Expert Workshop: Statistical and Epidemiological Issues in Active Medical Product Surveillance

Development of the Sentinel System will augment the U.S. Food and Drug Administration's (FDA) postmarket safety assessment capabilities by enhancing its capacity to conduct active surveillance at the population level. In order to inform the structure and operation of the full Sentinel System, FDA contracted with Harvard Pilgrim Health Care Institute to develop a pilot system known as Mini-Sentinel. One of Mini-Sentinel's key achievements in its first year was the establishment of a distributed data system, consisting of a coordinating center and various data partners – health plans and integrated health systems – that maintain physical and operational control over their patient-level data, leveraging data partner expertise and protecting patient privacy.

Accomplishing the goals of active medical product surveillance through the Sentinel System will require further development of new and existing methods and tools for drawing valid inferences about associations between medical products and health outcomes of interest (HOI). The most immediate need for these methods and tools relates to the phase of active surveillance known as signal refinement, which begins once a potential association between a medical product and an adverse HOI has emerged from other information available to FDA. In cooperation with the FDA, the Engelberg Center for Health Care Reform at Brookings convened a group of experts from academia, the life sciences industry, and relevant government agencies to discuss a set of specific methodological issues:

- Signal refinement in the context of large sample sizes
- Analyzing horizontally partitioned data (i.e., multiple data partners collect the same types of data for different cohorts of patients) or vertically partitioned data (i.e., different data on the same patients are held by multiple data partners)
- Evaluating the performance of active surveillance methods

The goal of this workshop was to identify and discuss alternatives in each of these areas, with a particular focus on how solutions may be applied within the context of Mini-Sentinel.

Addressing the Potential for Bias in Large Databases

Because increasing sample size decreases random error, the bias that remains is systematic in nature. Experts suggested several approaches to decrease the potential for this systematic bias, including:

- *Reducing bias by design:* Although sample sizes in Mini-Sentinel appear large, experts stressed that sample size may be dramatically reduced in order to evaluate more homogeneous subgroups of interest (e.g., those exposed to an uncommonly used drug), such that the effective sample size is not extremely large. As part of the design phase, sample size is also reduced given the need to ensure overlap in the distribution of potential confounders (e.g., pre-treatment variables and time-variation) across arms of the evaluation. In addition to these factors, it is also important to understand specific attributes of each site (data partner), as variation in attributes such as treatment protocols (e.g., payer formularies, hospital and physician practice guidelines) and patient demographics across sites may contribute to bias. In an ideal situation, investigators would be able to survey clinicians at each site and design studies to account for differences in

patterns of exposure. However, this may not be logistically feasible, underscoring the importance of leveraging data partner knowledge of treatment determinants by involving them in the design of the analysis.

- *Reducing bias by analysis:* When applied prudently, techniques such as propensity scores and instrumental variable analysis may help to reduce bias. However, it is important to recognize that not every database is equally suitable for answering every signal refinement question. Despite the availability of advanced techniques, in some circumstances, it may be appropriate to conclude that a certain signal refinement analysis cannot be done with a particular dataset without yielding an estimate with unreasonably high uncertainty. Finally, while generalized frameworks for signal refinement are desirable to achieve more rapid signal refinement results, these frameworks must remain flexible enough to be informed by considerations including biological plausibility of the potential medical product-HOI association.

- *Use of sensitivity analysis:* Even with careful design and analysis, some sources of bias will likely still exist and interpretation of results requires understanding assumptions about the data and its residual bias. Sensitivity analysis is a useful tool to estimate the magnitude of uncontrolled confounding that could prompt reconsideration of the results.

Understanding Effects of Heterogeneity Across Sites

A distributed data system introduces additional heterogeneity, given that estimates are drawn from disparate sites. While such heterogeneity offers the opportunity to identify important subgroup effects, it also poses a potential threat to validity. One option is to pool patient-level data to obtain a single estimate; doing so may enable extensive subgroup analysis and understanding of important site-specific effects. However, given privacy and data sharing concerns, pooling individual-level data will generally not be feasible in a distributed system. An alternative option includes generating site-specific estimates. If there is sufficient cross-site similarity, it may be appropriate to combine these estimates, by way of meta-analysis or other methods discussed below. However, if substantial heterogeneity exists, combining estimates may lead to unacceptable levels of uncertainty in the results. Further discussion will be required to elucidate more precisely the thresholds and circumstances under which combining site-specific estimates is appropriate as Mini-Sentinel and Sentinel conduct signal refinement evaluations.

Approaches to Resolving Partitioned Data

Using a distributed database model reduces risks to privacy and data security, but introduces the challenge of data partitioning. Mini-Sentinel currently faces the problem of horizontal partitioning, and once Mini-Sentinel begins linking data sources (e.g., a device registry with a claims database), it will also encounter the problem of vertical partitioning.

Meta-analysis is a technique commonly used in epidemiological studies to address horizontal data partitioning. Meta-analytic techniques confer a high degree of privacy protection since sites generate and transmit only an aggregated point estimate and variance to the coordinating center.[1] However, this approach requires that each site has sufficient statistical capability and limits flexibility to alter the analytic plan after queries are distributed.

Privacy preserving data mining is a collection of techniques that can be used to analyze partitioned data, and feasibility of these methods has been demonstrated in a number of fields. Privacy preserving data mining experts provided several examples of how these techniques can be applied to active surveillance,

---

[1] Rassen JA, Solomon DH, Curtis JR, Herrinton L, Schneeweiss S. Privacy-maintaining propensity score-based pooling of multiple databases applied to a study of biologics, Med Care 2010;48(6):83-9.

including a non-interactive model where all data files are sent directly to FDA, more complex non-interactive models that randomize responses and demographics to confer additional privacy protection, and an interactive cryptographic model, in which individual sites send encrypted data to the coordinating center to "sum" results but the coordinating center cannot determine individual site contributions.

However, privacy preserving data mining involves tradeoffs among utility of the results, data security, and time/resources required. Application of these methods also requires detailed upfront specification of what the analysis aims to achieve, representing a tradeoff of some flexibility for increased security. Some experts suggested that FDA's safety surveillance – a public health endeavor – does not require such rigorous privacy protection assurance.

Propensity score-based pooling of data is an alternative method that generates a single value for each patient that encapsulates all important information about a patient's disease status. Once a patient has been assigned a value, it is then possible to create two or more cohorts of patients with matching characteristics. While this does not ensure privacy to the same degree as privacy preserving data mining, it does offer a simple approach to signal refinement that offers some degree of privacy protection. Privacy preserving data mining experts, though, expressed concerns about patient privacy with the use of this method, particularly for rare diseases or outcomes, where patient identities could potentially be revealed.

Both sets of techniques offered different strengths and caveats. Mini-Sentinel's approach to horizontal data partitioning will depend upon balancing utility, practicality, and cost considerations with privacy requirements, on a case-by-case basis.

Approaches to Comparing Signal Refinement Methods

With a variety of methods in signal refinement available, FDA and investigators need a way to compare specific techniques and increase confidence in results. The Observational Medical Outcomes Partnership (OMOP) has been working on quantitatively evaluating a set of methods based on the ability to detect known associations between drug exposures and HOIs ("true positives" and negative controls) in a range of databases. OMOP's framework for comparing method performance focuses on method sensitivity and the false positive rate. Their research reveals that there is no single "best method" for detecting all evaluated drug-HOI pairs across the range of databases.

Another approach, referred to as event-based performance (EBP), takes into account event-based specificity and sensitivity as an alternative for comparing methods. The EBP approach incorporates time to signaling (i.e., time from medical product exposure to HOI), which is advantageous because time to signaling is highly relevant for public health decision-making.

Method selection, which balances the need for sensitivity and specificity, has real-world consequences. While false negatives unnecessarily expose patients to medication risks, conducting follow-on assessments for false positives is costly and unsustainable. Some experts suggested that high sensitivity is more important for pre-specified outcomes, while specificity becomes more important for non-pre-specified outcomes. This can ultimately serve as the beginning of a framework to determine specificity and sensitivity tolerance for particular medical product-outcome situations, which could inform method selection.

Summary and Next Steps

This meeting explored selected epidemiological and statistical issues that Mini-Sentinel is likely to encounter. Experts offered suggestions to reduce bias inherent in large sample size evaluations, as well as ways to deal with horizontally partitioned data. Finding the optimal balance among privacy, utility, and

practicality will help guide the approaches used to resolve horizontal partitioning. Selecting methods for signal refinement has public health consequences, and while experts proposed ways to evaluate methods, more investigation will be needed in this area to build a framework for signal refinement method selection that considers the consequences of false positives and negatives in a given setting. Finally, understanding variation in attributes (e.g., patient population, treatment selection) across Mini-Sentinel sites can inform the design and interpretation of signal refinement analyses. Further investigation to understand variations in patient care and practice patterns across data partners may help increase the validity of and confidence in signal refinement results.