MINI-SENTINEL METHODS

TAXONOMY FOR MONITORING METHODS WITHIN A MEDICAL PRODUCT SAFETY SURVEILLANCE SYSTEM: YEAR TWO REPORT OF THE MINI-SENTINEL TAXONOMY PROJECT WORKGROUP

Prepared by: Joshua J Gagne, PharmD, ScD,[1] Jennifer C Nelson, PhD,[2] Bruce Fireman, MS,[3] John D Seeger, PharmD, DrPH,[1] Darren Toh, ScD,[4] Tobias Gerhard, PhD[5,6] Jeremy A Rassen, ScD,[1] Azadeh Shoaibi, MS, MHS,[7] Marsha Reichman, PhD,[7] Sebastian Schneeweiss, MD, ScD[1]

Author Affiliations: 1. Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 2. Biostatistics Unit, Group Health Research Institute and Department of Biostatistics, University of Washington, Seattle, WA 3. Division of Research, Kaiser Permanente, Oakland, CA 4. Harvard Pilgrim Healthcare Institute and Department of Population Medicine, Harvard Medical School, Boston, MA 5. Department of Pharmacy Practice and Administration, Ernest Mario School of Pharmacy, Rutgers University, Piscataway, NJ 6. Institute for Health, Health Care Policy and Aging Research, Rutgers University, New Brunswick, NJ 7. Office of Medical Policy, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD

January 31, 2012

Mini-Sentinel is a pilot project sponsored by the U.S. Food and Drug Administration (FDA) to inform and facilitate development of a fully operational active surveillance system, the Sentinel System, for monitoring the safety of FDA-regulated medical products. Mini-Sentinel is one piece of the Sentinel Initiative, a multi-faceted effort by the FDA to develop a national electronic system that will complement existing methods of safety surveillance. Mini-Sentinel Collaborators include Data and Academic Partners that provide access to health care data and ongoing scientific, technical, methodological, and organizational expertise. The Mini-Sentinel Coordinating Center is funded by the FDA through the Department of Health and Human Services (HHS) Contract number HHSF223200910006I.

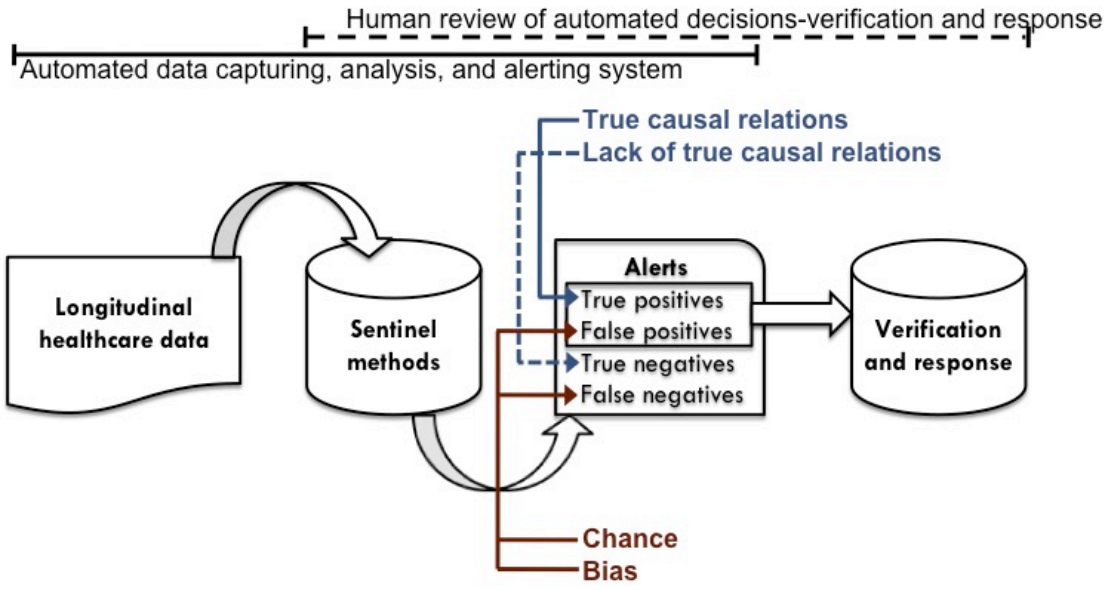
Mini-Sentinel Methods

Taxonomy For Monitoring Methods Within A Medical Product Safety Surveillance System: Year Two Report Of The Mini-Sentinel Taxonomy Project Workgroup

I. INTRODUCTION TO TAXONOMY

As with any public health surveillance system, a Sentinel System for medical product safety monitoring requires development, implementation, and orchestration of many tools, policies, procedures, and technical specifications.[1] At its core, an active medical product safety surveillance system is a series of methods for identifying and quantifying potentially causal relations among medical products and adverse health outcomes from longitudinal electronic healthcare data (**Figure 1**). As false positive and false negative alert generation can have untoward consequences for many stakeholders,[2] active monitoring systems require valid epidemiological and statistical methods to separate out relations that may be causal from those that may be due to chance and bias.

Figure 1. Schematic of major components of active medical product monitoring systems



Automated or semi-automated data analysis represents an important feature of large-scale public health surveillance systems.[3] However, all such systems require expert input from many stakeholders – including decision-makers, clinicians, methodologists, data experts, etc – to ensure that methods included in the system are valid and practical and are deployed in an intelligible, transparent, and appropriate manner for each monitoring activity. Various characteristics of specific medical product and

outcome pair combinations (i.e., scenarios) necessitate different methods, as each method, combination of methods, and constellation of parameters among a set of methods, requires certain assumptions that may be met in some scenarios, but can never be met in all scenarios. As such, it is unreasonable to expect that any single methodological approach to monitoring is suitable for all scenarios.

The Mini-Sentinel Taxonomy Workgroup sought to characterize analytic methods suitable for signal refinement, which FDA defines as an epidemiological process for evaluating the magnitude and clinical significance of a suspected association, and to provide clarity and practical advice for choosing the most appropriate signal refinement methodology for various medical product safety scenarios. Preemptive thinking about appropriate methodologies can enable collaborative, standardized, transparent, intelligible, consensus-designed, timely, and cost-efficient, decision-making to facilitate protocol development for one-off or sequential monitoring activities and can help outline the methodological needs of a system for routine medical product surveillance.

A. PROJECT PURPOSE AND OBJECTIVES

The overall goal of the Mini-Sentinel Taxonomy Workgroup is to categorize potential medical product safety scenarios that could be monitored within Sentinel according to select key characteristics, to map these categories, to the extent possible, to appropriate design and analytic methods suitable for active safety monitoring using electronic healthcare data, and to provide clarity and practical advice for choosing the most appropriate signal refinement methodology for the Mini-Sentinel program in order to support efficiency and transparency in decision-making. The specific aims of the project are: (1) to categorize potential medical product safety questions that might be evaluated within Mini-Sentinel according to characteristics that influence the choice of design and analytic approach; (2) to identify relevant methods for active safety surveillance; and (3) to map product-event pair characteristics to appropriate epidemiologic and statistical approaches for active safety surveillance using electronic healthcare data. The proposed mapping of methods is not intended to be prescriptive, but rather offer overarching methods guidance to increase efficiency and systems for methods selection. Moreover, the methods options are not exhaustive, but rather represent many of the common choices that we expect to be useful in Mini-Sentinel. Finally, the Taxonomy Workgroup seeks to identify, understand, and provide guidance on other issues that stakeholders might encounter when developing signal refinement

protocols or when developing methodological processes for expedited monitoring activities, and to identify gaps in methodology that could be filled by future Workgroups.

B. YEAR 1 ACCOMPLISHMENTS

The Taxonomy Workgroup was established during Year 1 of the Mini-Sentinel program and focused on identifying the most typical possible types of scenarios that may be subject to monitoring within the Sentinel System. Initially, key characteristics of exposures, health outcomes of interest (HOI), and the relations between them were identified.[4] The Workgroup then distilled the list down to scenarios defined by combinations of characteristics that influence monitoring design choice. Workgroup members then sought to identify the methodological design options for studying these scenarios and mapped a preferred design (or designs) to each scenario type. The key considerations for the design decisions were: (1) strength of within- and between-person confounding; (2) circumstances that may predispose to misclassification of exposure or misclassification of the timing of the HOI; and (3) whether the exposure of interest is transient or sustained.

The Year 1 Workgroup recommended that when the basic assumptions of self-controlled designs are fulfilled (i.e., transient exposure, lack of within-person, time-varying confounding, and abrupt HOI), self-controlled designs are to be preferred because of their inherent ability to avoid confounding by time-invariant confounding without having to measure those confounding factors. As scenarios deviated from those in which these assumptions were tenable, cohort-type approaches are generally preferred. When either self-controlled or cohort approaches are recommended (or when one is preferred but the other is listed as a possibility), several additional considerations are recommended, including whether absolute measures of risk (e.g., risk difference) can be estimated, and the availability of a reasonable active comparator.[4]

C. YEAR 2 AIMS

The main objective of the Year 2 Taxonomy Workgroup was to expand the decision table beyond design choices to include analytic approaches that are most suitable for each scenario. Specifically, the Workgroup aimed to: (1) identify current analytic methods most readily applicable to signal refinement activities in a distributed network of electronic healthcare databases; (2) map monitoring scenarios (as

defined by combinations of characteristics) to analytic options; (3) address additional specific methodological considerations likely to arise in routine active monitoring activities; (4) identify methodological gaps; and (5) develop a glossary of terms to help harmonize the language of active monitoring. To accomplish these aims, we convened a Workgroup comprising Mini-Sentinel investigators who are experts in epidemiological and statistical methods and who have considerable experience in understanding the challenges of active surveillance through involvement in other Mini-Sentinel Workgroups and related activities. Workgroup members included investigators involved in key Mini-Sentinel active surveillance activities for new molecular entities, including the diabetes drug saxagliptin[5] and the anticoagulant dabigatran.

II. SCENARIO CHARACTERISTICS

A key activity in the first year of the Taxonomy project was codifying scenario characteristics that might affect design and analytic methods decisions in an active surveillance system. **Table 1** summarizes these characteristics – which can be classified as being related to the exposure, related to the outcome, or characteristic of the potential link between exposure and outcome – with some modification to the Table in the Year 1 report.

Table 1. Scenario characteristics inherent to the specific exposure-outcome pair (i.e., scenario) that might affect design and analytic choice

Exposure characteristics			Characteristics of the (potential) exposure-HOI link				HOI characteristics	
Background frequency of use in population	Utilization trend in population	Use pattern	Onset of exposure risk window	Duration of exposure risk window	Strength of confounding		Background frequency	Expected degree of onset misclassification
					Between person	Within person		
More frequent	Uniform	Short-term (including intermittent)	Immediate	Short	Negligible	Negligible	Infrequent	Negligible (e.g., HOI is mortality captured by vital statistics)
Less frequent	Changing (increasing, decreasing, cyclical)	Long-term	Short	Long	Needs to be addressed	Needs to be addressed	Rare	Pertinent (e.g., cancer)

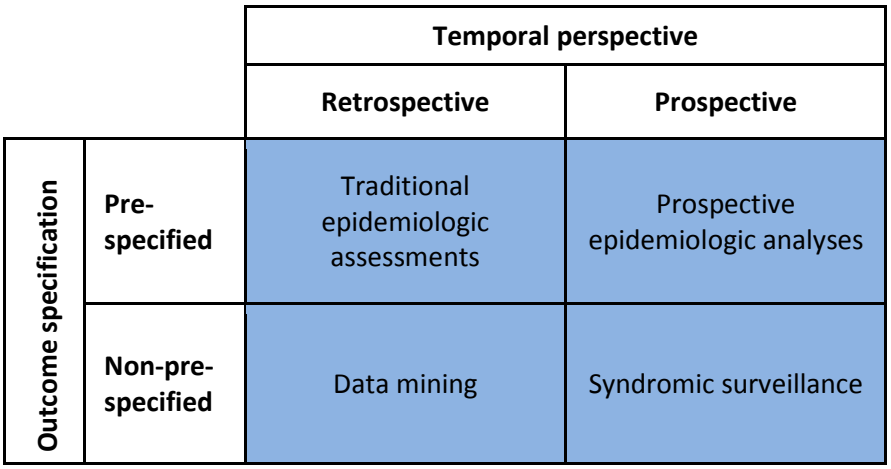
In addition to characteristics inherent to the exposure and outcome of interest, we identified three factors that may be characteristic of a monitoring scenario and influence methods selection, but which are specified by stakeholders in advance. These include the effect measure of interest, the number of comparison groups, and the comparison exposure (**Table 2**).

Table 2. Scenario characteristics determined by stakeholder/investigator that might affect design and analytic choice		
Effect measure of interest	**Number of comparison groups**	**Comparison exposure**
Difference measure	One	Active comparator
Relative measure	Multiple	Truly unexposed

III. KEY DECISION POINTS

Within the active monitoring framework, four types of surveillance activities are possible, as defined by combinations of temporal perspectives and outcome specification (**Figure 2**). Data mining activities, also referred to as "signal generation,"[1] are typically hypothesis-free analyses of a large number of potential adverse events (both suspected and unanticipated) among a large number of medical products. Data mining can be conducted retrospectively on static databases, or data mining can involve prospective identification of non-pre-specified outcomes following exposure to medical products, which more closely resembles syndromic surveillance.[6] A monitoring activity can also focus on pre-specified exposure-outcome pairs with suspected associations, which is also referred to as "signal refinement." Signal refinement can be a one-time analysis of pre-specified pairs, which resembles an ordinary retrospective epidemiologic assessment but is conducted in an expedited fashion,[7] or it can involve sequential monitoring where multiple analyses of accumulating epidemiologic data are conducted prospectively as new subjects become exposed over time.[8] This report focuses on monitoring activities for pre-specified outcomes and discusses aspects of the temporal perspective in the context of deciding between a one-time retrospective assessment versus prospective sequential monitoring.

Figure 2. Types of surveillance activities within an active medical product monitoring system



To develop the Taxonomy framework, we identified the decision points that investigators go through when developing a protocol for an assessment with a pre-specified outcome(s) and identified the options that they might consider at each decision point, keeping in mind the objective of using them for public health surveillance rather than for research, and within the context of both retrospective and prospective signal refinement activities executed across a distributed data network. The decision points and their options are not intended to represent an exhaustive list but rather the key decisions that depend on one or more scenario characteristics and common methods options at each decision node (**Table 3**).

Table 3. Key methodological decision points

Contrast	Methods to address exposure time trend	Methods to address baseline confounding		Estimation
		Confounder summarization	Incorporation in estimation	
Between-person	Case-time-control	No summarization	Stratification	No outcome model
Within-person	Bi-directional self-controlled case series	Propensity score	Matching	GLM (logistic, Poisson)
	Case-case-time-control	Disease risk score	As independent variable in model	Survival (Cox)
			Weighting	

Below, we describe each of the decision points, their options, and considerations for making choices among the options.

A. CONTRAST

Signal refinement designs can be conceptualized as different approaches to sampling person-time from an underlying population of interest. A fundamental design choice is whether it is of interest to compare exposure within individuals over time or between individuals.[9] Selection of the most appropriate contrast (i.e., within-person or between-person) is covered in considerable detail in the Year 1 Taxonomy report.[4] We briefly summarize those findings here.

Self-controlled observational designs make comparisons in exposure frequency between an "at-risk" period and a control period within the same individual among those who experience a HOI. Self-controlled designs are most valid when the exposure is transient, the HOI is abrupt, and risk factors for the HOI are fixed within individuals over the (often short) observation period.[10] Classic epidemiologic designs employ between-person comparisons, including the cohort designs and related approaches, such as the case-control and case-cohort design, that make comparisons between individuals by sampling from the underlying full cohort to ascertain exposure and covariate distributions.

The key considerations for choosing self-controlled versus cohort contrasts are: (1) strength of within- and between-person confounding; (2) circumstances that may predispose to misclassification of exposure or misclassification of the timing of the HOI, which leads to misclassification of exposure; (3) whether the exposure of interest is transient or sustained, which can reduce short-term exposure variation; and (4) desired interpretation.[11] When the key assumptions of self-controlled designs (i.e., transient exposure, lack of within-person, time-varying confounding, and abrupt HOI) are fulfilled, this approach is generally preferred to cohort-based approaches since self-controlled designs inherently avoid confounding by fixed, between-person factors. As scenarios deviate from those in which the self-controlled assumptions are tenable, cohort-type approaches are generally preferred. These include situations in which factors that affect misclassification are present and situations that may reduce variation in exposure; namely, when signal refinement questions pertain to sustained exposures. Other

aspects of particular scenarios must also be considered and are discussed in the Year 1 Taxonomy Report.[4]

B. METHODS TO ADDRESS EXPOSURE TIME TREND

Exposure time trends can bias unidirectional self-controlled designs when case- and control-windows are systematically ordered in relation to the timing of the HOI.[12] For example, the traditional case-crossover analysis considers as the case-window some period prior to the outcome and considers as the control-window some period prior to the case-window. In the absence of a true causal relation, but in the presence of an increasing trend in exposure in the population, the exposure can appear to cause the outcome since the exposure would occur more frequently in the case-window than in the control-window. This is particularly important when studying new medical products for which use increases rapidly in the early marketing period. Other forms of exposure trends can occur if use of a medical product decreases or if seasonal (e.g., as with antimicrobials or exposures related to elective surgeries) or other secular trends in use exist.

Several methods have been developed to correct bias introduced by exposure time trends. Here, we cover: (1) the case-time-control design;[12] (2) the case-case-time-control design;[13] and (3) the bidirectional self-controlled case series.[14] We focus our discussion on aspects of these methods that adjust for exposure time trends and the characteristics that should be considered in selecting the most appropriate approach in a given scenario. For a more thorough discussion of self-controlled designs, and of the traditional case-crossover and self-controlled case series designs in particular, refer to work completed by the Mini-Sentinel Case-Based Workgroup.[9]

Case-time-control design

Suissa proposed the case-time-control approach, which uses controls that did not experience the outcome of interest to adjust for an underlying exposure trend in a traditional case-crossover analysis.[12] Controls are matched to cases on time and on other important factors (e.g., age and sex). The ratio of exposures in the case-period to the exposures in the control-period among the case patients is divided by the corresponding ratio among the control patients. The key assumption of this approach is that the controls provide an accurate estimate of the background exposure time-trend among the cases.[12,15,16]

Case-case-time-control design

Rather than using controls that did not experience the outcome of interest, Wang et al. proposed using future cases as controls.[13] This approach requires less severe assumptions about control selection since all controls will eventually become cases, thus reducing the possibility of selection bias. This approach is also analytically simpler than the case-time-control design because it does not require identification and selection of additional patients as controls.

Bidirectional self-controlled case series

Unlike the case-crossover analysis and its variants, which involve sampling control-time in a period preceding the case-window, self-controlled case series analyses include control-time before and after a presumed effect period (i.e., case-period).[14,17] Thus, time trend bias is less likely to be problematic in this design. However, fatal HOIs can introduce immortal time bias because control time cannot be sampled after the case-period among those who died while exposed. In addition, this approach can be biased by reverse causation when the HOI affects future exposure. For example, if patients discontinue a therapy after experiencing a HOI, the exposure will be underrepresented in the control-period following the event, which could result in an upward bias.

Considerations

When using a self-controlled design to investigate medical product safety, some adjustment for an exposure time trend will almost always be needed since utilization trends commonly occur, especially in the early marketing period. When case- and control-windows are short and tied closely in time, bias due to the exposure time trend will be minimal. When time trend bias needs to be addressed, the most important considerations for choosing among methods are: (1) whether the HOI is related to death such that immortal time may be problematic; and (2) whether the HOI is likely to affect future exposure. If either of these conditions holds, then unidirectional approaches, such as the case-time-control and the case-case-time-control design, are preferred. In choosing between the case-time-control and the case-case-time-control, investigators should consider the potential for selection bias in control selection and the simpler analysis associated with the case-case-time-control design. However, the case-case-time-control design is less well established in the literature at this time. In scenarios in which immortal time

bias and reverse causation are unlikely, such as with childhood vaccinations, bidirectional approaches may be reasonable.

While cohort analyses are not subject to bias due to an exposure time trend *per se*, they can be biased if an exposure time trend coexists with trends in unmeasured risk factors for the HOI. This form of confounding can occur even if the risk factor is not a determinant of exposure. For example, if a signal refinement activity sought to quantify the cardiovascular risk associated with a new medication in electronic healthcare data, aspirin use might be an important, but unrecorded, risk factor for the HOI. Because aspirin use has increased substantially in the past two decades, a historical comparator group would comprise patients with less frequent baseline aspirin use than patients exposed to the new drug. One way to overcome this type of confounding is by using concurrent comparators. Matching patients exposed to the new drug to those exposed to a concurrent comparator, for example, can mitigate confounding by an aspirin time trend.

C. CONFOUNDER SUMMARIZATION

Confounding is a principal threat to false positive and false negative alerting in an active medical product safety surveillance system. Valid effect estimation often requires investigators to account for a large number of potential confounders.[18] Confounder summary scores, such as propensity scores (PSs) and disease risk scores (DRSs),[19] offer many advantages in medical product safety analyses and in analyses conducted across distributed data networks. Summary scores can enable efficient adjustment for many more variables than can be accommodated in traditional outcome regression models, which adjust separately for each covariate, and they can enable multivariable adjusted pooling across Data Partners without compromising data privacy.[20,21]

No summarization

When the number of confounders is small relative to the number of outcomes in an analysis, dimension reduction by confounder summary scores offers little or no advantage for valid and precise effect estimation over confounder adjustment in traditional outcome models. Cepeda et al. found that, while PSs produced estimates that were less biased, more robust, and more precise than logistic regression estimates when there were seven or fewer events per confounder, logistic regression empirical

coverage probability increased as the number of events per confounder increased and the PS empirical coverage probability decreased after eight or more events per confounder.[22] However, the number of variables that one might want to adjust for, even in the absence of known HOI risk factors, can grow quickly. In the Mini-Sentinel protocol for active surveillance of acute myocardial infarction in relation to anti-diabetic drug use, the investigators specified ten demographic and baseline healthcare utilization variables (age, sex, nursing home residence, hospitalization in prior month, hospitalization in past 31-365 days, emergency department visit in past month, emergency department visit in past 31-365 days, number of outpatient visits in prior year, number of unique medications in prior year, and total number of unique medications dispensed in prior year), which does not include the additional 16 conditions and four medications identified as potential confounders and the increase in number of required model parameters to estimate if any variables are discretized into more than two levels.[5]

Propensity scores

PSs are commonly used in medical product safety studies, owing primarily to the large number of confounders that usually need to be addressed relative to the often-infrequent HOI.[23] A patient's PS is his or her predicted probability of exposure to a particular product conditional on observed covariates.[24] Patients in different exposure groups who have the same PS will, on average, have similar distributions of covariates that entered the PS model. PSs were recently proposed as an approach to individual-level multivariable adjustment that maintains data privacy in distributed data analyses.[20,21] PS matching also enables analyses of multiple outcomes among the same PS matched cohorts, which can simplify analyses as compared to other methods that require separate models for each outcome. In addition, fixed-ratio PS matching allows for valid effect estimation without the need for additional statistical models, though accounting for matching in the analysis can increase precision.[25] PSs are particularly well suited for scenarios characterized by a single comparator group, a rare HOI, and when multiples outcomes are to be evaluated for individual products.

Disease risk scores

DRSs are the prognostic analogue to PSs, in which an investigator estimates each patient's probability or hazard of an HOI conditional on baseline risk factors for the outcome.[26,27] DRSs can be more practical than PSs when comparing more than two levels or categories of exposure since a single DRS model can

be used for all patients.[26] In addition, determinants of use of a new medical product can evolve rapidly in the period shortly after market authorization, complicating the use of PSs, particularly in sequential analyses. However, risk factors for the outcome remain relatively stable in a population over a period of a few years. For example, risk factors for myocardial infarction are similar and have similar associations with myocardial infarction today as compared to two years ago. This enables fitting of DRSs in recent historical populations.[28] A key limitation of DRSs is that they are less practical than PSs when investigating multiple outcomes per medical product since separate DRS models would need to be fit for each HOI.

A key consideration in the use of DRSs is in which population to develop the DRS model.[19,26] In retrospective inquires, typical options include the full cohort, the unexposed members of the full cohort, or an external population, such as in a historical population as mentioned above. In prospective assessments when the full cohort has not yet been enumerated, the first two options are unrealistic at least until a substantial number of events accrue to enable robust DRS calculation; therefore, an external population might be most appropriate. Further, development in the full cohort requires an assumption of absence of effect measure modification between the treatment and potential confounders. Alternatively, the DRS model could be fit in the whole cohort with treatment-by-covariate interaction terms. To our knowledge, this approach has not been explored empirically but should be considered for future Mini-Sentinel methods work. Additionally, future work should focus on combining PSs and DRSs to leverage the advantages of both approaches in a single analysis, particularly in the early marketing period when exposure determinants evolve rapidly.

Additional considerations

Variable selection is an important issue for all confounding control approaches.[29] In any conditioning strategy (e.g., matching, conditioning, restriction, modeling), regardless of whether confounder summary scores are used, investigators should aim to control all risk factors for the outcome.[30] For DRSs, this is a natural approach to variable selection. However, for PSs, this requires a shift in thinking from modeling the treatment selection process, to one aimed at controlling variation in exposure related to the outcome (i.e., “bad” exposure variation) but not variation in exposure independent of the outcome of interest (i.e., “good” exposure variation).[31] Analyses that employ any conditioning strategy

exploit the latter and are only possible in its presence since a perfectly discriminating PS model would render an analysis impossible because of completely non-overlapping patient populations (i.e., complete lack of exchangeability).

Theoretical analyses[30] have suggested and a statistical simulation performed by the Mini-Sentinel Signal Evaluation Workgroup[32] has confirmed that, in the presence of unmeasured confounding, conditioning on a variable that is associated with exposure but not associated with the outcome except through its association through exposure (i.e., an instrumental variable or instrument), can increase the amount of confounding bias. However, the simulation studies found that, since this phenomenon occurs only in the presence of unmeasured confounding, the magnitude of increase in bias due to conditioning on an instrument is generally trivial relative to the amount of residual confounding. While investigators should not include in the conditioning set those variables known or highly suspected of being instruments, removing as much confounding as possible mitigates residual confounding and bias amplification due to conditioning on an instrument whereas omitting suspected instruments from the conditioning set affects only the latter. Moreover, the small potential increase in bias related to adjusting for instruments affects all approaches to conditioning and not just PSs. For a more detailed discussion about bias amplification, refer to the Signal Evaluation Workgroup report.[32]

D. INCORPORATING CONFOUNDING CONTROL INTO ESTIMATION

Once investigators determine whether and how to summarize potential confounders, they must decide how to incorporate the individual confounders or summary scores into the effect estimation procedure. As confounder scores represent summaries of multiple confounders, they can be incorporated into the analysis in the same way as individual variables, including by restricting, stratifying, or matching on the score, and also by weighting (as with standardization approaches) by the score, and including the score as a covariate in an outcome model.

Restriction

If there is only interest in the treatment effect within a particular level of the PS or DRS, then restriction represents a suitable approach to address confounding. Restriction limits the comparison to patients with similar levels of the components of the summary score. Comparisons will have internal validity that

derives from the confounder control the summary score offers, but may have limited generalizability, especially in the setting of treatment effect heterogeneity.

Stratification

Stratification by confounder summary score is a good option for obtaining within-site estimates, which can then be meta-analyzed across sites in a distributed data setting to obtain an overall estimate. Alternatively, in a setting in which individual-level data can be pooled, one could stratify analyses by summary score and by site. Research is warranted to evaluate whether patients with similar DRS values are comparable across sites, which could simplify analyses by allowing stratification after pooling across sites. However, patients with similar PSs should not be pooled into strata across sites if PSs were estimated separately within each site. PSs may not be comparable across Data Partners since patients' PS values will depend on the prevalence of the exposure of interest in the population in which the PS is estimated. Specifically, the mean PS in a population will, in expectation, be equal to the prevalence of the exposure in that population. Many factors, including formulary restrictions and regional prescribing patterns, will influence the prevalence of drug use in a data partner. Thus, incorporating PSs into estimation should occur at the level of the Data Partner as exposed patients with high PSs in low-prevalence plans would likely not have distributions of baseline covariates similar to unexposed patients with low PSs in high-prevalence plans even if these patients have similar PSs.

An important limitation of stratification is that sequential analyses will result in a large number of strata (i.e., the number of strata times the number of monitoring periods), which can reduce estimation precision. One solution might be to re-stratify all patients at each time point, but this can complicate sequential testing procedures since a patient that contributes to the analysis in one period could theoretically not contribute to the analysis in a subsequent period if the patient is shifted to a non-informative stratum.

Matching

An implicit benefit of matching is that it restricts the population to those patients for whom positivity exists; that is, it restricts analyses to those patients in the range of overlap in confounder summary score distributions between treatment groups. Positivity is a required assumption for causal inference. Fixed

ratio matching enables simple assessments of observed covariate balance between treatment groups through visual examination of a Table 1 and by using various metrics (e.g., Mahalanobis distance, standardized differences, etc.). If matching achieves sufficient balance on measured confounders that contribute to the summary score, no additional statistical analysis is needed to control those variables when fixed ratio matching is employed. However, as in randomized trials, additional adjustment can be made for variables that are not well balanced. In addition, accounting for the matching in the analysis, such as by using a matched-pair analysis, can increase the precision of effect estimates.[25]

While matching implicitly restricts the study base to those with overlapping confounder summary score values, matching can also result in a loss of information when the number of eligible matches exceeds the matching ratio for any exposed patient. Variable ratio matching with no maximum ratio can alleviate this and is similar to fine stratification. In addition, as with stratification, patients should not be matched on the PS across sites if separate PS models were developed within each site.

Weighting

Another approach to overcoming the loss of information associated with fixed ratio matching is by reweighting the population by some function of the confounder summary score, such as by the inverse of the propensity score, as with inverse probability of treatment weights (IPTW).[33]

IPTWs create a weighted population (i.e., a "pseudo-population") comprising exposed and unexposed patients with similar distributions of covariates that were used to estimate the weights. A key consideration in using weighting is determining to which covariate distribution the pseudo-population will be standardized. In the presence of effect measure modification (or treatment effect heterogeneity) different standard populations will yield different results.[34] Reweighting to the distribution of covariates in the entire population provides marginal estimates (i.e., estimates pertaining to the effect of the treatment if everyone in the population received it versus the effect if the entire population did not receive it) while standardizing to the distribution of covariates among the exposed yields an estimate consistent with the effect of the treatment among those who actually received it.

While weighting allows retention of some information that is lost in fixed ratio matching, it cannot overcome non-positivity. That is, weighting (and also regression) cannot provide valid estimates for patients in the areas of non-overlap. Restricting patients to those in areas of overlap before weighting can avoid bias in effect estimates resulting from weighting when positivity does not hold.[35]

As independent variable in model

In addition to matching, stratifying, or weighting by confounder summary score, analyses can also be adjusted by including the score along with a treatment variable in an outcome model, such as in a Cox proportional hazards model. This approach adjusts for confounding to a similar degree as including all of the component variables in the outcome model, but it consumes fewer degrees of freedom. One concern about this approach, particularly for PSs, is that removal of confounding depends on specifying the correct functional relation between the summary score and the outcome of interest.

Including a summary score in an outcome regression model does not allow for a visual inspection of covariate balance, as in inspecting a Table 1 in a matched comparison. In addition, modeling extrapolates over areas of non-overlap of the confounder summary score, which can lead to bias. As with weighting, trimming should be considered prior to including the summary score in an outcome regression model.

Additional confounding adjustment

When stratifying, matching, or weighting by a confounder summary score, no further covariate adjustment may be necessary. For example, when matching by a fixed ratio, whether by individual covariates or by confounder summary scores, no further adjustment may be needed to obtain valid estimates if sufficient balance in observed confounders is achieved. Additional adjustment can be made in an outcome model for variables that exhibit imbalance between groups or when additional adjustment for strong risk factors is desired. When variable ratio matching is used, the matching must be accounted for in the analysis of cohort studies in order to produce valid effects. This is because matching by variable ratios produces valid estimates within strata defined by the matching ratio, but does not produce unbiased marginal estimates. Additionally, while use of fixed ratio matching does yield unbiased marginal estimates, accounting for the matching in the analysis, such as with McNemar's test,

can increase the precision of estimation. An analysis without further adjustment may be particularly advantageous in rapid assessments of multiple outcomes when matching is used.

E. ESTIMATION

The choice of an estimation procedure for the point and confidence interval estimation for safety effects depends on a number of factors. These include, but are not limited to, the effect measure of interest (e.g., difference measure [e.g., cumulative incidence difference] or relative measure [e.g., incidence rate ratio]), the way in which the HOI is measured (e.g., binary, count, continuous, time-to-event), the preferred assumptions regarding the distribution of the mean and variance of the outcome (e.g., binomial, normal, Poisson, robust, distribution-free), the HOI frequency (e.g., exact versus large-sample variance estimation methods), and whether correlation exists between outcomes. In this section, we briefly describe some characteristics of estimation options for common scenarios expected in a monitoring setting like Mini-Sentinel.

Cohort analyses, in which between-person comparisons are made, will be common within Mini-Sentinel. HOIs will often be binary, either observed acutely or over longer periods of time following exposure where person-time analyses are needed to account for variable follow-up time among patients due to many factors, such as treatment discontinuation and censoring for other reasons. The active surveillance protocol involving saxagliptin and acute myocardial infarction is one such example where person-time analyses of a binary endpoint are being conducted.[5] Estimation can be conducted using Poisson regression, which estimates quantities expressed as either incidence rate ratios (i.e., by using a log link function) or incidence rate differences (i.e., by using an identity link). A commonly used alternative to Poisson regression is the Cox proportional hazards model. The semi-parametric Cox model does not estimate the baseline hazard rate and instead assumes only that the hazard ratio is proportional over time between exposure groups or between other covariate groups. As such, Cox proportional hazards models will generally be preferred when only hazard ratios are desired, since they require fewer assumptions.

When the onset of the exposure risk window is immediate and its duration is short, such that events are observed acutely following exposure, person-time analyses may not be necessary and other regression

approaches may be appropriate. For example, linear regression models that use robust variance estimation methods can be used for binary outcomes to estimate risk differences. If ratio measures are of interest, Poisson or logistic regression (which produces odds ratios that approximate cumulative incidence ratios when the outcome is rare) models can be used. Assumptions regarding the variance specified by these generalized linear models (GLMs) can be loosened by applying robust sandwich variance estimation methods (i.e., generalized estimating equation or GEE methods).

Within-person analyses typically utilize generalized linear models, including the logistic regression model and the Poisson model. In particular, the self-controlled case series is usually analyzed with a Poisson model and case-crossover analyses, and their variants, are most often analyzed with conditional logistic regression models. These models are readily applicable with standard statistical software.[9,17]

IV. WORKED EXAMPLES

In applying the Taxonomy framework, multiple design and analytic approaches may be appropriate for a given scenario. The goal of this framework is to provide general guidance to facilitate expeditious and transparent decision-making. Below, we demonstrate how the Taxonomy framework can be used when applied to multiple scenarios. The important first step is to codify the scenario according to the key characteristics that are likely to influence methods selection. Ultimately, the utility of the framework depends on how well each scenario can be characterized, which can be a difficult task that will require input from multiple stakeholders. Note that not all of the examples below represent cases in which an association (causal or otherwise) is thought to exist.

A. INTRAVENOUS IMMUNE GLOBULINS (IVIG) AND THROMBOEMBOLISM

Production of IVIG begins with pooling of human plasma from thousands of donors. IVIGs are used to treat primary and secondary immune deficiencies, autoimmune disorders, and inflammatory disorders. Low dose IVIG (e.g., 300 to 400 mg/kg per month) protects against infection when used for passive immunization or repletion of deficient states and high dose (e.g., 1 to 2 grams/kg per dose) suppresses inflammatory or immune-mediated process. IVIGs are available as nine different intravenous types and four subcutaneous types. Subcutaneous formulations are given weekly over 2 hours over multiple body areas. IVIGs are usually administered in health care facilities, such as infusion centers and less commonly

administered at home. IVIG recipients are at an increased risk for thrombosis due to high blood viscosity, or infusion of plasma constituents that raise risk of clotting. Common uses of IVIG include chronic neuropathy, secondary hypogammaglobulinemia, idiopathic thrombocytopenia, primary immunodeficiency, and renal transplantation.[36]

Table 4 describes the characteristics of a scenario defined by IVIG for acute illness as the exposure and thromboembolic events as the outcome of interest.

Table 4. Example 1: Intravenous Immune Globulins (IVIG) *for acute illness* and Thromboembolic events (TEE)	
Characteristics determined by stakeholders/investigators	
Effect measure(s) of interest	Both difference and ratio measures
Comparator(s)	Assume we are interested in comparing TEE risk across the 4 subcutaneous types of IVIG
Exposure characteristics	
Background frequency of use:	Infrequent (<100,000 patients/year best guess)
Utilization trend in population:	Increasing
Use pattern	Short-term
Characteristics of the potential exposure-HOI link	
Onset of exposure risk window:	Immediate
Duration of exposure risk window:	Short (within 7 to 14 days)
Strength of confounding	
Between-person	Needs to be addressed
Within-person	Needs to be addressed
HOI Characteristics	
Background frequency	Infrequent (~1% of exposed)
Periodicity	Can be recurrent
Expected degree of onset misclassification	Negligible

Because both between- and within-person confounding needs to be addressed, there is no clear choice of contrast. However, in comparing different types of IVIG, a between-person analysis may sufficiently address baseline confounding and, since utilization is increasing in the population, exposure time trend adjustment would be necessary in a within-person analysis. Because the exposure of interest is

infrequent and the event of interest is uncommon, a confounder summary score may be necessary to facilitate adjustment of thromboembolic risk factors. With multiple IVIG exposure groups, confounder summarization by disease risk score (DRS) may be more straightforward than by propensity score. Any approach to incorporating the DRS into the analysis may be appropriate.

Table 5 presents the characteristics for a scenario defined by IVIG for chronic immune deficiency as the exposure and thromboembolic events as the outcome.

Table 5. Example 2: Intravenous Immune Globulins (IVIG) for chronic immune deficiency and Thromboembolic events (TEE)	
Characteristics determined by stakeholders/investigators	
Effect measure(s) of interest	Both difference and ratio measures
Comparator(s)	Assume we are interested in identifying TEE risk for a particular IVIG type versus no treatment
Exposure characteristics	
Background frequency of use:	Infrequent (<100,000 patients/year best guess)
Utilization trend in population:	Increasing
Use pattern	Long-term
Characteristics of the potential exposure-HOI link	
Onset of exposure risk window:	Immediate
Duration of exposure risk window:	Short (within 7 to 14 days)
Strength of confounding	
Between-person	Needs to be addressed
Within-person	Negligible
HOI Characteristics	
Background frequency	Infrequent (~1% of exposed)
Periodicity	Can be recurrent
Expected degree of onset misclassification	Negligible

In this scenario, where we assume that we are interested in comparing IVIG for chronic immune deficiency to no treatment, a within-person comparison may be more appropriate, especially since IVIG administration is not likely to be driven by acute disease flare-ups. Because IVIG use is increasing, a

method to address exposure time trends may be needed, such as a case-case-time control approach. A conditional logistic regression analysis would be appropriate to analyze a case-crossover type assessment.

B. SUSTAINED USE OF LISINOPRIL AND ANGIOEDEMA

The next five scenarios (examples 3-7) were included in the Year 1 Taxonomy report, in which we developed the most appropriate contrast. Below each table, we briefly describe considerations for each subsequent decision point.

Table 6. Example 3: Sustained use of lisinopril and risk of angioedema	
Characteristics determined by stakeholders/investigators	
Effect measure(s) of interest	Both difference and ratio measures
Comparator(s)	No treatment or treatment with active comparator not thought to be associated with angioedema (e.g., angiotensin receptor blockers)
Exposure characteristics	
Background frequency of use:	More frequent
Utilization trend in population:	Uniform
Use pattern	Long-term
Characteristics of the potential exposure-HOI link	
Onset of exposure risk window:	Immediate
Duration of exposure risk window:	Long
Strength of confounding	
Between-person	Negligible
Within-person	Negligible
HOI Characteristics	
Background frequency	Rare
Periodicity	Assume we are interested in the first event following initiation (but patients may have had events prior to treatment)
Expected degree of onset misclassification	Negligible (within days)

As the monitoring question was formulated to evaluate the safety of sustained lisinopril use, rather than pertaining to initiation of lisinopril, a cohort-type contrast was preferred. Since lisinopril is a fairly

frequent exposure and angioedema is rare, a propensity score approach is preferred over disease risk score. However, since risk factors for angioedema are not well established, no summarization may also be appropriate, depending on the observed number of events. A propensity score could appropriately be incorporated into the effect estimation using stratification, matching, as a regressor in a model, or via weighting. If using the propensity score in an outcome model or simply adjusting for independent covariates in an outcome model, either a Poisson or Cox model may be appropriate, as either will enable a person-time analysis.

C. MEASLES, MUMPS, AND RUBELLA VACCINATION AND FEBRILE SEIZURES

Table 7. Example 4: Measles, mumps, and rubella vaccination and febrile seizures	
Characteristics determined by stakeholders/investigators	
Effect measure(s) of interest	Both difference and ratio measures
Comparator(s)	Unexposed
Exposure characteristics	
Background frequency of use:	More frequent (among population of interest, i.e., children)
Utilization trend in population:	Uniform
Use pattern	Short-term
Characteristics of the potential exposure-HOI link	
Onset of exposure risk window:	Immediate
Duration of exposure risk window:	Short
Strength of confounding	
Between-person	Needs to be addressed
Within-person	Negligible
HOI Characteristics	
Background frequency	Rare
Periodicity	Assume we are interested in the first seizure following vaccination
Expected degree of onset misclassification	Negligible

Self-controlled designs are well suited for assessing the safety of childhood vaccines. Since the utilization trend among children is relatively stable and because issues of reverse causation and immortal time bias are likely negligible, either a traditional case-crossover approach or a self-controlled case series is appropriate. A case-crossover analysis could use a conditional logistic regression model or a Mantel-Haenszel approach. A self-controlled case series could use a Poisson model.

D. ROSUVASTATIN AND RHABDOMYOLYSIS

Table 8. Example 5: Rosuvastatin and rhabdomyolysis	
Characteristics determined by stakeholders/investigators	
Effect measure(s) of interest	Both difference and ratio measures
Comparator(s)	Other statins (excluding cerivastatin)
Exposure characteristics	
Background frequency of use:	More frequent
Utilization trend in population:	Changing (increasing)
Use pattern	Long-term
Characteristics of the potential exposure-HOI link	
Onset of exposure risk window:	Immediate
Duration of exposure risk window:	Long
Strength of confounding	
Between-person	Negligible (when compared to other statins)
Within-person	Negligible
HOI Characteristics	
Background frequency	Rare
Periodicity	Once
Expected degree of onset misclassification	Negligible (within days)

Owing to the intended long-term use pattern of statins and to the fact that the monitoring question pertains to the comparative safety of rosuvastatin to other statins, which limits the amount of between-person confounding, a cohort-type approach is preferred. As with angioedema, risk factors for rhabdomyolysis are not well established. As such, a propensity score approach or no summarization may be appropriate. A robust disease risk score model will be difficult to construct because of the uncommonness of rhabdomyolysis. A person-time analysis will be important and could be accommodated in either a Poisson model or a Cox regression model.

E. AMPHOTERICIN B AND ACUTE LIVER FAILURE

Table 9. Example 6: Amphotericin B and acute liver failure	
Characteristics determined by stakeholders/investigators	
Effect measure(s) of interest	Both difference and ratio measures
Comparator(s)	No treatment or treatment with active comparator not thought to be associated with acute liver failure
Exposure characteristics	
Background frequency of use:	Less frequent
Utilization trend in population:	Cyclical (assume some seasonality in use)
Use pattern	Short term
Characteristics of the potential exposure-HOI link	
Onset of exposure risk window:	Immediate
Duration of exposure risk window:	Short
Strength of confounding	
Between-person	Needs to be addressed
Within-person	Needs to be addressed
HOI Characteristics	
Background frequency	Rare
Periodicity	Once
Expected degree of onset misclassification	Negligible

If the infection prompting treatment with amphotericin B is itself a risk factor for acute liver failure, then self-controlled approaches and cohort-approaches that use truly unexposed time as the comparator will be subject to confounding by indication. If a reasonable active comparator exists, such as another antimicrobial used to treat the same infection, then a between-person contrast is preferred. If a within-person approach is selected, adjustment for the cyclical exposure time trend would be necessary. The low outcome incidence argues in favor of propensity scores in a between-person comparison, particularly since the analysis will likely need to account for a large number of confounders. If the follow-up period is very short, then a logistic regression model could be considered since the odds ratios will approximate risk ratios. Otherwise, any approach to incorporating confounder into estimation and any of the models may be appropriate.

F. MECHANICAL HEART VALVE AND THROMBOEMBOLISM

Table 10. Example 7: Mechanical heart valve and thromboembolism	
Characteristics determined by stakeholders/investigators	
Effect measure(s) of interest	Both difference and ratio measures
Comparator(s)	No heart valve replacement
Exposure characteristics	
Background frequency of use:	Less frequent
Utilization trend in population:	Changing (increasing)
Use pattern	Long-term (assume we are interested in long-term effects rather than effects related to placement)
Characteristics of the potential exposure-HOI link	
Onset of exposure risk window:	Immediate
Duration of exposure risk window:	Long
Strength of confounding	
Between-person	Needs to be addressed
Within-person	Needs to be addressed
HOI Characteristics	
Background frequency	Infrequent
Periodicity	Recurrent
Expected degree of onset misclassification	Negligible

Because placement of the heart valve may be a risk factor for thromboembolism and because the monitoring question relates to the long-term effects of exposure to the heart valve, a cohort approach is strongly preferred. Identifying a good comparison group will be difficult and important confounders may not be recorded in the database. However, the number of measured variables will likely be large relative to the outcome frequency, so a propensity score approach may be preferred. A person-time analysis will be important to account for variable follow-up and either a Poisson or Cox model would be appropriate.

V. ADDITIONAL SPECIFIC METHODOLOGICAL CONSIDERATIONS IN ROUTINE ACTIVE MONITORING

In addition to the decision points represented in the table, several other considerations are necessary in active safety monitoring, including: (1) whether to focus on new users or also include prevalent users of the medical product; (2) whether to consider advanced methods; (3) diagnostic and sensitivity analyses; and (4) whether to perform a one-time assessment or sequential analyses. If performing sequential analyses, investigators must also determine how to space monitoring periods.

A. ROLE OF PREVALENT USERS IN ACTIVE SAFETY SURVEILLANCE

The Taxonomy Workgroup considered when it might be acceptable to include prevalent users in medical product safety monitoring programs. When establishing a monitoring program for a particular medical product, Mini-Sentinel investigators must decide whether to focus on *incident users* or whether to include both incident and *prevalent users*. We define incident users in a database(s) as those patients with recorded exposure to the medical product following an observable period of a minimum pre-specified length with no recorded exposure to the product (or, sometimes, to similar products). We define prevalent users as those patients who persist on therapy such as those in the database(s) with a recorded exposure to the medical product but who do not meet the requirement of an exposure-free baseline period. For example, prevalent users may be those patients who are exposed to a medical product prior to enrollment in the database and who continue their exposure to the product thereafter.

Benefits of focusing on incident users

Incident user only designs can mitigate important types of bias: (1) so-called "survivor" bias (i.e., when patients who remain on treatment represent a special, less susceptible, subgroup of all treated patients); (2) bias due to inability to address baseline confounding (i.e., inability to observe in the database treatment initiation such that pre-treatment confounders cannot be measured and addressed); and (3) biases due to conditioning on factors affected by the medical product (i.e., adjusting for factors that are downstream of exposure and that may be on the causal pathway between exposure and outcome.[37] Focusing on incident users allows the investigator to identify and include in the analysis those events that occur shortly after drug initiation. When the outcome hazard function varies with time

since initiation, which occurs frequently when evaluating the outcomes of medical interventions, prevalent users represent a heartier subgroup of patients than the entire cohort of patients who initiated the medical product since those who experienced events shortly after initiation are implicitly excluded. This partly explains the discrepancy between observational studies and randomized controlled trials that examined cardiovascular events in relation to hormone therapy. Patients included in the observational studies were those who did not experience events shortly after initiation, a time during which randomized trials found a large elevated risk of events related to hormone exposure.[37]

Defining the start of follow-up at the first exposure to a medical product enables the investigator to establish clear temporality among confounders, exposures, and outcomes, to facilitate proper handling of these variables in the analysis. On the other hand, covariate adjustment in prevalent user studies will necessarily involve adjustment for "on treatment" variables,[38] which can both obscure relations between the exposure and an outcome,[39] when the factor is an intermediate on the causal pathway, and produce spurious associations when the factor is associated with the outcome in other, non-causal, ways.[39]

Situations in which prevalent users might be needed

Certain situations may present that could preclude investigators from monitoring outcomes among incident users only, such as if:

(1) Most patients enter the database(s) already using the medical product of interest, such that an insufficient number of incident users can be identified to produce useful results; or
(2) Patients identified as incident users in the database(s) leave the database(s) before experiencing the outcome(s) of interest, rendering monitoring of only incident users futile.

In these cases, monitoring may be impossible without including prevalent users. Therefore, investigators may wish to initiate monitoring with prevalent users as long as the following assumptions and implications are made explicit.

Assumptions and implications of including prevalent users

(1) When focusing monitoring evaluations on prevalent users, investigators must be mindful of time-varying hazards and depletion of susceptibles. Observational studies of hormone therapy did not produce the wrong answer, *per se*, but they produced an answer to a very different question than that answered by the randomized trials. The observational studies answered the question: "What is the risk of cardiovascular events among women using hormone therapy who survived the initial period of elevated cardiovascular risk shortly after starting treatment?" Monitoring prevalent users does not provide insight into the general safety of medical products, but rather into the safety of medical products among patients who "survive" on them.

(2) Adjustment for "on-treatment" variables requires either (a) the assumption that these factors are not influenced by prior exposure; or (b) appropriate methods (e.g., g-methods[40]) to handle factors that are affected by prior treatment. Otherwise, monitoring results may be biased in either direction. In addition, pre-treatment values of potential confounders remain unadjusted and can result in residual bias.

(3) Prevalent users are generally a special subgroup comprising those patients who persist with (i.e., "adhere" to) treatment. Even independent of the survivor effect noted above, patients who persist on their treatment regimens are generally healthier than those who do not. In the absence of an active comparator group with similar persistence tendency, outcomes may be a function of the patients' propensity toward a healthy lifestyle. Even in the presence of an active comparator group that neutralizes bias due to healthy adherer effects, the resulting effect estimate may not be generalizable since observed event rates in these groups may be substantially lower than in the general population of users, which comprises both "persisters" and "non-persisters."

Recommendation for inclusion of prevalent users

When a sufficient number of incident users are available with enough follow-up time to yield enough outcomes to produce useful results, prevalent users should generally be avoided. Inclusion of prevalent users should be limited to the setting in which investigators are monitoring outcomes following long-term exposure to a medical product, such that few incident users can be identified in the database(s). In addition, the assumptions and implications of prevalent user biases must be considered and a

determination must be made about whether monitoring should proceed despite the potential for these biases.

B. ROLE OF NASCENT AND ADVANCED ANALYTIC METHODS IN CURRENT SAFETY SURVEILLANCE

The Workgroup considered the role of advanced and nascent analytic methods (e.g., marginal structural models, targeted maximum likelihood, etc) in current safety surveillance activities. While these methods are theoretically sound and offer great promise for epidemiologic analyses, they can add considerable complexity to surveillance protocols. When time-varying confounders are known and measured, these methods should be considered. However, in near-real-time surveillance activities in which analyses must be completed quickly and with limited resources, such advanced methods may currently be more appropriate as sensitivity analyses performed at the end of monitoring. Aspects of these approaches, such as inverse probability of censoring weighting to address informative censoring, should also be considered, as appropriate. Future Mini-Sentinel methods work should more fully explore advanced approaches in the electronic healthcare data environment and specifically for routine active surveillance, focusing on ways to validly, but expeditiously, employ these strategies.

C. DIAGNOSTIC AND SENSITIVITY ANALYSES

In addition to primary analyses, signal refinement activities should consider diagnostic, secondary, and sensitivity analyses. Diagnostic and secondary analyses can be performed alongside the primary analysis while sensitivity analyses will generally be performed at the end of monitoring. Diagnostics analyses can help identify cases in which programs do not work as expected, such as when a coding error occurs or when variables that should not enter propensity score models are inadvertently included. Secondary analyses can provide information to support decision-making in addition to results from primary analyses. Sensitivity analyses can help test certain methodological assumptions and test the robustness of study findings.

Different outcome definitions

Substantial effort has been dedicated to validating algorithms for defining outcomes in the Mini-Sentinel Distributed Database.[41,42] Algorithmic outcome definitions employed in electronic healthcare databases can vary with respect to the specificity of clinical event that they identify. For example, an outcome

could be defined as serious bleed, comprising severe upper gastrointestinal (GI) bleeding and intracranial hemorrhage. A drug that increases patients' proclivity to all serious bleeding events, such as an anticoagulant, could affect both GI and intracranial bleed. However, other drugs, such as non-selective non-steroidal anti-inflammatory drugs (NSAIDs), might act only on one of the component events, such as GI bleed. Varying outcome definitions that yield results that vary in accordance with expected biological mechanisms, can increase one's confidence in the results. For example, we may be interested in a finding that a new NSAID increases the risk of GI bleeding but would be wary of the finding if the drug also increased the risk of intracranial hemorrhage. One the other hand, if the new NSAID increases only GI bleed risk, then the association may be obscured if the original outcome definition included both types of bleed.

"Negative control" exposures and outcomes

Negative controls can be useful tools for detecting confounding and bias in observational analyses.[43] Active comparators are a form of negative control exposures and are addressed in more detail in the Taxonomy Year 1 report. Negative control outcomes can be useful for identifying systematic bias such as that due to "healthy user" effects. Dormuth and colleagues found that, as compared to those who do not adhere to statins, patients who adhere to these drugs are less likely to be involved in motor vehicle and workplace accidents and were more likely to use screening services.[44] Including preventive screening services, as long as they are unrelated to the exposure of interest, can help flag the presence of bias in signal refinement activities. Similarly, "positive controls" can also tip off the investigator to anomalous situations. For example, in evaluating the association between statins and an adverse event of interest, investigators could simultaneously evaluate the drugs' association with cardiovascular events. A lack of such an association, or an observed association that is inconsistent with clinical trial evidence, could suggest a flaw in the design or analysis or their implementations.

Model and balance diagnostics

When using outcome regression models, usual diagnostics (e.g., examining residuals, assessing proportional hazards assumptions, etc) should be applied. When applying confounder summary scores, one can also assess covariate balance and empirically assess positivity by examining overlap in the summary score distributions between treatment groups. A number of metrics exist to quantify covariate

balances between treatment groups. For example, the Mahalanobis distance is often used to quantify the difference in two groups across a range of covariates by taking into account the correlated nature of the covariates. Other metrics, such as the standardized difference or absolute differences in covariate values, can be used to quantify the balance in individual covariates.

D. ONE TIME ASSESSMENT VERSUS SEQUENTIAL MONITORING

The Workgroup considered factors that determine whether a signal refinement activity should be a one-time assessment or a sequential monitoring activity, with repeated looks at accumulating data. Rather than a methodological problem, the question is one about decision-making with implications on methods selection, particularly with respect to selecting testing methods. The fundamental determination is whether a single look provides sufficient information for decision-making. While this may, to some extent, be predictable prior to the implementation of a monitoring activity, the decision of whether to continue monitoring beyond the first look or to terminate monitoring for a specific product-outcome pair must be made conditional on the information produced in the first monitoring period. Terminating monitoring for a specific product-outcome pair after the first period is then an extreme case of sequential monitoring. Thinking about sequential monitoring along this continuum, where a one-time look represents an extreme case of sequential monitoring, is a way to unify this framework.

From a decision-making standpoint, whether to initiate sequential monitoring to gain continuous near-real-time information or to wait until more information has accrued, hinges on two factors: (1) the importance of timeliness of the information; and (2) resource availability. If the safety information does not need to be received immediately and resources are constrained, then it might make sense to wait and plan for a one-time assessment. However, if resources are unlimited and safety information is needed immediately, then a sequential analysis would be advantageous. Most monitoring questions will likely fall somewhere in between these two scenarios, where more rapidly available safety information will facilitate timely public health decision-making, but resources are nonetheless limited. Thus, the decision invokes the notion of a value of information problem.

Sequential monitoring involves a higher resource cost than a one-time assessment since it requires repeatedly querying the distributed data network and thus involves both human effort and computing

time. It also involves additional resources to interpret results at each analysis. On the other hand, there may be a public health toll associated with not analyzing and acting upon accruing information when an important safety issue exists.

Other practical considerations include when the decision to initiate a signal refinement activity is made in relation to the amount of time the medical product has been on the market. Current Mini-Sentinel signal refinement activities are divided by the time of activity implementation relative to the market authorization date of the product of interest. Specifically, activities are considered for drugs that have been on the market for less than two years and, separately, for drugs that have been on the market for two or more years. If stakeholders determine that a signal refinement activity is needed for a product that has been on the market for several years, a one-time assessment may provide sufficient information for decision-making without the need for any additional time waiting for data accrual.

Only after policy, resource, and public health factors have been taken into account should methodological aspects be considered in deciding whether sequential monitoring is needed beyond a one-time assessment. Also, it is important to recognize that sequential effect measure estimation (i.e., continuously updating effect estimates as data accrue in the distributed database with or without formal statistical testing) and sequential testing (i.e., repeatedly testing a formal statistical hypothesis) can be separated and provide different information for decision-making. Whereas sequential testing provides support about whether an estimate is statistically different from a null hypothesis, sequential effect measure estimation can be used to refine estimates of effect between formally defined sequential testing intervals and even beyond the set number of monitoring periods specified for the sequential test.

Methodological considerations

The value of a one time assessment versus sequential monitoring will be a function of the number of exposed patients in the distributed data network that would be eligible for analysis at the time of protocol implementation and the rate of growth of the number of exposed patients. The rate at which patients exposed to a new medical product accrue in a distributed database will be a function of many factors, including incidence/prevalence of treatment indication, size of the target population (e.g., may

only be interested in older patients), rate of product adoption in clinical practice, and formulary restrictions among Data Partners. The number of patients exposed to a comparator product would be another important, but secondary, consideration. The number of outcomes anticipated among analysis-eligible patients (or number of outcomes with exposure crossover for self-controlled designs) is another key consideration, and will depend on the incidence of the outcome as well as on the relevant exposure-risk period. Uncommon outcomes or those associated with a long exposure-risk period may necessitate sequential monitoring.

The decision as to whether to perform a one-time assessment or sequential monitoring will likely be product and outcome specific. A one-time assessment may be sufficient for a particular product-outcome pair if the outcome is common, but sequential monitoring may be needed for another outcome related to the same product if the outcome is less frequent. For some activities, the decision will be obvious. For example, sequential monitoring will be needed if FDA is interested in developing evidence in a prospective manner as soon as a product enters the market. Other situations may involve a one-time assessment with the option for additional sequential monitoring; i.e., a one-time assessment could be performed with a subsequent determination about whether additional data are needed, in which case the data from the "one-time assessment" could become the first monitoring period.

This discussion also raises the question of when monitoring should end for a particular product-outcome pair. There may be value in monitoring an association even after a sequential statistical test has rejected or formally failed to reject the null hypothesis. Formal sequential testing procedures require investigators to specify the intended run-time, usually defined in terms of an observed number of outcomes and based on pre-defined alpha and beta levels. For example, a sequential boundary might be set to constrain overall Type I error at 0.05 (i.e., alpha = 0.05) and allow 10 looks at the data (i.e., monitoring periods) until 100 outcomes have accrued. The test might be designed to have 80% power (i.e. beta = 0.20) to detect an association of a particular magnitude. If, after 100 outcomes accrued, the test failed to reject the null hypothesis, an alert would not be generated. However, continued monitoring might reveal that the failure to reject the null hypothesis was due to a Type II error, or that the composition of treated patients changed over time such that the medical product had little or no effect on the HOI of interest among the patients treated in the early marketing period but had a large

effect among patients using it later on because of differences in patient characteristics that modify the treatment effect.

E. DETERMINING OPTIMAL WINDOW LENGTH IN SEQUENTIAL DRUG SAFETY MONITORING

If sequential monitoring is to be performed beyond a one-time assessment, an important next question is how long each observation period should be. Many of the factors that must be considered in determining optimal window length are closely related to factors influencing whether sequential analyses are needed beyond a one-time assessment. Here we describe general considerations for making this decision.

Numbers of outcomes

The number of outcomes that are expected in a monitoring period is a key consideration in determining window length since number of exposed outcomes is the main driver of statistical power. Ideally for sequential analyses, each monitoring period would add sufficient outcomes to generate an alert without spending alpha on non-informative sequential tests.

Practicability and costs

Data updating in distributed data networks (e.g., Mini-Sentinel) will likely occur on a regular calendar-time-based schedule (e.g., quarterly), which is logistically most feasible. While this may be suboptimal for near real time monitoring of frequently occurring events, it may be sufficient for rare outcomes. It may be possible that sequential tests would occur even less frequently (e.g., twice yearly) than the data-updating schedule. Even within single databases, defining window length by other times scales (e.g., X number of exposed patients or Y number of outcomes) can be complex and requires continuous background processing (e.g., continuous new user and eligibility identification, outcome follow-up only among eligible patients, etc). Generally, more frequent monitoring will require more programming hours and computing time and will require more person-hours to review outputs.

Alerting time

The robustness of the sequential monitoring design and the practicability and costs of monitoring must be balanced with the issue that longer windows can delay identification of safety issues. This raises

policy questions about how quickly safety information on a product is needed and about the potential public health burden of waiting to analyze accruing data when a product is truly causing harm.

VI. FUTURE

As the object of the Taxonomy project is to promote efficiency and transparency in methodological decision-making within an active medical product surveillance system, the Year 1 and Year 2 reports may be useful in assisting Mini-Sentinel investigators in developing and implementing future signal refinement protocols. Meanwhile, a number of Mini-Sentinel Workgroups continue to develop, test, and refine methods for medical product monitoring and these methods should be incorporated into the Taxonomy framework as appropriate.

In future activities, the Taxonomy Workgroup is considering applying the framework to a large number of examples, including application to completed and ongoing Mini-Sentinel assessments, with the objectives of: (1) evaluating the alignment between the Taxonomy framework and what has been or is being done in actual assessments; and (2) identifying which methodological approaches (i.e., which combinations of methods) will likely be most frequently applied in signal refinement activities. Findings from the latter could be used to help prioritize the development of modular programs. The Workgroup will also consider the development of a Taxonomy interface to facilitate the use of the Taxonomy framework in future Mini-Sentinel assessments.

A. TAXONOMY'S CURRENT ROLE IN ONGOING SURVEILLANCE ACTIVITIES (E.G., DABIGATRAN)

Several Taxonomy group members are participating in the Workgroup to develop an active monitoring protocol for dabigatran, a new anticoagulant. Below, we describe the scenario characteristics using the Taxonomy framework. Based on these characteristics, we were able to quickly eliminate several methods options. For example, within-person confounding will need to be addressed for most (if not all) outcomes. For this reason, and because a meaningful active comparator exists, a between-person comparison is most appropriate. Also, because of the large number of potential confounders and the multiple outcomes of interest, a propensity score should be used.

Table 11. Example 8: Dabigatran and multiple outcomes (severe upper GI bleed, hemorrhagic stroke, ischemic stroke)	
Characteristics determined by stakeholders/investigators	
Effect measure(s) of interest	Both difference and ratio measures
Comparator(s)	Warfarin
Exposure characteristics	
Background frequency of use:	More frequent
Utilization trend in population:	Increasing
Use pattern	Long-term
Characteristics of the potential exposure-HOI link	
Onset of exposure risk window:	Immediate for all outcomes
Duration of exposure risk window:	Long for all outcomes
Strength of confounding	
Between-person	Needs to be addressed for all
Within-person	Needs to be addressed for all
HOI Characteristics	
Background frequency	Infrequent for ischemic stroke and rare for all other outcomes
Periodicity	Assume we are interested in the first of each event following initiation (but patients may have had events prior to treatment)
Expected degree of onset misclassification	Negligible for all

B. LINK TO PRISM 'TAXONOMY'

The Taxonomy project is intended to provide a framework for conceptualizing and guiding methods selection for specific types of monitoring scenarios. Certain exposure types might require more nuanced considerations beyond the scope of this report. The Post-Licensure Rapid Immunization Safety Monitoring (PRISM) program, originally launched by the federal government to monitor 2009 H1N1 vaccine safety, is now a component of Mini-Sentinel and is intended to provide routine near real-time active vaccine surveillance within the Mini-Sentinel infrastructure. A current PRISM task order involves, in part, describing methodological considerations when evaluating vaccine outcome pairs and the advantages and limitations of immunization safety monitoring using the PRISM framework. Other types

of exposures that could benefit from more specific methodological attention in projects like PRISM include medical devices and biologics.

C. METHODS GAPS

An important objective of the Taxonomy Workgroup is to continually identify areas in which additional methodological work is needed for an active safety surveillance system. In addition to the methods gaps acknowledged in the Year 1 Taxonomy report, we have identified the following issues that warrant further consideration:

(1) How many variables can go into a propensity score model relative to the number of exposed patients?
(2) When is a confounder summary score model too predictive (e.g., is there a propensity score model c-statistic that is too high)?
(3) What are the best approaches for applying confounder summary scores in distributed data settings?
(4) How can confounder summary scores (i.e., PSs and DRSs) be combined to leverage the advantages of both in the early marketing period?

VII. ACKNOWLEDGEMENTS

We thank many individuals who have contributed to this work, including Malcolm Maclure from the University of British Columbia, Vancouver, BC Canada, Meghan Baker from Harvard Pilgrim Healthcare Institute and Department of Population Medicine, Harvard Medical School, Boston, MA, as well as many from FDA who attended an in-person Taxonomy meeting at the 27th International Conference on Pharmacoepidemiology & Therapeutic Risk Management on August 15th, 2011 in Chicago, IL, including: Tarek Hammad, Moni Houstoun, Danica Marinac-Dabic, Jerald Mullersman, Michael Nguyen, Judy Racoosin, Mary Beth Ritchey, Judy Staffa, and Gwen Zornberg.

VIII. REFERENCES

1. Platt R, Carnahan RM, Brown JS, et al. The U.S. Food and Drug Administration's Mini-Sentinel program: status and direction. Pharmacoepidemiol Drug Saf 2012;21 Suppl 1:1-8.

2. Gagne JJ, Walker AM, Glynn RJ, Rassen JA, Schneeweiss S. An Event-Based Approach for Comparing the Performance of Methods for Prospective Medical Product Monitoring. Pharmacoepidemiol Drug Saf 2012.

3. Gagne JJ, Rassen JA, Walker AM, Glynn RJ, Schneeweiss S. Active Safety Monitoring of New Medical Products Using Electronic Healthcare Data: Selecting Alerting Rules. Epidemiology 2012.

4. Gagne JJ, Fireman B, Ryan PB, et al. Design considerations in an active medical product safety monitoring system. Pharmacoepidemiol Drug Saf 2012;21 Suppl 1:32-40.

5. Fireman B, Toh S, Butler MG, et al. A protocol for active surveillance of acute myocardial infarction in association with the use of a new antidiabetic pharmaceutical agent. Pharmacoepidemiol Drug Saf 2012;21 Suppl 1:282-90.

6. Walker AM. Signal detection for vaccine side effects that have not been specified in advance. Pharmacoepidemiol Drug Saf 2010;19:311-7.

7. Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. Pharmacoepidemiology and Drug Safety 2010;19:858-68.

8. Nelson JC, Cook AJ, Yu O, et al. Challenges in the design and analysis of sequentially monitored postmarket safety surveillance evaluations using electronic observational health care data. Pharmacoepidemiol Drug Saf 2012;21 Suppl 1:62-71.

9. Maclure M, Fireman B, Nelson JC, et al. When should case-only designs be used for safety monitoring of medical products? Pharmacoepidemiol Drug Saf 2012;21 Suppl 1:50-61.

10. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. Am J Epidemiol 1991;133:144-53.

11. Maclure M. 'Why me?' versus 'why now?'--differences between operational hypotheses in case-control versus case-crossover studies. Pharmacoepidemiol Drug Saf 2007;16:850-3.

12. Suissa S. The case-time-control design. Epidemiology 1995;6:248-53.

13. Wang S, Linkletter C, Maclure M, et al. Future cases as present controls to adjust for exposure trend bias in case-only studies. Epidemiology 2011;22:568-74.

14. Farrington CP. Relative incidence estimation from case series for vaccine safety evaluation. Biometrics 1995;51:228-35.

15. Suissa S. The case-time-control design: further assumptions and conditions. Epidemiology 1998;9:441-5.

16. Greenland S. Confounding and exposure trends in case-crossover and case-time-control designs. Epidemiology 1996;7:231-9.

17. Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: the self-controlled case series method. Stat Med 2006;25:1768-97.

18. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology 2009;20:512-22.

19. Arbogast PG, Ray WA. Performance of disease risk scores, propensity scores, and traditional multivariable outcome regression in the presence of multiple confounders. Am J Epidemiol 2011;174:613-20.

20. Rassen JA, Avorn J, Schneeweiss S. Multivariate-adjusted pharmacoepidemiologic analyses of confidential information pooled from multiple health care utilization databases. Pharmacoepidemiol Drug Saf 2010;19:848-57.

21. 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:S83-9.

22. Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. Am J Epidemiol 2003;158:280-7.

23. Sturmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. J Clin Epidemiol 2006;59:437-47.

24. Austin PC, Mamdani MM, Stukel TA, Anderson GM, Tu JV. The use of the propensity score for estimating treatment effects: administrative versus clinical data. Stat Med 2005;24:1563-78.

25. Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. Stat Med 2008;27:2037-49.

26. Cadarette SM, Gagne JJ, Solomon DH, Katz JN, Sturmer T. Confounder summary scores when comparing the effects of multiple drug exposures. Pharmacoepidemiol Drug Saf 2010;19:2-9.

27. Hansen BB. The prognostic analogue of the propensity score. Biometrika 2008;95:8.

28. Glynn RJ, Gagne JJ, Schneeweiss S. Role of disease risk scores in comparative effectiveness research with emerging therapies. Pharmacoepidemiol Drug Saf 2012;In press.

29. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. Am J Epidemiol 2006;163:1149-56.

30. Pearl J. On a class of bias-amplifying variables that endanger effect estimates. In: Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence (UAI 2010). Corvallis, OR: Association for Uncertainty in Artificial Intelligence; 2010:425-32.

31. Schneeweiss S, Gagne JJ, Glynn RJ, Ruhl M, Rassen JA. Assessing the comparative effectiveness of newly marketed medications: methodological challenges and implications for drug development. Clin Pharmacol Ther 2011;90:777-90.

32. Myers JA, Rassen JA, Gagne JJ, et al. Effects of adjusting for instrumental variables on bias and precision of effect estimates. Am J Epidemiol 2011;174:1213-22.

33. Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol 2008;168:656-64.

34. Kurth T, Walker AM, Glynn RJ, et al. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. Am J Epidemiol 2006;163:262-70.

35. Platt RW, Delaney JA, Suissa S. The positivity assumption and marginal structural models: the example of warfarin use and risk of bleeding. Eur J Epidemiol 2011.

36. Darabi K, Abdel-Wahab O, Dzik WH. Current usage of intravenous immune globulin and the rationale behind it: the Massachusetts General Hospital data and a review of the literature. Transfusion 2006;46:741-53.

37. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. Am J Epidemiol 2003;158:915-20.

38. Weinberg CR. Toward a clearer definition of confounding. Am J Epidemiol 1993;137:1-8.

39. Cole SR, Hernan MA. Fallibility in estimating direct effects. Int J Epidemiol 2002;31:163-5.

40. Robins J. The control of confounding by intermediate variables. Stat Med 1989;8:679-701.

41. Carnahan RM. Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative data: summary of findings and suggestions for future research. Pharmacoepidemiol Drug Saf 2012;21 Suppl 1:90-9.

42. Carnahan RM, Moores KG. Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative and claims data: methods and lessons learned. Pharmacoepidemiol Drug Saf 2012;21 Suppl 1:82-9.

43. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology 2010;21:383-8.

44. Dormuth CR, Patrick AR, Shrank WH, et al. Statin adherence and risk of accidents: a cautionary tale. Circulation 2009;119:2051-7.

IX. GLOSSARY

***Abrupt onset**:* refers to a sudden occurrence of a health outcome of interest with a sharp contrast between absence and incidence such that the time of the event is easy to define and ascertain (e.g., mortality).

Background frequency: refers to the frequency with which a phenomenon occurs in the general population. For example, *background frequency of use in the population* refers to the commonness of use of a particular medical product among all patients in the base population (e.g., in the Mini-Sentinel Distributed Database). Background frequency as a health outcome of interest characteristic refers to the incidence (or prevalence) of the event in the entire population.

Design choices: are defined as constraints on observation time intended to yield the most valid comparisons. For example, constraints may be used to restrict signal refinement to a particular patient population defined by a certain age range or by presence of a specific underlying medical condition. Constraints could also be used to define a period of observation time to serve as a basis for comparison, such as the identification of a comparator group (through matching or restriction, for example) or ascertainment of an alternate observation period in a patient's history.

Difference measures of association: quantify the relation between a medical product and a health outcome of interest on an absolute scale. Examples include incidence rate differences and cumulative incidence differences (a.k.a., risk differences).

Exposure time trend: refers to the trend in use of a medical product in the general population. For example, shortly after market authorization use of new drugs in a population increases rapidly. Use of a medical product can also be relatively stable, decreasing, or cyclical, such as occurs with antibiotic utilization, which mirrors seasonality of common infections.

Incident users (or new users): are patients who become exposed to a medical product for the first time. Operational definitions typically require an observed "washout" period preceding medical product initiation to ensure that exposure is indeed new use at least for the minimum washout time. It can be

difficult to ascertain true first use for patients since investigators cannot observe patients' histories before they enter a database. Therefore, a washout period should be long enough so that exposures preceding the washout period cannot have a biological effect on the outcome of interest.

***Insidious onset**:* refers to an outcome that develops gradually and for which the definition of onset time may be ambiguous (e.g., incidence of multiple sclerosis).

***Monitoring scenarios**:* refer to pairs of pre-specified medical products and health outcomes of interest categorized by a unique constellation of exposure characteristics, outcome characteristics, and characteristics of the links between them.

Prevalent users: are patients who have continued exposure to a medical product of interest. Generally, prevalent users refer to patients who enter a database with immediate recorded exposure to the medical product of interest such that they do not achieve the requisite washout period to qualify as a new user.

Relative measures of association: quantify the relation between a medical product and a health outcome of interest on a ratio scale. Examples include incidence rate ratios, cumulative incidence ratios (a.k.a., risk ratios), and odds ratios.

Sequential monitoring: refers to repeated assessments of prospectively accruing data (as compared to a retrospective one-time assessment after all data are accrued). A *sequential test* involves repeated tests of a statistical hypothesis as the data accrue.

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 health outcome of interest.

Signal generation:* includes a collection of methods for identifying potential associations between medical products and health outcomes of interest.

***Signaling methods**:* imply analytic approaches used to determine when sufficient evidence – beyond chance – exists, indicating a product-health outcome of interest association requiring further attention (e.g., a test statistic or a decision rule).

Signal refinement:* is an epidemiological process for evaluating the magnitude and clinical significance of a suspected association.

***Transient exposure**:* is defined as an exposure lasting only for a short time. It is important to note that this does not preclude subsequent periods of exposure (for example, each as-needed dose of an anti-inflammatory drug would be considered a transient exposure if the health outcome of interest pertained to the initiation of the medication).

*The terms "signal evaluation," "signal generation," and "signal refinement" are suggested and defined by FDA.