SENTINEL DATA QUALITY ASSURANCE PRACTICES

COMPLIANCE WITH "GUIDANCE FOR INDUSTRY AND FDA STAFF: BEST PRACTICES FOR CONDUCTING AND REPORTING PHARMACOEPIDEMIOLOGIC SAFETY STUDIES USING ELECTRONIC HEALTHCARE DATA"

Prepared by: Sentinel Operations Center

February 27, 2017

Sentinel Data Quality Assurance Practices

Compliance With "Guidance for Industry and FDA Staff: Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data"

Table of Contents

I.   PURPOSE

The Food and Drug Administration (FDA) set forth its current recommendations for data quality assurance (QA) in the following document: "Guidance for Industry and FDA Staff: Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data" (Guidance), section IV.E "Best Practices – Data Sources: Quality Assurance (QA) and Quality Control (QC)," in May 2013.[1]

This Guidance describes best practices that particularly apply to observational studies designed to assess the risk associated with a drug exposure using electronic healthcare data. While the guidance specifically mentions that it does not address real-time active safety surveillance assessments such as Sentinel on page 3, many of its recommendations regarding data QA apply more broadly to include the practices and standards used by the Sentinel Coordinating Center's (SOC). This document describes the ways in which the SOC upholds the FDA's standards regarding data quality assurance.

II.   COMPLIANCE WITH THE BEST PRACTICES FOR DATA SOURCES

As part of the SOC, the Data Management and Quality Assurance (DMQA) Team addresses the following topics recommended by the FDA in section IV.E. The topics are in bolded font and the manner in which they are addressed by the DMQA Team is in italicized font.

- **The general procedures used by the data holders to ensure completeness, consistency, and accuracy of data collection and management**

  While the SOC has no access to the source data of its Data Partners (DPs), each DP transforms local source data into the Sentinel Common Data Model (SCDM) format to be included in the Sentinel Distributed Database (SDD). The purpose of the SOC data QA activities is to assess whether the SDD meets reasonable standards for data transformation consistency and quality, including reviewing data integrity across data tables as well as characterizing data trends and patterns.

- **The frequency and type of any data error corrections or changes in data adjudication policies implemented by the data holders during the relevant period of data collection**

  To evaluate data characteristics and quality, SOC developed distributed code to query the content of SCDM formatted tables. The distributed code generates aggregate output tables that help determine whether the data conform to SCDM specifications, maintain integrity across variables and across tables, and trend as expected over time. Execution of all sections of the data quality review and characterization program package generates up to 244 output files: 164 Core output tables, up to 59 Lab output tables (depending on laboratory tests present in the data), and 21 Vital Signs output tables.

---

[1] http://www.fda.gov/downloads/drugs/guidances/ucm243537.pdf, accessed on February 14, 2017.

Output tables are designed to evaluate one or more data checks, i.e., pre-defined data quality measures or characterizations. Approximately 1,200 data checks are evaluated during each DP data refresh. Each data check is designated a "level 1," "level 2," "level 3," or "level 4" data quality check depending on the complexity of a data characteristic/issue:

- o Level 1 data checks review the completeness and content of each variable in each table to ensure that the required variables contain data and conform to the formats specified by the SCDM specifications (e.g., data types, variable lengths, SAS formats, acceptable values, etc.).

- o Level 2 data checks assess the logical relationship and integrity of data values within a variable or between two or more variables within and between tables (e.g., variable ADMITTING_SOURCE in the Encounter table is populated only for inpatient and institutional encounters).

- o Level 3 data checks examine data distributions and trends over time, both within a Data Partner's database (by examining output by year and year/month) and across a Data Partner's databases (by comparing updated SCDM tables to previous versions of the tables). For example, a level 3 data check would ensure that there are no large, unexpected increases or decreases in records over time.

- o Level 4 data checks examine the occurrence and prevalence of nonsensical diagnoses and examine variations in care practices across Data Partners (e.g., the proportion of prostate cancer diagnoses among women). Level 4 checks are designed to provide more targeted data analyses and profiling of Data Partner data, and are not necessarily designed to detect and correct errors.

Once the DMQA team receives the output from the data quality review and characterization programs provided by each DP, the following steps are implemented at the SOC, as part of DMQA standard operating procedures, to achieve uniform performance of the QA processes across all DPs and timeframes:

1. A Data Quality Analyst does a primary review of the output to ensure that data quality acceptance criteria are met.

2. The Data Quality Analyst who performs the primary review prepares a Data Quality Findings Report (hereafter referred to as the Report).

3. Another Data Quality Analyst does a secondary review of the output and the Report to ensure that data quality acceptance criteria are met.

4. The Data Quality Analyst performing the secondary review annotates the Report with additional findings or corrections.

5. The Data Manager reviews the Report and the output, finalizes the Report, and transmits the Report to the Data Partner using the Sentinel Secure Portal or other approved secure mechanism.

If data issues are found in the Report, the DP investigates and provides a written response either explaining the results or proposing corrective action. All decisions and discussions are documented in the Report in order to develop a knowledge repository about each DP's data. The SOC and its DPs work closely together to resolve QA-related data issues in order to approve the data for use in the FDA's data requests.

- **A description of any peer-reviewed publications examining data quality and/or validity, including the relationships of the investigators with the data source(s)**

  The SOC data quality assurance is not geared toward any one study type or outcome of interest, thus the DMQA team does not maintain its own list of peer-reviewed publications.

  However, Sentinel as a whole is committed to publishing findings in journals and sharing information at relevant conferences. "Publications and Presentations" section of the Sentinel Initiative website provides summary information about Sentinel activities that have appeared in peer-reviewed journals or conference materials.[2] Additionally, "Health Outcome of Interest Validations and Literature Reviews" section of the Sentinel Initiative website lists literature reviews and validation studies of a number of safety outcomes.[3] Lastly, some published articles specifically focused on building and maintaining a framework and infrastructure for data quality assessments in distributed data networks.[4,5]

- **Any updates and changes in coding practices (e.g., ICD codes) across the study period that are relevant to the outcomes of interest**

  The SOC data quality assurance is not geared toward any one study type or outcome of interest. Any such updates and changes are documented by the project teams leading various evaluations using the SDD.

  In general, some of the principles that guide the development and maintenance of the SCDM[6] ensure flexibility designed to capture various drug, diagnosis and procedure code types, including the new and evolving ones:

  - Principle 2: The SCDM is able to incorporate new data types and data elements as needs indicate.

  - Principle 6: The SCDM leverages evolving healthcare coding standards.

  The DMQA team uses licensed databases of diagnosis, procedure and NDC codes to validate a list of codes included in each DP's data, incorporates the results of this medical code verification process into the summary QA review report and communicates the findings to the DPs.

- **Any changes in key data elements during the study time frame and their potential effect on the study**

  Regardless of outcomes of interest and any specific time frames, the DMQA team routinely monitors data trends and patterns of the key data elements (i.e., the SCDM variables needed to make sure that the FDA's data queries can be executed properly). If any data anomalies are

---

[2] https://www.sentinelinitiative.org/communications/publications, accessed on February 22, 2017.

[3] https://www.sentinelinitiative.org/sentinel/surveillance-tools/validations-lit-review, accessed on February 22, 2017.

[4] http://repository.edm-forum.org/cgi/viewcontent.cgi?article=1052&context=egems, accessed on February 27, 2017.

[5] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4306391/pdf/nihms-491345.pdf, accessed on February 27, 2017.

[6] https://www.sentinelinitiative.org/sites/default/files/data/DistributedDatabase/Mini-Sentinel_CommonDataModel_GuidingPrinciples_v1.0_0.pdf, accessed on February 27, 2017.

found, the SOC works together with the DP to investigate the issues and find suitable solutions in order for the DP's data to be included in the SDD.

- **A report on the extent of missing data over time (i.e., the percentage of data not available for a particular variable of interest) and a discussion on the procedures (e.g., exclusion, imputation) employed to handle this issue. Investigators should also address the implications of the extent of missing data on study findings and the missing data methods used**

The DMQA team routinely collects information about the missingness of all variables in the SCDM and communicates the findings to each DP. Any procedure to address missingness for a particular variable of interest is done by the project team conducting a study of any specific health outcome of interest using the SDD. All Sentinel studies include a study-specific data quality/fitness-for-use assessment of the fields to be included in that study.