ENGELBERG CENTER for Health Care Reform at BROOKINGS

September 5, 2012

Meeting Summary

Overcoming Statistical Challenges to the Reuse of Data within the Mini-Sentinel Distributed Database

With passage of the Food and Drug Administration Amendments Act of 2007, Congress mandated that the U.S. Food and Drug Administration (FDA) develop a system for postmarket risk identification and analysis using existing electronic health data. In response to this charge, FDA launched the Sentinel Initiative in 2008. Mini-Sentinel, a pilot project of the Sentinel Initiative, currently has 18 data partners and over 130 million covered lives within its Mini-Sentinel distributed database.[1] Mini-Sentinel offers FDA the ability to quantify the occurrence of many possible outcomes within a defined population for whom medical product exposure is known. As a result, a range of public health surveillance activities are possible, ranging from calculating crude incidence rates to conducting full epidemiologic hypothesis testing using data from a significant fraction of the United States (US) population. Notably, these activities can be conducted in much less time and at much lower cost than previously possible.

The development of the Mini-Sentinel pilot for medical product safety surveillance marks an important expansion of FDA's surveillance capabilities with the potential for delivering more rapid and actionable information to inform regulators, health professionals, and patients. Accomplishing the goals of active medical product surveillance will require consideration of possible statistical challenges that arise within Mini-Sentinel's distributed database approach to conduct rapid safety surveillance activities. Among these statistical challenges, the issue of data reuse will require the development of a framework for appropriate procedures and analysis.

In cooperation with FDA, the Engelberg Center for Health Care Reform at Brookings hosted an expert workshop, "Overcoming Statistical Challenges to the Reuse of Data within the Mini-Sentinel Distributed Database." This workshop was convened to provide an opportunity for experts from academia, industry, and relevant government agencies to discuss challenges related to the reuse of data in Mini-Sentinel analyses and to identify potential options to address the statistical concerns.

Understanding Mini-Sentinel Analyses

Mini-Sentinel collaborating partners have established a distributed database system of electronic health care data, which includes administrative claims and clinical data. The system allows partners to maintain physical and operational control over electronic data in their existing environments. Mini-Sentinel scientists have developed a set of analyses capable of interrogating the distributed database to evaluate potential safety signals.

In collaboration with the Mini-Sentinel Operations Center, data partners execute standardized programs developed by Mini-Sentinel investigators to perform a range of analyses. Data partners use the Mini-Sentinel Common Data Model (MSCDM), which is a set of data formats that standardizes administrative

[1]Mini-Sentinel Distributed Database "At A Glance." (December 2012). Retrieved March 8, 2013, from http://mini-sentinel.org/about_us/MSDD_At-a-Glance.aspx.

and clinical information across systems. Data partners execute analyses within their individual data sets and then return aggregate results to the Operations Center for pooled analyses. Two of the most common types of programs are modular programs and protocol-based assessments.

Modular programs run rapid assessments of potential safety signals that are conducted in near real time. Modular programs[2] are SAS programs designed to run against the MSCDM and are executed by each data partner behind their firewall.

Protocol-based assessments are formal and more detailed evaluations, which are capable of greater sophistication in analyzing product-outcome pairs. Using customized study designs and protocols, these assessments allow for the measurement and comparison of exposure groups and outcomes while controlling for potential confounding factors.[3] Conducting protocol-based assessments, however, requires significantly more time and resources than modular programs.

Modular programs, specifically Modular Program 3 (capable of calculating the frequency of select events during exposure to a medical product group of interest), can provide FDA scientists with information regarding the strength of suspected associations between product-outcome pairs.[4] Stratification by age, gender, and data partner is also possible with modular programs. When considered along with other information about a suspected association, these results can be helpful to regulators for various activities, including determining if a more formal assessment would be appropriate. If a more thorough analysis is indicated, protocol-based assessments can be performed to further reduce the systematic bias associated with confounding.

FDA and Mini-Sentinel scientists are currently working to develop a system with analytical components called modules that will allow for the semi-automated routine active surveillance of newly approved products. This expansion of surveillance capabilities will allow multiple analyses on a list of predetermined exposure and outcome characteristics.

Modular programs and protocol-based assessments are just one of the tools that FDA may utilize when conducting drug safety evaluations. Information gained from the evolving Sentinel System is considered along with all other data about the safety signal coming from a variety of other sources including the premarketing development program, spontaneous reports, and other postmarketing studies to support FDA regulatory decisions.

---

[2] "Data Activities." Retrieved March 8, 2013, from http://www.mini-sentinel.org/data_activities/.

[3] "Health Outcomes Among Individuals Exposed to Medical Products." Retrieved March 8, 2013, from http://www.mini-sentinel.org/assessments/medical_events/default.aspx.

[4] Modular Program 1 describes outpatient pharmacy medication use during a defined period. Output is stratified by calendar window, age group, and sex. The output includes counts of exposed members, dispensings, and days of supply of the drug or drug group of interest.
Modular Program 2 describes outpatient pharmacy medication dispensings and use among individuals who had a specified diagnosis or a group of diagnoses before the initial drug use. Any number of drugs or drug groups can be included. The program output includes counts of unique users, dispensings, and days of supply of the drug or drug group of interest among those members with the specified diagnosis of interest.

Statistical Challenges with Data Reuse

Statistical challenges may arise when modular programs and protocol-based assessments are used to consecutively investigate the same or similar safety signals within a dataset. This is because protocol-based assessments have the potential to "reuse" some or all of the same data from a previous rapid assessment with a modular program. Some experts have expressed concern that the reuse of data to further refine associations that arose from the modular programs may lead to incorrect results or otherwise inappropriate conclusions. It is important to note that any approach to overcome these statistical issues must account for FDA's public health responsibilities and be conscious of FDA's overall decision making processes.

In an effort to address this issue and to explore solutions for mitigation, FDA charged an expert panel formed from Mini-Sentinel's Safety Science Committee with identifying specific data reuse issues and developing an initial set of recommendations for data reuse procedures.

This expert workshop was convened to explore the recommendations from a variety of experts, including those of the Mini-Sentinel's Safety Science Committee. The workshop was organized to facilitate discussion of the following issues:

- Statistical concerns related to the reuse of data within Mini-Sentinel
- Proposed solutions for overcoming statistical challenges
- Practical considerations with implementation in Mini-Sentinel

The expert workshop participants discussed a series of potential solutions to mitigate potential data reuse errors and various options for moving forward with these statistical challenges.

Addressing the Potential for Data Reuse Errors

Participants discussed the potential for data reuse to increase type I and/or type II errors, but were divided on the extent to which this is an issue. Participants suggested various approaches for mitigating the errors. While no agreement has been reach on the best procedure for mitigating data reuse, several approaches were discussed as potential solutions moving forward. These approaches, which include procedures recommended by the expert panel derived from the Mini-Sentinel Safety Science Committee, are as follows.

Classifying Activity by Strength of Product-Outcome Associations: Using prior knowledge of suspected product-outcome associations, scientists might classify the strength of a hypothesis which in turn, would outline the acceptable reuse conditions.

With this approach, Mini-Sentinel activities are classified into three categories, based on the *a priori* strength of the knowledge of the suspected associations ("prior"):

- Activity #1: Signal generation, or analysis with no prior
- Activity #2: Signal refinement, or analysis with a weak or moderate prior
- Activity #3: Signal evaluation, or analysis with a strong prior

"Signal generation" refers to a collection of methods for identifying potential, non pre-specified associations between medical products and adverse outcomes. While not currently a Mini-Sentinel activity, it is anticipated that it will be in the future. "Signal refinement" refers to a process for evaluating the magnitude and clinical significance of a suspected association. "Signal evaluation"

consists of the implementation of a formal epidemiological analysis to more definitively establish or refute causality between exposure to the medical product and the adverse outcome.

For activities 1 and 2, participants proposed a split-sample approach with different datasets used for derivation (analysis on first sample) and replication (analysis on second sample) or refinement (analysis on second sample with additional confounding adjustment) of the signal. For activity 3, participants suggested the use of a full protocol-based assessment but noted that modular programs might be useful under certain circumstances. Participants also commented that under this framework, modular programs should be avoided when performing sample size calculations, when statistical power is too low, and when systematic error is too severe to provide actionable information.

Stakeholders noted, however, that determining a set of clear and reproducible definitions for each level of priors may prove very challenging to develop and apply consistently, as the prior knowledge of a product-outcome association is not objectively quantifiable. Participants raised concerns regarding certain circumstances within activity #3 where it is possible that the use of modular programs in conjunction with protocol-based assessments will be deemed necessary to protect the public. Specifically, participants expressed concern that the initial modular program results, once known, will impact the design of the statistical analysis plan and allocation of resources for full protocol-based assessments. More specifically, initial results with a positive signal are more likely to be followed with a full protocol-based assessment than results with no signal, creating an unbalanced approach to safety surveillance activities.

Intent of Analysis: Determining the purpose for which the analyses are executed may provide scientists with a framework for appropriate data reuse procedures. Participants suggested that this framework accounts for the FDA decision-making context and avoids issues associated with quantifying the strength of product-outcome associations.

Participants stated that by classifying analyses into two categories, signal development and hypothesis testing, scientists can determine the appropriate method for analysis. “Signal development” refers to an activity which seeks to generate or refine safety signals (i.e., signal generation and signal refinement). “Hypothesis testing” refers to an activity which evaluates safety signals (i.e., signal evaluation).

Workshop participants suggested that under this framework, hypothesis testing activities be completed through full protocol-based assessments. Signal development activities may use modular programs to characterize a potential signal. However, a pre-specified protocol and statistical analysis plan must be in place prior to using a modular program in order for a hypothesis testing activity to follow. Given the resources required to design pre-specified protocols and statistical analysis plans, some participants remarked that these procedures may reduce FDA’s ability to rapidly query safety signals, and the results of modular programs would inevitably influence the content of the protocol and statistical analysis plan afterwards.

Data Splitting: The use of data splitting in general, regardless of *a priori* strength of association, was discussed at length. Participants noted that there is an increased risk of obtaining false positive results when using very large databases, such as the data used in Mini-Sentinel (i.e., high power to detect a very small chance difference as statistically significant).

Two options for conducting data splitting were proposed. The first option involves dividing each data partner's dataset randomly into two parts to provide truly homogenous data for modular programs and protocol based assessments in both datasets. Some participants commented on the relative ease of implementation and suggested that this approach would be a valuable way to generate two independent samples. It was noted that if the outcome of interest is not too rare, the loss of power and ability to control for confounding may not be a concern. Others challenged that this approach offers no advantage to reducing systematic error and only serves to reduce random error.

The second option discussed was a non-random partitioning of the data partners, which would set aside more complex datasets (i.e., those with clinical data in addition to claims data) to be used in a second analysis to control for additional confounding. This approach was seen by some as an advantage over the random split approach because of its ability to control for systematic error. However, some participants noted that the heterogeneity between the datasets may conflict with this method. As the structure of care delivery systems and the methods used to generate data are different in each dataset, a non-random separation will produce two disparate sample populations. The heterogeneity between these samples might impact results (i.e., differences that stem from dissimilarities between the datasets, not from any underlying casual relationships in the data).

A number of participants did not endorse the partitioning of Mini-Sentinel data at all. Regardless of approach, many workshop participants cautioned that the loss of statistical power associated with split samples may affect the ability of Mini-Sentinel database to test associations when rare exposure and/or rare events are involved, particularly, with respect to newly approved medical products that have not reached a large enough uptake. Furthermore, participants questioned the relative value that data-splitting would bring to a surveillance program. It was argued that repeating the same assessment in two portions of one database and making a conclusion based on the combined results does not address the potential for increased probability of type I and II errors as compared to conducting one analysis in the entire database. The reason is that the combined results of two analyses cannot be considered independent replication and confirmatory versus that of one analysis on the sum of the data. As a result, many participants encouraged the use of all available data at once. It was also unclear what conclusions, if any, could be drawn from a situation when the split-sample approach produces differing or opposing results. In the event of disagreement in the results between two split samples, this approach would serve to increase the chance of a false negative (i.e., type II error). Consideration of the trade-off between false positives and false negatives should be made. Another concern for the split-sample approach was that if the two analyses were in agreement, this could give a false sense of accuracy and reliability of the results, which may not be appropriate if the original modular program analysis did not control for confounding and thus only random error was reduced.

Nominal p-value: Participants proposed adjusting the nominal p-value to a smaller value as a solution for data reuse and mitigating the potential for increased false positive results. This method helps account for the multiple testing between modular programs and protocol-based assessments when querying the entire distributed database. Participants commented that this method allows for Mini-Sentinel investigators to continue the current use of modular programs, which can be a helpful tool for evaluating safety signals. It was also suggested that increasing the study power for a protocol based assessment could reduce the potential for increased false negative results.

Participants cautioned, however, that this approach may conflict with the practical aspects of FDA decision-making process. While Mini-Sentinel is just one tool to inform FDA's decision-making, the

agency may have an interest to act when analyses return results that would otherwise be statistically significant (e.g., p-values of 0.01 or 0.05), but were deemed *a priori* to not be significant. Acting on such an association would return to the same statistical concerns related to data reuse.

Additional Data Reuse Considerations

Workshop participants emphasized the importance of transparency with Mini-Sentinel analyses and suggested that all analyses be made publicly available. Participants also suggested that the repeat of some surveillance activities in other data sources (e.g., CMS, DOD, non-US databases) may be an effective means of ensuring accuracy and providing an appropriate independent confirmation. However, it was noted that reconfirmation of results may not always be possible, due to existing data limitations and resources, and FDA's regulatory timeframe constraints.

The accumulation of data from each partner may also have a significant effect on data reuse. The data continue to evolve as data partners update or "refresh" their datasets on a quarterly or annual basis. Some participants suggested that the reuse of a data after a "data refresh" may lessen concerns for potential statistical issues, as the new data will change the composition of each dataset. Participants noted that further exploration into the effects of data refresh on modular program and protocol-based analyses might prove beneficial for moving forward.

Summary and Next Steps

This expert workshop evaluated issues surrounding the reuse of data within a large distributed network and helped to identify general characteristics of a framework for overcoming the potential challenges. Participants emphasized the need to implement a practical solution. In order to overcome the statistical issues, it was generally agreed that a decision framework should be used to guide procedures. Experts stressed the importance of maximizing the available resources, both in terms of data availability and FDA resources, before limiting analysis due to data reuse issues. Overcoming this and future statistical issues within the database will require solutions that are consistent with Mini-Sentinel capabilities and FDA's regulatory and public health mandate.