e.g., exposures or outcomes), and the output contains summary-level counts (e.g., number of members with an incident exposure to a drug, number of members with a specific diagnosis/condition, at-risk populations) stratified by various parameters (e.g., age group, sex, year). All programs and documentation is posted on the on the Mini-Sentinel website ([Data Activities](#)) when available.

[Modular Program 1](#) *(medication/procedure use):* Characterizes the use of specified products or groups of products (defined by National Drug Codes (NDC)) dispensed in the outpatient pharmacy setting or procedures/diagnoses recorded in any setting. **For example:** Prevalent and incident users of statins by age group, sex, and year; number of members who received an influenza vaccination by age group, sex, and year; number of members with who had gastric bypass surgery.

[Modular Program 2](#) *(medication/procedure use among those with a specific condition):* Characterizes the use of specified products or groups of products (defined by National Drug Codes (NDC)) dispensed in

the outpatient pharmacy setting or procedures/diagnoses recorded in any setting, among a cohort of individuals with observation of a specified condition defined by ICD-9-CM diagnosis codes. **For example:** Use of asthma medications among those with an asthma diagnosis by age group, sex, and year; use of anti-TNF agents among those with a psoriasis diagnosis.

Modular Program 3 (incident medication/procedure use and outcomes): Evaluates the rate of specified outcomes (defined by ICD-9-CM diagnosis codes, procedure codes, or medication dispensings) among those with incident exposure to medications, procedures or diagnoses, with or without a pre-existing condition defined by ICD-9-CM diagnosis codes or procedure codes (ICD-9-CM or HCPCS). **For example:** Rate of stroke during exposure to an antidiabetic medication among new users of the medication who also had a prior diabetes diagnosis.

Modular Program 4 (concomitant medication/procedure use): Characterizes concomitant use of specified products or groups of products (defined by National Drug Codes (NDC)) dispensed in the outpatient pharmacy setting or procedures/diagnoses recorded in any setting, among those with incident use of specified products or procedures/diagnoses with or without a pre-existing condition, defined by ICD-9-CM diagnosis codes or procedures codes (ICD-9-CM or HCPCS). **For example:** Characterization concomitant use of atypical antipsychotic drugs and selective serotonin reuptake inhibitors among those with a diagnosis of depression.

Modular Program 5 (*background rate of health outcomes of interest*): Provides prevalence and incidence rates of diagnoses, procedures, or outpatient medication dispensings among at-risk populations. **For example:** prevalence and incidence rates of type 2 diabetes stratified by age groups, sex, and year.

Modular Program 6 (*medication/procedure use following a diagnosis*): Provides rate of medication/procedure use among at-risk, diagnosed populations, as well as metrics on time to first medication/procedure use from diagnosis index date. Optional features include: ability to restrict to incident diagnosis and/or naïve-to-treatment (i.e., medication and/or procedure) patients, and ability to add pre-existing conditions. **For example:** rate of oral antidiabetic medication use following first diagnosis of diabetes; rate of hip replacement surgeries following a fall at home among female patients aged 65+ with osteoporosis.

Modular Program 7 (*most frequently used codes prior & post index event*): Characterization of the "Top XX" (user-defined) most frequently observed diagnosis, procedure, and drug codes during a user-defined period before and after an index date. Index event of interest can be defined using any type of code, and results are provided for both prevalent and incident patients of the index event code(s). Standard output provides "Top XX" rankings using both number of users and events, and rates for both prevalent and incident use of each most frequently used codes are provided. **For example:** Top 10 dispensings observed in the 30 days before and after a heart transplant.

1. **Modular Program Query Request Process**

Use of each modular program is detailed in Section VII.A (Modular Programs). FDA requesters can request the use of any modular program using a set of input forms containing detailed information on all required and optional input parameters. Input forms are exchanged between FDA and MSOC via a secure file transfer system. Prior to distributing modular program query

requests to the Data Partners, the MSOC staff ensures that specifications of the request will meet the requester's expectation and the program package with the request specifications is tested by MSOC. Data Partners are expected to execute and return results of the query within 5 business days. MSOC typically returns a report to the requester within 5-10 business days after receipt of complete results from all participating Data Partners. MSOC has developed a Query Request and Fulfillment SOP to guide response to modular program (and other) query requests. A diagram of the Query Request and Fulfillment SOP is provided in **Figure 2** below.

Figure 2. Query Request and Fulfillment Process
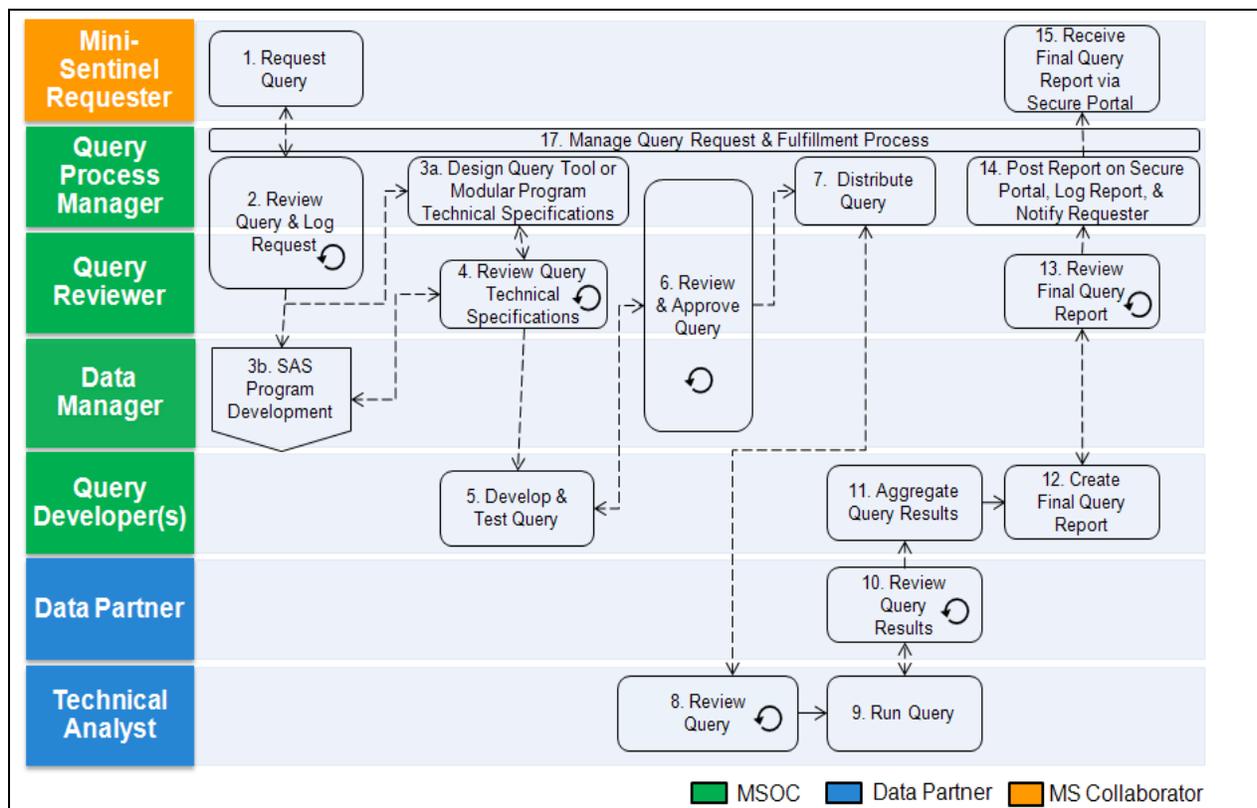
B. MODULAR PROGRAM REVISIONS

Year Three modular program revisions and updates included 1) enhancements of the existing programs with new features and capabilities, 2) modularizing existing code to add to the pool of modular programs, and 3) further development of input form capabilities.

1. Enhancement to Modular Programs

Two types of enhancements were undertaken in Year 3. First, MSOC implemented code revisions to all modular programs with new features to improve flexibility in defining cohorts of interest. Second, a pilot using Modular Program 3 (incident use and outcome) was implemented to add standalone modules that enhance result stratification. Summary descriptions of each enhancement are listed below.

a. New feature to modular programs

- **Population exclusion:** allows exclusion of members with selected diagnoses, procedure or medication exposures prior to a relevant index date to modular programs with a pre-existing condition module (i.e., modular programs 2, 3, 4 and 6). Example: exclude all members with a code for cancer in the 180 days before treatment initiation.

b. Standalone modules as pilot with Modular Program 3

- **High-dimensional propensity score:** a pilot project has been implemented to assess feasibility of use of various methods of multivariate adjustment for medication-outcome associations in a distributed database environment like Mini-Sentinel. The pilot focuses on incorporating a propensity score (PS) adjustment method, and will also include a semi-automated high-dimensional propensity score (hd-PS) approach. Input forms will allow end users of the PS standalone module to select a range of PS and hd-PS related parameters. Various drug-event scenarios will be tested and validated by selected Data Partners. A report describing the specifications of the PS module as well as results and description of testing and validation activities developed.
- **Utilization-based stratification:** in addition to the standard stratification capabilities (by Data Partner, age group, sex, and year/year-month), a new module allowing stratification by pre-index medical care utilization is being implemented. Options available include commonly used stratification metrics in observational assessments: number of inpatient visits, number of emergency department visits, number of outpatient visits, any care setting visits, number of dispensings for different drug products, number of different ICD-9-CM code categories, and number of procedures. Example: rate of AMI during anti-diabetic drug treatment among new users stratified by age group, sex, and number of hospitalizations in the 365 days before new user.
- **Charlson Comorbidity Index (CCI) stratification**: similar to utilization-based stratification, MSOC is adding a module to stratify cohort of interest into groups based on their medical complexity using the Deyo adaptation of the Charlson Comorbidity Index.

2. Conversion of Existing Code into Modular Programs (New Modular Program)

As part of CDER Task Order #7 (***Mini-Sentinel Operations Center Response to Potential Exposure/Outcome Associations***), the MSOC has revised the Year Two program for the **Drug Use Studies** project **(Comparison to Nationally Projected Databases)**. This program aims to assess uptake and persistence patterns for New Molecular Entities (NMEs). Work conducted for this revision has been recycled to convert the program into a modular program that will be available to FDA for routine data requests in the MSDD. The new modular program will go through the same quality assurance process and will have its own input form and documentation available.

3. Testing Phase

Before being used in production mode, all new modular programs go through a rigorous internal Quality Control process by developers: test cases are manually built (and documented) to stress-test the

modular program SAS code to ensure it only selects desired cohorts of interest and generates the expected output. Once a modular program has passed the Quality Control process, it is then shared with at least two Data Partners for additional testing and validation. Documentation is revised to ensure compliance with specification. Any feedback or suggested modifications from the Data Partners are handled by the MSOC; then the modular program is shared with all Data Partners. The Partners then 1) run it using a test scenario to confirm it can run within their local IT environment, 2) inspect output and log files to confirm they are valid and error-free, and 3) authorize the MSOC to routinely use it with FDA data requests. MSOC accepts a Modular program for use once all Data partners have approved it.

4. Input Form Software Development

During Year Three, MSOC has continued working on new input forms software: the Modular Program Query Interface (MPQI). Designed as a response to increased complexity of what modular programs can offer to FDA and as a safer mechanism to exchange the modular program query request parameters between FDA and MSOC, the MPQI prototype has gone through multiple cycles of internal testing to make it consistent with structure of new modular programs.

C. OVERVIEW OF SUMMARY TABLES

A second analytic tool used by the MSOC is the Mini-Sentinel Distributed Query Tool and Portal, described in greater detail in the next section. This software application allows the MSOC to quickly create and securely distribute queries to network Data Partners. Data Partners are then able to quickly review, execute, and securely return results of those queries within two business days to the requestor via a web-based Portal. Queries are run off of each Data Partner's "Summary Tables."

All Mini-Sentinel Data Partners create a set of 12 summary tables from their distributed database. They are created using distributed programs developed by MSOC programmers. Summary tables include prevalence counts of dispensings, procedures, diagnoses, and enrollment stratified by year, sex, age group, and where applicable, care setting. Specifically, the nine prevalence summary tables represent prevalence counts of diagnoses (3-, 4-, and 5-digit ICD-9-CM), procedures (3- and 4-digit ICD-9-CM and HCPCS), drug exposures (ingredient name and drug category), and enrollment. The code set used for the specifications for HCPCS, ICD-9-CM Diagnosis (3-, 4-, and 5-digit) and ICD-9-CM Procedure (3- and 4-digit) query types are provided by Ingenix, Inc. A description of each summary table is provided here:

ICD-9-CM Diagnosis Summary Table (3-Digit): Provides a count of unique members with a specific 3-digit diagnosis observed during the period and a count of events experienced within each stratum. The counts are stratified by setting of visit (inpatient, outpatient, emergency department, any), age group, sex, year, and 3-digit ICD-9-CM code.

ICD-9-CM Diagnosis Summary Table (4-Digit): Provides a count of unique members with a specific 4-digit diagnosis observed during the period and a count of events experienced within each stratum. The counts are stratified by setting of visit (inpatient, outpatient, emergency department, any), age group, sex, year, and 4-digit ICD-9-CM code.

ICD-9-CM Diagnosis Summary Table (5-Digit): Provides a count of unique members with a specific 5-digit diagnosis observed during the period and a count of events experienced within each stratum. The

counts are stratified by setting of visit (inpatient, outpatient, emergency department, any), age group, sex, year, and 5-digit ICD-9-CM code.

ICD-9-CM Procedure Summary Table (3-Digit): Provides a count of unique members with a specific 3-digit procedure observed during the period and a count of events experienced within each stratum. The counts are stratified by setting of visit (inpatient, outpatient, emergency department, any), age group, sex, year, and 3-digit ICD-9-CM code.

ICD-9-CM Procedure Summary Table (4-Digit): Provides a count of unique members with a specific 4-digit procedure observed during the period and a count of events experienced within each stratum. The counts are stratified by setting of visit (inpatient, outpatient, emergency department, any), age group, sex, year, and 4-digit ICD-9-CM code.

HCPCS Summary Table: Provides a count of unique members with a specific HCPCS code observed during the period and a count of events experienced within each stratum. The counts are stratified by setting of visit (inpatient, outpatient, emergency department, any), age group, sex, year, and HCPCS code.

Generic Name Summary Table: Provides a count of unique members who had a drug dispensing during the period, a count of dispensing received by all of these members, and total days supplied by strata. Counts are stratified by generic drug name, age group, sex, quarter-year, and year.

Drug Category Summary Table: Provides a count of unique members who had a drug dispensing during the period, a count of dispensing received by all of these members, and total days supplied by strata. Counts are stratified by drug category, age group, sex, quarter-year, and year.

Enrollment Summary Table: Provides a count of unique members and days covered stratified by age group, sex, year, drug coverage status and medical coverage status. The count of unique members or days covered can be used as denominators to calculate crude prevalence rates.

Summary tables and the Query Tool are not set up for ICD-10-CM diagnoses and procedures because these codes are not yet being used by Mini-Sentinel Data Partners. Summary tables are housed in a Microsoft Access Database held by each individual Data Partner, and are connected to the Mini-Sentinel Distributed Query Tool and Portal, which allows querying by the MSOC.

D. SUMMARY TABLE REVISIONS

MSOC implemented major revisions to simplify creation of the summary tables that are used for rapid querying via the Mini-Sentinel Query Tool. There is now a single distributed program with nested macros that improve efficiency through re-use of intermediate files for multiple purposes. MSOC also developed a new SAS program for the creation of summary tables for incident counts (events and members) for three different types of outcomes: (1) incident outcome by 3-digit ICD-9-CM diagnosis code; (2) incident exposure by generic name; and (3) incident exposure by drug category.

Both the revised prevalent and incident programs were reviewed and tested in accordance with the Mini-Sentinel SAS Program Development SOP. The SOP included MSOC testing, beta-testing by several Data Partners, and iteration until the programs are accepted as final. Testing also included verification that the output generated was compatible with the Query Tool software. Part of the process included

development of software specifications and changes needed to incorporate the new summary tables, implementation of the specifications by the software developers, user acceptance testing, and a new release of the software to enable rapid distributed querying of the newly-designed summary tables.

E. MINI-SENTINEL DISTRIBUTED QUERY TOOL

1. Overview of Query Tool

The FDA Mini-Sentinel Distributed Query Tool and Portal allows MSOC staff to create and securely distribute "queries" to Data Partners and enables Data Partners to review, execute, and securely return the results of those queries. The distributed architecture allows Data Partners to maintain control of their data and all its uses. The system allows different levels of query automation that can be set at the discretion of the Data Partners. The network is hosted in a private cloud environment in a Federal Information Security Management Act of 2002 (FISMA)[ix] compliant TIER III data center. The Mini-Sentinel Query Tool and Portal is based on the PopMedNet[TM] software platform. The implementation design and architecture are detailed in the Mini-Sentinel Distributed Query Tool: System Description and Technical Documentation.

The Mini-Sentinel Distributed Query Tool (see screenshot of the login screen in **Figure 3**.) currently allows rapid distributed querying of preprocessed summary tables. Using preprocessed summary tables speeds the querying process because it:

- Obviates the need to access person-level data, thereby avoiding local privacy and patient–confidential, data-release authorization procedures
- Allows use of a simple menu-driven querying tool interface
- Allows nontechnical Data Partner staff to execute and return results
- Avoids the need to specify, create, and validate new SAS programming codes to answer simple questions

The expected response time for these queries is 2 business days. The system includes three broad query types: prevalent queries, incident queries, and most frequent utilization queries. The capability to support the incident count and most frequent utilization queries were added during this contract year. The nine prevalence queries represent prevalence counts of diagnoses (3, 4, and 5 digit ICD-9-CM), procedures (3 and 4 digit ICD-9 and HCPCS), drug exposures (ingredient name and drug category), and enrollment. The incident queries represent diagnoses (3-digit ICD-9-CM) and drug exposures (ingredient name and drug category). For diagnoses and procedures, the system generates rates per 1000 enrollees, events per 1000 enrollees, and the number of events per person. For drug queries, the system generates users per 1000 enrollees, dispensings per 1000 enrollees, days supplied per dispensing, and dispensings per user. The tables have been revised to include the number of enrolled days per year by age group and sex to enable more precise calculation of prevalent rates. The most frequent utilization queries return the most frequently observed utilization (drug exposures, diagnoses, or procedures) defined by events or number of users by age group, sex, and year. The Mini-Sentinel Distributed Query Tool Investigator's Guide, a description of the Mini-Sentinel Summary Tables, and additional documentation

---

is available on the Mini-Sentinel website and has additional details on the summary tables and a description of how to create and distribute queries.

The Mini-Sentinel Distributed Query Tool architecture is consistent with the standards promulgated by the Standards and Interoperability (S&I) Framework supported by ONC. Mini-Sentinel staff are working actively with the S&I Framework Query Health team and actively participating in the ONC Query Health Initiative as a pilot program. Through this engagement we continue to communicate the lessons learned from implementation and operation of the Mini-Sentinel distributed querying system. These lessons include the need for detailed technical documentation and user training material, the need for security documentation and clearance by each Data Partner, and barriers faced related to installation of external software on local computers.

Figure 3. Distributed Query Tool Login Page
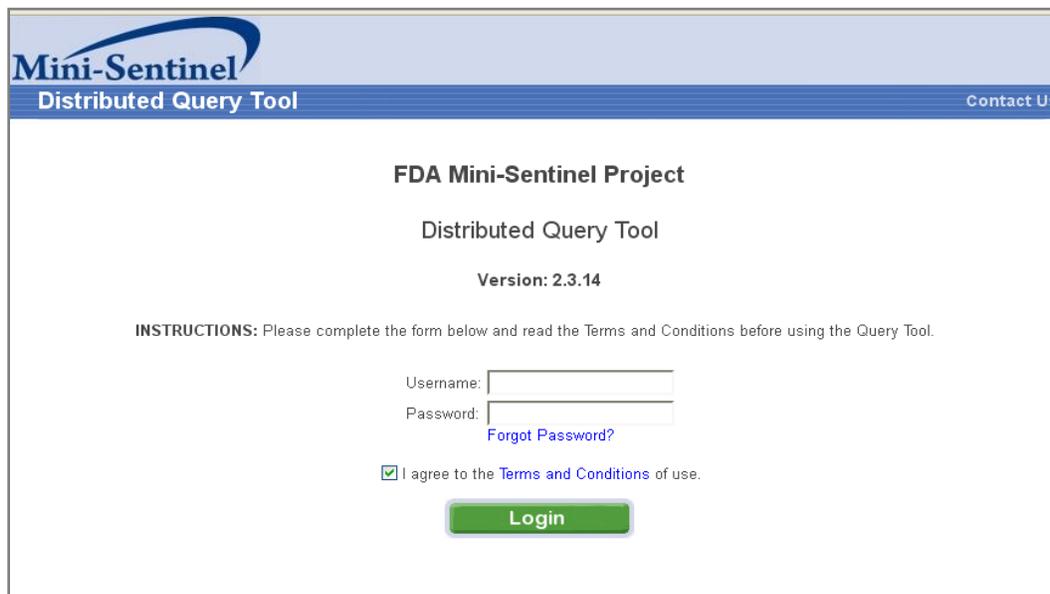
2.   Network Implementation

The distributed querying network was established in partnership with the MSOC, Mini-Sentinel information technology vendors, and the Data Partners. The implementation process involved establishment of a "staging" network that allowed testing of governance, security, and querying capabilities of the software platform, development of a series of user manuals, and implementation of a production site to allow secure distribution of queries. Use of the system has led to several revisions and enhancements. During Year Three, query tool enhancements focused on improving system architecture and functionality and better alignment with national querying standards as developed by ONC.

3.   User Setup and Testing

All Data Partners have login credentials to the Mini-Sentinel staging site (the beta-testing environment) to enable them to investigate system updates, set permissions, and otherwise evaluate the acceptability and usability of the software platform updates and system enhancements. MSOC provided Data Partners with testing scripts, role-based user manuals (e.g., *DataMart Administrator Manual*,

Investigator Manual, Overview and Technical Document) and detailed setup instructions. MSOC also provided one-on-one site support for system updates, site administration, and technical issues through telephone calls, webinars, and email. Technical questions about the software and security architecture were answered by the Mini-Sentinel IT vendor responsible for creating and operating the system.

Formal testing of system upgrades involved extensive testing by MSOC on the Mini-Sentinel staging network. The MSOC reviewed sample results to confirm proper system functionality. After MSOC approved all upgrades and enhancements, Data Partners were transitioned to the updated secure production server and Portal, and software upgrades were installed. Once transitioned to the production server, MSOC issued test queries for each query type to ensure the upgraded system was functional and operating as expected.

MSOC and the software developer provide ongoing support as new sites and users are added, questions arise, and enhancements are requested and developed. All software upgrades and revisions are accompanied by Release Notes to inform the Data Partners of the changes implemented. There are currently 17 unique Data Partners using the Mini-Sentinel Distributed Query Tool and Portal.

4. Platform Enhancements for Mini-Sentinel Query Tool Version 3.0

The Mini-Sentinel Query Tool software platform has undergone a series of enhancements and updates to improve the software platform to better conform to software development standards, enable modularization of enhancements, improve scalability and extensibility, make the system easier to maintain, and simplify system modifications. Enhancements also were made to better align our infrastructure with national querying standards described by the ONC S&I Framework Query Health Initiative. The specific enhancements to the technical architecture have been implemented to allow for a broader and more efficient use of the Query Tool software to improve:

- **Maintainability**. The platform upgrades make the Query Tool more efficient and sustainable to maintain as the system activity grows.
- **Enhancements**. The upgrades allow modularization of enhancements using a plug-in design.
- **Scalability**. The Query Tool can cultivate and support new networks, projects and users.
- **Extensibility**. The plug-in design allows for development of new features that can be added without impacting other parts of the system.

The specific platform re-architecture enhancements adopted by the Mini-Sentinel Query Tool are outlined below:

- **.NET 4 Framework**. The .NET 4 Framework is an application platform that is comprised of common language runtime and class library features, providing higher efficiency for overall code management and updates.
- **Entity Framework**. A type of object-relational mapping used as part of the new Query Tool platform. As part of this framework, an Entity Manager view, Data Access Layer and Structure Business Layer Class were developed.
- **Common Controls.** Information that is presented on the Query Tool portal, including functions that present the user with information that is presented in grids and lists.
- **Complex Controls.** Controls used to perform functions based on a specialized set of data, such as the controls for roles. Specifically, the Query Tool is able to understand all the information associated with the identified defined roles.

- **PopMedNet Library.** This is a library within the software platform that contains a set of helper and utility functions and services that are used across the entire Query Tool application.
- **Hub Background Service.** These are services that are outside the application and are essential to keep the application running.
- **User Interface (UI).** The entire UI for the Query Tool has been enhanced as part of the platform work. The UI contains all the graphical and textual information that the Query Tool presents to the user. The main function of the UI is to translate tasks and results into a format that the user can understand as they navigate through the system. Examples of new UI improvements include:
  - New Menu Layout - Menu names have changed to streamline navigation throughout the Query Tool. For example, menu tabs include pages for Home, Requests, Profile, Resources, Reports and Network.
  - New Code Selector Controls - A pop-up window was added for code selection and improved search functionality with codes and wildcards.
  - Newly Designed UI Buttons throughout the Query Tool
  - New Master Page Template and Home Page Layout
  - New Request (Query) Summary Page - New streamlined format for submitting query requests.
  - New Request (Query) Result Detail Page - Data Partners have a new enhanced view of their query request results.
  - New Request (Query) Status Page - Data Partners can view a list of their query request status on the Query Tool Portal.
  - New Response Page - Updated DataMart Client for Data Partners to view their workflow of outstanding or completed query requests. A Model Administration Page was added. Data Partners may have access to multiple types of data models and query types.
  - New Profile Page - Easily accessible user settings pages for Data Partners to administer their user profile.
  - New DataMart Administration Page - New Administration page for Data Partners to manage access and rights and download the latest version of the DataMart Client software.
- **Business Layer.** This layer incorporates and implements the business logic, located between the data access layer and the user interface. This layer coordinates the application, processes commands, makes logical decisions, and performs calculations.
  - Enhanced Access Controls
  - Event Manager
  - Authentication
  - Code Event Logger
  - Business Rules
  - Notification Manager
  - Request Scheduling Manger and Request State
  - Web Services
  - File Distribution or File Transfer feature
- **Model Adapter Construction.** The platform upgrade includes the newly designed concept for Model Adapters. This feature abstracts the request implementation from the system platform into a "Model Plugin". This new type of architecture separates the concerns of the network

platform from the details of the requests (i.e., queries) that travel through it. The result is a network that forms a tunnel through which requests and responses travel.

- o **Summary Table Model Adapter.** The Mini-Sentinel summary query functionality has been adapted to the model plugin software for the 3.0 platform.
- o **i2b2 Model Adapter.** Creation of an i2b2/PopMedNet plugin adapter that allows for use of the i2b2 Query Composer to construct queries that can be executed against i2b2 data sources within the Mini-Sentinel platform. The design and development of a PopMedNet/i2b2 Model adapter currently fulfills the standards for the FDA to participate in the ONC Pilot Program.

5. **Portal Enhancements**

- **Single Sign On.** To improve the management of queries and to allow for a single point of entry for multiple Mini-Sentinel applications, a single sign on landing page has been created. The single sign on for Mini-Sentinel web-based applications will allow a user to sign in through a secure landing page and gain access to all available applications due to a Lightweight Directory Access Protocol (LDAP) or Active Directory (AD) based single sign on service. The MSOC is currently pilot testing this feature with the Query Tool, Mini-Sentinel Public Website administration page, and the Mini-Sentinel Data Catalog. Next, MSOC will transition all Mini-Sentinel Data Partners with access to the Query Tool and the secure file transfer portal to using the single sign on feature for access to both applications.
- **Mini-Sentinel Data Catalog.** To better integrate Mini-Sentinel applications, the Data Catalog is now hosted within the FISMA compliant TIER III data center. MSOC has adapted the single sign on features to integrate the Data Catalog application into the Mini-Sentinel single sign-on landing page.
- **Summary Table Updates.** The software has been adapted to work with the new format of the prevalent and incident tables. These updates have been adapted to both the 2.3.14 version of the software currently in use and will be implemented as part of the new 3.0 platform roll out. The updates specifically include:
  - o Inclusion of the setting 'AN' which tracks how many individuals had a code for at least one of the available settings (e.g., IN, IP, AV, ED).
  - o Three new incidence tables which include 90-, 180- and 270-day look-back periods.
  - o A new age groups table, which is used to simplify the SQL script for each query. There is a new table in the sample database.
  - o Stratification of results by setting, and user ability to select more than one setting in one query.
  - o Inclusion of the new 'Days Covered' column within the prevalent and incident tables
  - o Most Frequent Utilization query type that allows capture of the most frequently observed utilization, listed by events or users and stratified by age group, sex, and year.

F. **MINI-SENTINEL DATA CATALOG**

As Mini-Sentinel's distributed querying capabilities grew during the first two years of the pilot program, MSOC developed the need for a system capable of organizing metadata regarding the execution of distributed data requests. During Year Two, MSOC designed the Mini-Sentinel Data Catalog (MSDC), a software system that tracks data flows within the Mini-Sentinel distributed data network. During Year

Three, MSOC implemented the use of the MSDC for the tracking of the results of all Mini-Sentinel modular program and workgroup requests, modular program beta-testing, and data quality checks.

1. Function of the Mini-Sentinel Data Catalog

The MSDC performs several tasks in the distributed data request process. The MSDC allows tracking of all queries, including the query metadata such as the sites involved, the project name, project number, query requester, query identifier, and dates for all activities.

When Data Partners return results via the Mini-Sentinel Secure Portal, emails listing Data Partner name, file name, file location, and upload date are generated automatically and sent to MSOC for notification and recording. In addition, the notification email is sent to the MSDC application that parses the email and automatically creates a record for each file uploaded by Data Partners. When MSOC accepts a results file uploaded by a Data Partner, the MSDC retains the record it automatically generated, allowing the MSOC to track all completed and current data requests.

2. Expansion of the Mini-Sentinel Data Catalog during Year Three

In Year Three, the regular use of the MSDC was helpful in directing its development as a powerful tool for MSOC. Since its implementation, the MSDC has been used to track more than sixty data requests. This volume of tracking demonstrates the utility of the MSDC as an organizational tool for the Operations Center.

During Year Three, several technical improvements were made to the MSDC. The reports function was expanded to allow users to select from a variety of metrics and filters in order to allow customization of the reports for different audiences. Additionally, the MSDC was moved from its hosting site at MSOC to a server at a vendor site. The vendor that now hosts MSOC also manages the development of the Mini-Sentinel Secure Portal, and this move has allowed for better integration of the MSDC and the Secure Portal and has facilitated integration of the MSDC with the single sign on tool. Finally, MSOC initiated work that will lead to the modification of the structure of MSDC input forms. This work will allow for the tracking of summary table requests in the MSDC, in addition to the modular program, workgroup, and data checking requests that are tracked currently.

3. Future Work

The MSDC has proven to be a valuable tool in tracking MS projects. However, further enhancements can be made to improve usability, tracking details, and reporting functionality:
- More flexibility in structure of input forms and results (for instance, if Data Partners submit two results for the same workplan due to an error in the original results, the MSOC would like to track both results)
- Enhanced privileges structure to allow FDA and Data Partners to see relevant information.

G. WEB-BASED LIBRARY

1. Overview of Web-Based Library

As part of Year Three activities, MSOC was tasked to build a web-based library of tools. The availability of infrastructure tools built and used by MSOC for FDA data queries via the project's public website will help current and future Mini-Sentinel investigators and developers with MSDD work.

All modular programs, data checking algorithms, and related tools are posted to the Mini-Sentinel website upon completion. These programs and the related documentation are updated as needed. The Mini-Sentinel tools also include SAS 'macros' that can be re-used by programmers. Programmers writing SAS programs running against the MSDD will be able to use these macros to speed their development work. The advantage of doing so is twofold. First, it reduces programming effort and cost to design, write, and test these commonly used procedures, thus speeding the development phase, reducing the potential for programming errors, and minimizing quality check time. Second, it implicitly allows Mini-Sentinel investigators to use algorithms (e.g., the Mini-Sentinel stockpiling algorithm) already validated by other investigators or FDA requesters and executed by all Data Partners, thus ensuring consistency across different projects.

As more Mini-Sentinel programmers use these macros, the MSOC will collect feedback on how to fine tune them and enhance their flexibility, and they will be revised appropriately. Future plans for the web-based library also include the posting of more SAS programs (modular or not) as MSOC finalizes them, creation of more standalone macros as implemented in various SAS programs, and the sharing of various template reports for modular program and summary table queries.

2. Description of Currently Available Tools

All SAS code posted to the Mini-Sentinel library includes a user guide and documentation. In addition each standalone macro comes with examples and test datasets to be used as test scenarios to speed development work. **Table 4** contains a list of all programs and macros posted as Year Three activities. More macros will be posted as they become available, and all new modular programs will be posted once finalized.

Table 4. Description of Year 3 Web-Based Library

Program Name	Short Description
Modular Program 1 v2.0	Year Two version of MP1: medication/procedure use
Modular Program 2 v2.0	Year Two version of MP2: medication/procedure use by condition
Modular Program 3 v2.0	Year Two version of MP3: incident medication/procedure use and outcomes
Modular Program 4 v2.0	Year Two version of MP4: concomitant medication/procedure use
Modular Program 5 v2.0	Year Two version of MP5: *background rate of health outcomes of interest*
Modular Program 6 v2.0	Year Two version of MP6: *medication/procedure use following a diagnosis*
Modular Program 7 v2.0	Year Two version of MP7: *most frequently used codes prior & post index event*
MS_AgeStrat v1.0	Age & time stratification tool
MS_CreateEpisodes v1.0	Creation of continuous treatment episodes with maximum allowable treatment gap
MS_Denominator v1.0	Reconciliation (i.e., bridging) of enrollment episodes with maximum allowable gap
MS_Envelope v1.0	Reclassification of Encounter Type value to IP for non-IP encounters identified during actual IP stays
MS_GetPharmacy v1.0	Extraction of outpatient pharmacy records with drug codes of interest
MS_GetMedical v1.0	Extraction of medical records with diagnosis and/or procedure codes of interest
MS_FreezeData v1.0	Creation of snapshot/frozen MSDD datasets for cohort of patients of interest
MS_Stockpiling v1.0	When an outpatient pharmacy dispensing is filled in early, make the next dispensing start at the end of the previous
MS_ConfirmElig v1.0	Confirm that medical and pharmacy records of interest or episodes within eligibility periods

H. ELECTRONIC SUPPORT FOR PUBLIC HEALTH (ESP) DATA MODEL AND SOFTWARE

One goal in Year Three of the Mini-Sentinel project was to enhance the MSOC's ability to create MSCDM-compliant test data for vital signs and laboratory test results. Building on the work during the Year Two, MSOC is using an open source electronic medical record public health surveillance platform known as Electronic Support for Public Health (ESP) as a for creation of test vital sign and laboratory results data in the MSCDM format. Several features of the tools and processes were selected for augmentation, documentation, and training.

1. Enhancing ESP's driver to Create Fake Lab Data

The ESP platform was updated to accommodate the revised laboratory data model and to incorporate the new test types added Year Three. To accommodate the creation of test data for these new labs tests, the "driver" table used by ESP to randomize the fake lab data generation was updated with new

lab test names. In addition, several formats for these new lab tests (names, codes, lows, highs, and units) were designed.

2. ETL Process and Tool and Enhancements of Mini-Sentinel Test Data

One of the challenges encountered in the Year Three work was that ESP and Mini-Sentinel do not have fully compatible schemas for lab data. The processes used to generate fake test lab data for Mini-Sentinel that were developed during Year Two resulted in null values for many of columns of lab data due to absences of relevant schema items in ESP's lab tables.

To overcome gaps in the schema overlap between ESP and Mini-Sentinel, the ESP driver data, used by the fake data generation, was expanded for use by the ETL tool. When a specific lab test is selected for fake data generation and has been mapped by the ETL tool, the ETL tool's batch transfer process expands the random data set by "looking up" values in the ESP driver table for each specific lab test and generating appropriate values for these fields.

3. Installation, Configuration, Documentation of the Toolsets and Process on a Windows Environment

The ESP product is designed for use on UNIX systems; however, for this project, the "fake data generation" process of ESP was ported for use on a Windows environment. The "make fakes" utility was enhanced to employ a user-friendly properties file for input parameters to the fake data generation (i.e. how big a record set).

The process was installed and configured end-to-end on a windows environment at the MSOC and a set of documented training materials was presented to allow MSOC staff to be able to independently create test data sets for Mini-Sentinel use.

I. LESSONS LEARNED

MSOC successfully implemented a series of updates and enhancements to the Mini-Sentinel analytic tools. The tools are activily used to respond to Mini-Sentinel queries (see Section VII), and through this use have identify opportunities to improve their use and efficiency. The procedures developed by MSOC to implement changes to the analytic tools proved valuable in helping to ensure reliable transitions to new and updates tools. A summary of lessons learned related to the key Mini-Sentinel analytic tools is below.

1. Modular Program

MSOC developed and implemented a more formal system for development and implementation of new and revised modular programs. The approach is detailed in the SAS Program Development SOP. The process includes distinct steps to ensure that the programming is well specified, tested, and beta-tested by Data Partners. Although this process has proved useful, modular program development is still complex and time-consuming. MSOC would benefit from a more streamlined and semi-automated testing platform that could be used for each release of modular programs, and improved documentation and systems to facilitate modular program requests.

2. Summary Tables and Distributed Query Tool Software

The Mini-Sentinel Distributed Query Tool is the most actively used tool within Mini-Sentinel and has proved very useful in quickly generating high-level information regarding exposures, diagnoses, procedures, and enrollment. To date, the Query Tool has been used issue over 100 summary table queries that generated information on several hundred drug exposures, diagnoses, and procedures.

The increasing importance of the tool has highlighted our need to tightly manage software upgrades to ensure that the tool is available for use. The Query Tool is a complex software application that now involves dedicated software management and support by MSOC.

3. Web-based Toolkit

The web-based tool kit has proved invaluable to Mini-Sentinel programmers and helps ensure consistency across Mini-Sentinel activities. Common procedures such as identification of a new user of a medical product or creation of a continuous enrollment period can be accomplished using the Mini-Sentinel tools, obviating the need for Mini-Sentinel program developers to implement their own unique approaches to these common tasks. Going forward, MSOC will continue to build the analytic toolkit with more tools and improved documentation to enable support to a wider group of programmers.

VI.   OTHER DATA CORE ACTIVITIES

A. MSOC COMMUNICATIONS

MSOC holds a weekly teleconference to maintain regular contact with and between the Data Partners. In addition to regularly scheduled meetings, MSOC is available by email, phone, and teleconference to deal with concerns and questions as they arise.

During Year Three, the MSOC continued to expand its work with various workgroups. MSOC helps ensure that workgroups utilize the MSDD effectively, efficiently, and properly. MSOC Data Core members are available to the workgroups during regular meetings or by email and phone as needed. In particular, MSOC reviews all workgroup plans to ensure that sensitive information is appropriately protected. MSOC also maintains a secure system used to communicate sensitive information with Mini-Sentinel Collaborators. This system has been designed to be compatible with all Mini-Sentinel Collaborators to continually facilitate data exchange.

Dissemination Activities, shown in **Table 5**, include presentations given by MSOC Data Core members during Year 3.

Table 5. Mini-Sentinel Data Core Meetings and Presentations (Year 3)

Date	MSOC Data Core Staff	Venue	Presentation Title
9/26/11	Jeff Brown	Brookings Institute	FDA Sentinel Initiative Strategic Review: Mini-Sentinel Querying Capabilities and Lessons Learned from Recent Assessments
10/11/11	Jeff Brown	3rd Annual Great Lakes cGMP & Regulatory Science Forum	FDA Mini-Sentinel and the Common Data Model
2/8/12	Jeff Brown	Mini-Sentinel Methods/Protocol Core Call	Mini-Sentinel Distributed Database and Rapid Querying Capabilities
4/30/12	Nicolas Beaulieu, Lisa Trebino, Jim Marshall, Ashley Wong, Tiffany Woodworth	2012 HMO Research Network Conference	Organizing and Tracking Multi-site Data Network Project Activities
4/30/12	Jeff Brown	2012 HMO Research Network Conference	Mini-Sentinel Modular Programs: Bringing Data Closer to Investigators
6/15/12	Nicolas Beaulieu, Jeff Brown	FDA Webinar	Mini-Sentinel Modular Programs and Summary Tables
6/25/12	Jeff Brown	Mini-Sentinel Data Partner Call	Mini-Sentinel Data Core Mid-Year Review
6/28/12	Jeff Brown	DIA Annual Meeting	Distributed Electronic Health Data Networks for Medical Product Safety Surveillance
7/19/12	Mark Weiner	FDA Webinar	Content and Capabilities of the Mini-Sentinel Clinical Additions (Laboratory Results and Vital Signs)
7/20/12	Lesley Curtis	FDA Center of Biologics Evaluation and Research	Blood Safety Continuous Active-Surveillance Network Feasibility Assessment

VII. MSDD QUERY REQUEST SUMMARY

A. MODULAR PROGRAMS

Modular programs (MP) were executed to fulfill 33 data requests by FDA in Year Three. CDER was responsible for 30 requests; CBER was responsible for three requests; and MSOC initiated one request (**Table 6**). MP1 was used in eight requests, MP2 in one request, MP3 in 26 requests, MP4 in one request, and MP5 in two requests (**Table 7**). MP6 and MP7 were not available until the end of Year Three.

The requests had varying levels of complexity, ranging from a straightforward MP1 request with one run to a complex request consisting of a combination of MP1 and MP3 with pre-existing conditions and

incidence input files. For example, the prasugrel request consisted of two executions of MP1 and four executions of MP3 (total of six executions) to assess overall use and use among those with certain pre-existing conditions. It used one drug and two pre-existing condition input files. In another example, the smoking cessation request required two separate runs of MP3 with six executions each (total of 12 executions). Both runs used several drug and outcome input files. Each scenario generates a unique set of output files that must be audited by the Data Partners and MSOC, aggregated and checked by MSOC, and put in a report format for FDA. The more complex the request the more complex it becomes to create the report. In addition, because the modular programs are so flexible, it is difficult to create a standard reporting template for re-use.

The modular programs are being revised to allow incorporation of multiple input parameter scenarios within the same run of a modular program. The output generated would be the same, but the number of input files and output datasets will be reduced. For example, the scenario above with 2 washout periods and two cohort definitions will require only one execution using the new code and still generate the same output. Though the number of FDA requests fulfilled for Year Three was 32, the total number of unique scenarios queried was 887.

Table 6. Number of Completed Modular Program Requests, Executions, and Reports by Requester in Year 3 (September 23, 2011 to September 22, 2012)

Center/ Requester	Number of Requests	Number of Scenarios	Number of Reports
CDER	30	879	31
CBER	3	28	3
MSOC	1	40	1
Total	**34**	**947**	**35**

Table 7. Number of Completed Modular Program Requests and Executions by Modular Program in Year 3 (September 23, 2011 to September 22)

Modular Program (MP)	Number of Requests	Number of Scenarios
MP1	8	292
MP2	1	52
MP3	26	550
MP4	1	7
MP5	2	6
TOTAL	**38***	**907**

*Number of Requests in Table 7 does not match Table 6 because 4 requests used multiple modular programs.

Data Partners have five business days to complete every request. However, MSOC occasionally distributed multiple requests concurrently but staggered the due dates to keep consistent with Data Partners' workload expectations. Of the 19 requests, 12 were completed on time by all Data Partners. Of the 5 remaining requests, the average number of days past the due date these requests were completed was 1.8 days. Overall, response time by Data Partners was similar to Year Two and well within expectations.

All reports were created in Microsoft Excel® and included both tables and figures along with an overview sheet describing the request specifications and contents. Most reports presented the number of users, dispensings, total days supplied, dispensings per user, days supplied per user, days supplied per dispensing, and events (for MP3) for either prevalent users, incident users, or both. Additionally, the reports showed percent contribution of each Data Partner to the total number of users, dispensing, days supplied, and events (for MP3). The revised modular programs will include denominators and person-time when appropriate.

The average time from receipt of all data from the Data Partners to report submission was 18.8 business days and the median time was 12 days. The increasing use of modular programs has given requesters more experience with the capabilities of the programs, and in turn generated more complex requests. Complex requests usually require additional consultation with FDA regarding specifications, more "scenarios" and more data received from the Partners, and more complicated reports. Additionally, some requests required investigation and revision of errors or unexpected data in the output at one or more of the 17 Data Partners, and prioritization of other requests and activities.

B. SUMMARY TABLES AND QUERY TOOL

A total of 120 summary table queries were performed to respond to 34 requests during Year Three (**Table 8**). These 120 distinct queries included over 210 different drug products, 5 drug classes, 289 diagnosis-setting combinations, 16 procedure-setting combinations, and 569 HCPCS-setting combinations, each stratified y age group, sex, and year. CDER was responsible for 14 requests and the Center for Devices and Radiological Health (CDRH) submitted three. The FDA leadership team submitted one request. MSOC initiated the remaining requests for the purposes of: investigating counts as background information for task order activities (8 requests); investigating counts as background information for PRISM analyses (4); investigating counts as background information for a Modular Program request (3); and obtaining updated enrollment numbers for each Data Partner (1).

Data Partners always responded to query requests before the assigned due dates. Data Partners were typically given two business days to complete each query. However, when MSOC distributed multiple query requests at the same time, MSOC would stagger the due dates by a couple of days. In addition, occasionally, MSOC would wait for a Data Partner to update its data before asking the Data Partner to respond to a particular request, especially if the request was for more recent data. In these cases, one or two Data Partners might be given a slight extension of the due date.

Twenty-three of the 34 requests involved sets of summary reports that were created by MSOC and submitted to the requester (Table 8). All reports were created in Microsoft Excel and included both pivot tables and figures along with an overview sheet describing the tables and figures presented in the report. Most requests involved more than one Excel file report because reports were grouped by type of query. For example, if a request involved three generic name queries and two HCPCS queries, two reports would be created—one for the generic name queries and one for the HCPCS queries. For generic name queries and drug class queries, reports displayed counts of users, prevalence rates (users per 1,000 enrollees), days supplied per user, dispensings per user, and days supplied per dispensing. For diagnosis and procedure queries, reports displayed counts of patients, prevalence rates (patients per 1,000 enrollees), and the number of events per patient.

Table 8. Number of Summary Table Query Requests Completed in Year 3 (September 23, 2011, to September 22, 2012), by Requester

Center/Requester	Number of Requests (Broad Categories)	Number of Queries	Number of Requests Involving Reports to FDA
CDER	14	32	14
CDRH	3	8	3
CBER	0	0	0
MSOC	16	72	5
FDA Leadership	1	8	1
TOTAL	**34**	**120**	**23**

Table 9 displays the number of queries completed during Year Three divided by requester and query type. Most queries were HCPCS queries (41), generic name queries (27), or four-digit ICD-9-CM diagnosis code queries (22). In addition, there were 12 five-digit ICD-9-CM diagnosis queries, seven enrollment queries, five three-digit ICD-9-CM diagnosis queries, three drug class queries and three four-digit procedure code queries. There were no requests for three-digit procedure code queries.

Table 9. Number of Summary Table Queries Completed in Year 3 (September 23, 2011, to September 22, 2012), by Requester and Query Type

Requester	Enrollment	Generic Name	Drug Class	3-Digit Diagnosis Code	4-Digit Diagnosis Code	5-Digit Diagnosis Code	3-Digit Procedure Code	4-Digit Procedure Code	HCPCS	TOTAL
CDER	---	18	---	1	2	2	---	---	9	32
CDRH	---	---	---	---	2	---	---	2	4	8
CBER	---	---	---	---	---	---	---	---	---	0
MSOC	7	8	2	4	18	10	---	1	22	72
FDA Leadership	---	1	1	---	---	---	---	---	6	8
TOTAL	7	27	3	5	22	12	0	3	41	120

C. EXPLORATORY REQUESTS

As of September 22, 2012, MSOC distributed three exploratory requests in Year Three, all of which were requested by the Clinical Data Elements workgroup. After reviewing output from the Year Two Laboratory Data Quality Assurance (QA) programs, the workgroup thought it was necessary to understand and explore laboratory result units before resuming Year Three work. The first exploratory request executed custom code which was developed to examine range of result values associated with each result unit for each lab. Results from this request were used to create reports of the distribution of lab results for each laboratory result unit for each Partner. These reports educated the workgroup on how to enhance the MSCDM Laboratory table.

Laboratory QA and Vital Signs QA programs, both enhanced in Year Three, were also distributed as exploratory requests. The data request is still active at the time of this report, and thus data from all participating sites has not been collected. These programs were designed to characterize laboratory and vital sign data. As data are collected and summarized, the workgroup will review and make revisions as necessary. The Laboratory QA and Vital Signs QA programs are expected to be integrated into the "main" Data QA programs at the end of Year Three.

D. AD HOC REQUESTS

Four requests requiring ad hoc programming were completed during Year Three. FDA first requested a rerun of the 65 and over data characterization program developed earlier in Year Two. This program aims at characterizing use of medical services and outpatient pharmacy dispensing for the 65 and over population of the MSDD by collecting demographic information and using the top 100 most common diagnosis and procedure codes and drug generic names and classes. The analysis stratified patients by sex and age group (65-74, 75+) and was conducted for four calendar years (2007-2010). The request was distributed to Data Partners on October 18, 2011 and the report was submitted to FDA on February 7, 2012.

Then FDA requested the execution of an ad hoc program to enable the analysis of stability (i.e., data changes and status) in Data Partners' Mini-Sentinel Common Data Model (MSCDM) databases across successive data refreshes. Stability of the data was assessed by comparing selected information (e.g., trends, levels) between data refreshes. The request was distributed to all Data Partners on February 21, 2012 and the report was sent to the FDA November 14, 2012.

Finally FDA requested execution of a revised MP3 to assess risk of severe acute liver injury (SALI) among members exposed to oral antifungal agents. Additional programming was needed to enhance the usual modular program outcome definition to allow combinations of diagnosis and procedure codes, and care settings. Exposed members were further flagged for various severity indicators and two different sets of exclusion criteria were applied. The results were stratified by sex, age group (0-18, 19-50, 51-64, and 65+), and calendar year. The SALI report required two runs of the same ad hoc program, with the second run including a revised definition of antifungal agents. The second and final request was distributed to Data Partners on March 19, 2012 and the report was submitted to FDA on March 30, 2012.

All ad hoc requests followed the Mini-Sentinel SAS Program Development SOP, requiring formal specification of the revisions from FDA, SAS program development and testing, MSOC testing, and Data Partner beta-testing before final programs were finalized.

E. REPORT POSTINGS TO MINI-SENTINEL WEBSITE

During Year Three, MSOC began posting to the Mini-Sentinel website the reports generated from summary table and modular program requests (reports completed during both Year Two and Year Three). All Data Partner-specific information was removed. The reports were sent to the requesting center for approval before posting. In Year Three, 31 reports were posted. All 31 reports appear in the "Assessments" tab on the website with 17 under the sub-tab "Diagnoses and Medical Procedures", 12 under the sub-tab "Exposures to Medical Products", and 2 under the sub-tab "Health Outcomes Among Individuals Exposed to Medical Products". The titles of the reports are shown below.

Under "Assessments: Diagnoses and Medical Procedures":

- Acute Myocardial Infarction Diagnoses
- Arthritis Diagnoses
- Aseptic Necrosis of Bone Jaw Diagnoses
- Asthma Diagnoses
- Cardiopulmonary Resuscitation (CPR) Not Otherwise Specified Procedures

- Heart/Lung Resuscitation (CPR), Injection Diphenhydramine Hydrochloride (HCL), Injection Adrenaline Epinephrine Procedures
- Hip Implant Procedures and Diagnoses
- Hyperlipidemia Diagnoses
- Milk Allergy Diagnoses
- Negative Pressure Wound Therapy (NPWT) Procedures
- Occurrence of selected HCPCS codes 1
- Occurrence of selected HCPCS codes 2
- Progressive Multifocal Leukoencephalopathy Diagnoses
- Regional Enteritis and Ulcerative Enterocolitis Diagnoses
- Serious Cutaneous Adverse Reaction (SCAR) Diagnoses
- Unspecified Allergic Reaction Diagnoses
- VEGF and Bone Resorption Inhibitor Procedures

Under "Assessments: Exposures to Medical Products":

- Bone Resorption Inhibitor Use (by Drug Class)
- Clozapine Use
- Dalfampridine use
- Infliximab use
- Injection infliximab procedures
- Lindane Use
- Occurrence of selected generic drugs 1
- Occurrence of selected generic drugs 2
- Occurrence of selected vaccinations 1
- Propylthiouracil and methimazole use
- VEGF Inhibitor Use (by Generic Drug Name)

Modular Program Reports Under "Assessments: Health Outcomes Among Individuals Exposed to Medical Products":

- Angiotensin II receptor blockers & celiac disease
- Smoking cessation drugs & cardiovascular outcomes

MSOC is working with FDA to post the remainder of reports that have been created for summary table requests and will continue to work with FDA to post reports as new requests are completed and new reports are created.

F. LESSONS LEARNED

MSOC successfully implemented a fully functioning rapid response querying for modular program, summary tables, and ad hoc request. Over 150 requests have been fulfilled, representing information for several hundred drug products, diagnoses, and procedures. The process is guided by the Query Fulfillment SOP. MSOC staff members are dedicated to reviewing and responding to rapid response requests (modular programs or summary tables), which has improved efficiency and standardization as use of the system has grown. Modular program and summary table reporting templates have been

developed and updated and based on feedback from FDA and others. Open and effective communication with rapid response requesters, from the initial request through the delivery of the report (and afterwards) is essential to effective use of the system. Proactively requesting feedback on the content and structure of the reports has led to improvements in the reporting format.

The increasing use and complexity of the modular program has highlighted the need to continue and improve effective communication between MSOC and the requester. Based on use and feedback, MSOC has updated the review process for throughout the modular program lifecycle, from initial request to testing to reporting. This updated process will minimize the opportunity for errors in both execution and logic, thereby leading to decreased response time.

VIII.   AUTHORSHIP

This report was prepared by members of the Mini-Sentinel Data Core and Data Partners.

Data Core Leads

Lesley H. Curtis, PhD, Duke Clinical Research Institute, Duke University School of Medicine, Durham, NC, USA

Mark G. Weiner, MD, Department of Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

MSOC Members

Elizabeth Balaconis, BA
Adrian Banerji
Nicolas U. Beaulieu, MA
Jeffery S. Brown, PhD
Jillian Lauer, BS
James Marshall, MPH
Megan Mazza, MPH
Richard Platt, MD, MSc
Lisa Trebino, MPH
Tiffany S. Woodworth, MPH

Data Partners

Aetna
HealthCore, Inc.
HMO Research Network
Humana
Kaiser Permanente Center for Effectiveness and Safety Research
Vanderbilt University

IX.   REFERENCES

1. 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. Available at: http://www.regulations.gov/#!documentDetail;D=FDA-2009-N-0192-0005. Accessed 3/16/11.

2. Maro JC, Platt R, Holmes JH, et al. Design of a National Distributed Health Data Network. *Ann Intern Med*. 2009; 151: 341-344.

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.

4. Brown JS, Holmes JH, Shah K, Hall K, Lazarus R, Platt R. Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. *Medical Care*. 2010; 48: S45-51.

5. Brown J, Holmes J, Maro J, et al. Design specifications for network prototype and cooperative to conduct population-based studies and safety surveillance. Effective Health Care Research Report No. 13. (Prepared by the DEcIDE Centers at the HMO Research Network Center for Education and Research on Therapeutics and the University of Pennsylvania Under Contract No. HHSA29020050033I T05.) Rockville, MD: Agency for Healthcare Research and Quality, July 2009. Available at: http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=150. Accessed 3/16/11.

6. Brown J, Holmes J, Syat B, et al. Proof-of-principle evaluation of a distributed research network. Effective Health Care Research Report No. 26. (Prepared by the DEcIDE Centers at the HMO Research Network and the University of Pennsylvania Under Contract No. HHSA29020050033I T05.) Rockville, MD: Agency for Healthcare Research and Quality, June 2010. Available at: http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayProduct&productID=464. Accessed 3/16/11.

7. Brown J, Syat B, Lane K, et al. Blueprint for a distributed research network to conduct population studies and safety surveillance. Effective Health Care Research Report No. 27. (Prepared by the DEcIDE Centers at the HMO Research Network and the University of Pennsylvania Under Contract No. HHSA29020050033I T05.) Rockville, MD: Agency for Healthcare Research and Quality, June 2010. Available at: http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayProduct&productID=465. Accessed 3/16/11.

8. Hornbrook MC, Hart G, Ellis JL, et al. Building a virtual cancer research organization. *J Natl Cancer Inst Monogr*. 2005:12-25.

9. Electronic Primary Care Research Network (ePCRN). Available at: http://www.epcrn.bham.ac.uk. Accessed 3/16/11.

10. Mini-Sentinel Coordinating Center Data Core Year Two Common Data Model (CDM) Report. Section III.A.2. Available at:

http://www.mini-sentinel.org/data_activities/distributed_db_and_data/details.aspx?ID=152 [11 July 2013]

11. Mini-Sentinel Common Data Model Expansion Plan. Available at: http://www.mini-sentinel.org/data_activities/distributed_db_and_data/details.aspx?ID=150 [11 July 2013]