

RE-USE OF MINI-SENTINEL DATA FOLLOWING RAPID ASSESSMENTS OF POTENTIAL SAFETY SIGNALS USING CUSTOMIZABLE MODULAR PROGRAMS

Prepared by: The Mini-Sentinel Data Re-use Committee

November 30, 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.

Re-Use Of Mini-Sentinel Data Following Rapid Assessments Of Potential Safety Signals Using Customizable Modular Programs

Table of Contents

I. EXECUTIVE SUMMARY	- 1 -
II. THE MINI-SENTINEL PROGRAM	- 2 -
III. RAPID ASSESSMENT VIA CUSTOMIZABLE MODULAR PROGRAMS IN MINI-SENTINEL.....	- 2 -
IV. IMPLICATIONS OF RAPID ASSESSMENTS ON SUBSEQUENT ANALYSES OF THE SAME PRODUCT-OUTCOME PAIRS USING THE SAME DATA	- 3 -
V. CHARGE TO THE COMMITTEE.....	- 3 -
VI. DELIBERATION PROCESS	- 3 -
VII. SUMMARY OF DELIBERATIONS	- 4 -
A. KEY CONCEPTUAL FRAMEWORK: FALSE POSITIVE FINDINGS VERSUS FALSE NEGATIVE FINDINGS	- 4 -
B. ANALYSIS FOR WHICH RAPID ASSESSMENTS USING MODULAR PROGRAMS ARE APPROPRIATE	- 4 -
1. <i>Signal generation</i>	- 4 -
2. <i>Signal refinement</i>	- 4 -
3. <i>Signal evaluation</i>	- 5 -
C. DETERMINING A PRIORI CATEGORY OF MODULAR PROGRAM ANALYSIS AND ASSOCIATED FOLLOW-UP PLANS ..	- 5 -
D. SIGNAL GENERATION (ANALYSIS WITH NO PRIOR)	- 6 -
E. SIGNAL REFINEMENT (ANALYSIS WITH A WEAK OR MODERATE PRIOR).....	- 6 -
F. SIGNAL EVALUATION (ANALYSIS WITH A STRONG PRIOR)	- 8 -
G. RELEASING RESULTS FROM RAPID QUERIES USING MODULAR PROGRAMS	- 8 -
H. ANALYSIS FOR WHICH RAPID ASSESSMENTS USING MODULAR PROGRAMS ARE NOT SUITABLE	- 8 -
I. REPLICATION IN NON MINI-SENTINEL DATA SOURCE.....	- 9 -
J. WORKED EXAMPLES.....	- 9 -
1. <i>Olmесartan and celiac disease</i>	- 9 -
2. <i>Dabigatran and hemorrhage and ulcer</i>	- 10 -
VIII. CONCLUSION	- 10 -
IX. ACKNOWLEDGEMENTS	- 12 -
X. TABLES AND FIGURES	- 12 -
XI. REFERENCES	- 14 -

I. EXECUTIVE SUMMARY

The U.S. Food and Drug Administration (FDA)'s Mini-Sentinel program uses pre-written and pre-tested modular analytic programs that allow comparisons of event rates between different exposure groups to perform rapid assessments of suspected safety signals within a distributed system of electronic healthcare databases. The Agency asked Mini-Sentinel's Safety Science Committee, which serves as the senior expert panel for the scientific activities of the pilot, to convene a committee to assess the implications of such rapid assessments on subsequent analyses of the same product-outcome pair using the same data. The non-binding recommendations of the committee to FDA summarized in this report are based on a distinction between three broad categories of activities envisioned for Mini-Sentinel, according to the strength of the knowledge of the suspected association before running the analysis ("prior"): 1) signal generation (an analysis with *no* prior), 2) signal refinement (an analysis with a *weak* or *moderate* prior), and 3) signal evaluation (an analysis with a *strong* prior). The committee believes that modular programs are most useful for signal refinement activities.

The first recommendation is that before any modular program analysis begins the FDA must determine under which of these categories the analysis falls, i.e., the planned use of the Mini-Sentinel data, and a planned course of actions based upon the results.

The second recommendation is that activities 1 and 2 may use a split-sample approach, with different datasets used for derivation and replication or refinement of the signal; the process should then not be considered complete until after the analysis of the second dataset is finished. Some committee members, however, disagree with this recommendation and instead suggest alternative approaches that use all available data in one analysis, with additional adjustment for the signaling threshold or use of an empirical Bayesian approach to minimize false positives, or with reporting of measures of association and variability (e.g., relative risks and confidence intervals) and leaving the interpretations to FDA or other stakeholders. The entire committee agrees that with strong priors (activity 3), a protocol-based assessment using all available data and with robust approaches to adjust for systematic errors (i.e., confounding, selection bias, information bias) should be used; in that situation rapid modular program analyses may be worthwhile only uncommonly and under certain limited circumstances.

The third recommendation is that results should always be disclosed when the set of analyses planned a priori is completed, even if additional analyses will then be forthcoming.

The final recommendation is that rapid modular program assessments of possible product-outcome associations should not be performed: 1) if the goal is just study size determination, 2) if systematic errors will predictably be too severe to provide actionable information; or 3) if the precision will predictably be too low for the analysis to provide useful information.

Under this general framework, some members of the committee believe that the same Mini-Sentinel data analyzed rapidly by modular programs may be re-used to refine suspected signals if the goals of the subsequent analyses are to reduce further systematic errors or to investigate further the signal (e.g., examine the signal in specific subgroups). If a full protocol-based assessment follows a modular program analysis using the same data, this second assessment should *not* be interpreted as independent

confirmation of the association, such as would be established via replication of the same product-outcome association in two different populations. Instead, the protocol-based assessment should be interpreted as an analysis that has reduced insofar as possible systematic errors that may have been present or residual in the original modular program analysis.

II. THE MINI-SENTINEL PROGRAM

The Mini-Sentinel program (<http://www.mini-sentinel.org>) is part of the Sentinel Initiative, an effort by the U.S. Food and Drug Administration (FDA) to develop a national system for monitoring the safety of medical products as mandated by the FDA Amendments Act of 2007.¹⁻³ Mini-Sentinel currently focuses on signal refinement, a term used to describe initial, ideally rapid, assessment of the magnitude of suspected associations between selected medical products and specific adverse health outcomes.⁴ Specific product-outcome pairs are chosen by FDA, based on knowledge of the product, its class, its intended use, its public health impact, and reported experience with it.

FDA distinguishes between signal (hypothesis) generation, signal refinement, and signal evaluation.⁴ *Signal generation* uses statistical methods to identify potential safety signals among non pre-specified medical product–adverse outcome pairs. *Signal refinement* assesses an identified potential safety signal to determine more clearly whether evidence exists to support a product-outcome association. *Signal evaluation* attempts to investigate the causal relationship between medical products and adverse outcomes through full epidemiological analysis.

The boundary between signal refinement and signal evaluation is not always well demarcated, as both often use similar epidemiological and statistical methods. However, signal refinement is intended as a set of usually rapid analyses to estimate more precisely previously suggested signals, while signal evaluation is intended as a more definitive assessment of a previously noted association. As will be seen, the committee’s recommendations for the methods to be used in each of these settings will be different. Mini-Sentinel does not currently perform data mining or other signal generation activities as a standard practice. However, such activities are part of its Congressional mandate,¹ its Post-Licensure Rapid Immunization Safety Monitoring (PRISM) program is exploring this capability,⁵ and ultimately Mini-Sentinel hopes to be able to conduct activities that would encompass the full spectrum of surveillance, which will encompass signal generation, signal refinement, and signal evaluation.

III. RAPID ASSESSMENT VIA CUSTOMIZABLE MODULAR PROGRAMS IN MINI-SENTINEL

A goal of Mini-Sentinel is to provide FDA the ability to conduct near real-time safety assessments using its distributed data system, which as of December 2011 include data on over 125 million lives and nearly 350 million person-years of observation time in its Distributed Database. A valuable tool that has been created to realize this capability is the use of customizable modular analytic programs that allow for definitions of cohorts of new users of medical products of interest, their exposures, outcomes, date ranges, age ranges, and other attributes.⁶ These modular programs allow rapid response because the analytic codes have been written, tested, and documented, and approved by all data partners.

The modular programs currently permit confounding adjustment through stratification by age, sex, and health plan, but do not generate effect estimates that are adjusted for other confounders. Additional modular programs are being developed that will allow more sophisticated “routine” adjustment for measured confounders, using methods such as matching treated and untreated patients by propensity score.

Despite these planned improvements, rapid queries via modular programs are not designed to replace formal protocol-based assessments of suspected signals. The main difference between these two types of analyses is that modular program assessments rely on “off-the-shelf” but customizable analytic tools to obtain safety information rapidly, while protocol-based assessments are tailored to specific product-outcome pairs and typically include more sophisticated approaches to adjust for confounding, as well as approaches to evaluate outcome misclassification. It is expected that signal evaluation in Mini-Sentinel always be guided by a formal protocol, while signal refinement may or may not be protocol-based.

IV. IMPLICATIONS OF RAPID ASSESSMENTS ON SUBSEQUENT ANALYSES OF THE SAME PRODUCT-OUTCOME PAIRS USING THE SAME DATA

Mini-Sentinel currently uses its entire Distributed Database for rapid assessments to maximize timeliness and statistical information. However, if and when FDA decides to conduct a formal evaluation of causality for a signal through “conventional” protocol-based epidemiological assessments, it may have to rely on the same data partners that participate in Mini-Sentinel because in the U.S. there are few alternatives that are not part of Mini-Sentinel. Thus, the Agency asked the Mini-Sentinel Safety Science Committee, which serves as the senior expert panel for the scientific activities of Mini-Sentinel,⁷ to convene a committee with experts in epidemiology, biostatistics, surveillance, and other relevant expertise (Table) to provide advice on the implications of using modular programs for rapid assessments of suspected safety signals on subsequent analyses of the same product-outcome pair using the same data.

V. CHARGE TO THE COMMITTEE

The expert committee, led by the Safety Science Committee, was charged with deliberating on the issues identified above. The committee members focused solely on Mini-Sentinel modular programs that allow for comparison of adverse event rates between different exposure groups.⁸ The committee was not charged with discussing other Mini-Sentinel activities, such as protocol-based, one-time or sequential safety assessments of approved medical products,⁹⁻¹² or evaluations of the effect of FDA regulatory actions.¹³ It was agreed at the outset by the committee and FDA that the committee members did not have to reach consensus on any of the issues or offer any recommendations to FDA. However, if and when there was an agreement, the committee might offer general, non-binding recommendations to the Agency. FDA is not required to follow these recommendations.

VI. DELIBERATION PROCESS

The committee deliberated over a three-month period from December 2011 to February 2012 via teleconferences. Additional discussions occurred through email correspondence, during that time and

thereafter. All meetings were recorded; all materials including audio recordings and meeting summaries were shared through a secure WebOffice. Several FDA scientists, identified by the Agency, were invited to participate as observers. The committee reached all decisions described here by consensus, except where it is specified otherwise.

VII. SUMMARY OF DELIBERATIONS

This section summarizes the issues discussed during the deliberations process. Where appropriate, recommendations are provided. Many of the recommendations are meant to serve as general principles; they require further deliberations to be refined or operationalized.

A. KEY CONCEPTUAL FRAMEWORK: FALSE POSITIVE FINDINGS VERSUS FALSE NEGATIVE FINDINGS

The committee discussed what could give rise to a safety signal in a rapid assessment. An apparent safety signal may arise from a real causal relationship, but also from random errors or systematic errors, the latter including selection bias, information bias, and confounding.¹⁴ In an approach that the committee does not encourage, assessment results may conventionally be dichotomized into “positive signal” or “no signal”. This approach obscures quantitative assessment and degrades the information in the data.¹⁵ Those who practice this approach refer to two types of error, type I errors (false positive findings) and type II errors (false negative findings), and commonly place greater focus on type I errors, thus leading to concern about adjustment for multiple comparisons. In addition to reducing a quantitative exercise to a qualitative one, this approach is problematic because it ignores the prior likelihood of type I and type II errors. The prevalence of type I versus type II errors depends on the prevalence of false versus true null hypotheses in a body of data. A rational assessment would consider not only these prior probabilities but also their relative costs before applying blanket decision rules.

B. ANALYSIS FOR WHICH RAPID ASSESSMENTS USING MODULAR PROGRAMS ARE APPROPRIATE

The three broad categories of activities envisioned in Mini-Sentinel are characterized by what is already known about the strength of the product-outcome association of interest before running the analysis (“prior hypothesis” or “prior”), which then dictates the primary goal of the analysis (Figure).

1. Signal generation

Signal generation is an analysis in the *absence of a prior* hypothesis. As signal generation in Mini-Sentinel is largely an uncharted territory, the applicability of modular programs for such activities is unknown. However, the committee believes that the primary goal of signal generation – should it be performed in Mini-Sentinel rapidly using modular programs or by others outside of Mini-Sentinel using the same datasets – will be to reduce false positives while preserving nearly all important signals in the data.

2. Signal refinement

Signal refinement refers to an analysis with a *weak or moderate prior*. The committee believes that signal refinement is what modular programs are most useful for in Mini-Sentinel – to rapidly refine signals

identified from other sources. The committee expects that this activity will make up the majority of the modular program analyses. The primary goal of the analysis is to reduce false positives without reducing greatly the ability to detect important signals, through some control of confounding.

3. Signal evaluation

Signal evaluation is an analysis with a *strong prior*. The committee recommends that a protocol-based assessment be used to investigate signals with strong priors. As noted above, modular programs – even though they are highly customizable – are not designed to adjust fully for systematic errors for every specific product-outcome pair. This will likely remain true even with the incorporation of more sophisticated methods for confounding adjustment and more valid algorithms for identifying exposure, outcome, or other cohort characteristics in future versions. In the context of signal evaluation, rigorous control of systematic errors is critical to yield valid and precise effect estimates when testing a specific hypothesis that emerged as a strong prior. Therefore, a protocol-based assessment is strongly preferred. However, more rapid and less rigorous queries may be worthwhile under certain limited circumstances, such as when there is a public-health emergency and FDA needs to obtain some useful information quickly for urgent regulatory decisions.

The committee does not provide functional definitions of the strength of the prior, but recommends that it be judged by FDA based on evidence available at the time the analysis starts from other sources, such as pre-market randomized controlled trials, spontaneous adverse event reports, post-market randomized and observational studies, and biological plausibility. For example, an analysis of a medical product-outcome pair would be classified as a *signal generation* activity if neither existing data nor the known biological mechanism of the product predicts such an association. If there are a few anecdotal reports of uncertain quality or any other weak hint (e.g., unproven mechanism or reasoning by analogy), then the analysis may be considered a *signal refinement* activity with a *weak prior*. If there are multiple spontaneous reports of the association, and perhaps also some observational studies that are judged not to be sufficient or some biological reasoning that is not yet supported by firm data, then the analysis may be categorized as a *signal refinement* analysis with a *moderate prior*. If the association is observed in pre-market randomized controlled trials or post-market observational studies of sufficient quality, then the analysis may be considered a *signal evaluation* activity with a *strong prior*.

C. DETERMINING A PRIORI CATEGORY OF MODULAR PROGRAM ANALYSIS AND ASSOCIATED FOLLOW-UP PLANS

Before running a modular program analysis, FDA should determine the purpose of the analysis, the category under which the analysis falls, and develop plans for associated follow-up plans based on how much the results strengthen or weaken the suspected association. As described in greater detail in the sections below, follow-up plans depend on the “posterior” strength of the suspected signal, and therefore may differ for positive vs. null results. The committee’s recommendations about follow-up plans are intended to serve as general guidelines. They take into account, but do not quantify, the posterior probability that these suspected signals are true to help FDA prioritize its activities and allocate its resources. The committee does not think the decisions should be automated. FDA will still always need to apply regulatory judgment on a case-by-case basis. The committee is simply suggesting that the potential regulatory avenues be decided upon a priori, as this will dictate which type of analysis is then performed. Thus, if FDA recognizes a priori that even if an association is found (or not found) in an analysis using a modular program, it would not be sufficiently convinced of the results to act on them (e.g., because of the

residual risk of systematic errors), then embarking on a modular program analysis is not wise; rather, other options like a full protocol-based assessment may make more sense. To guide in these judgments, FDA is encouraged to seek input from FDA and non-FDA scientists, as it finds useful. The committee's recommendations emerge from both frequentist and Bayesian perspectives;^{16,17} FDA may therefore want to consider including both when interpreting the modular program results and planning for the next steps.

D. SIGNAL GENERATION (ANALYSIS WITH NO PRIOR)

Signal generation usually involves a large number of non pre-specified product-outcome pairs. The committee feels that signal generation using electronic healthcare databases is largely an uncharted territory. As Mini-Sentinel does not currently perform signal generation on a routine basis, as part of its contractual scope of work, the committee is unclear about what form it will take. For these reasons, the committee only briefly deliberated on issues related to rapid signal generation using modular programs. However, should FDA wish to use modular programs to generate signals quickly in Mini-Sentinel, the Agency may employ approaches recommended for signal refinement, discussed in greater detail in the next section.

E. SIGNAL REFINEMENT (ANALYSIS WITH A WEAK OR MODERATE PRIOR)

When modular programs are used to refine a weak or moderate signal, most committee members recommend using a split-sample approach.¹⁸ The approach will partition the Mini-Sentinel Distributed Database into two samples. To reduce false positives, an analysis is first performed in one sample and then the same analysis is repeated in the second sample to reduce chance findings. The process should not be considered complete until after the analysis of the second sample is finished. Interpretation of the findings should consider results from both samples together although the two samples are never aggregated analytically.

Further, the committee believes that the split-sample approach could also reduce systematic errors arising from measured confounders in signal refinement activities, if a more robust approach is used to analyze the second sample than the first. For example, the first split sample may adjust only for age, sex, and health plan, but the second sample can further adjust for other confounders using more sophisticated approaches (e.g., matching on or stratification by propensity score¹⁹ or disease risk score²⁰), or when appropriate, use other techniques like an instrumental variable analysis or natural experiment designs (e.g., comparison between institutions with different policies). Of course, this may become less relevant if routine, sophisticated confounding adjustment become available in modular programs and can be used in both samples. In both cases, the process is only considered complete after the analysis of the second split sample is finished.

Should the split-sample approach be used, the committee discussed possible ways to partition the data. One way is to have each data partner split their data randomly into two datasets of roughly equal size; the two samples will be similar in patient characteristics and size. Although this approach offers some protection against false positives, it also reduces statistical efficiency.^{21,22} The committee recommends non-random splitting of data, more specifically, splitting by data partner and including sites with additional variables (e.g., laboratory test results, vital signs) in the second sample. Although such splitting is also susceptible to the loss of statistical power, it offers some advantages. First, it is logistically easier. More important, it can increase the internal validity of the analysis if the association is observed in two

different samples, derived from institutions with different patient characteristics and practice.²¹ Validity can be further improved if the analysis of the second sample adjusts for more confounders as afforded by sites with richer data. Theoretically, it might be worthwhile to have the second analysis done by a different group of investigators; however, the committee does not recommend this as a feasible routine approach because this may lead to delay in timeliness and additional logistical challenges.

Follow-up plans may differ for positive vs. null results, and may further vary by the strength of the prior and the strength of the positive results. For example, if a modular program analysis of a weak prior produces credible evidence of a positive association, FDA may consider moving forward quickly with a full protocol-based assessment before making other regulatory decisions. Yet, if a modular program analysis of a weak prior produces null results, FDA may bring the issue to closure. Of course, new analyses can still be initiated if more information about the potential signal later arises through other existing channels. In contrast, if a modular program of a moderate prior produces credible evidence of a positive association, FDA may consider certain regulatory actions (e.g., changing a label). If it produces only weak or no evidence of an association, FDA may choose instead to follow the product-outcome association through other existing channels (e.g., ongoing randomized controlled trials or its spontaneous adverse event reporting systems) and gather more data.

Questions arose about whether any of the Mini-Sentinel data can be re-used for subsequent assessments of the same product-outcome pairs once they are analyzed by the split-sample approach. The committee generally agreed that the answer is yes, but only if the goal of the subsequent analysis is to employ more robust methods to reduce systematic errors (e.g., adjusting for more confounders, reducing outcome misclassification through chart review) in one or both of the initial split-sample analyses. It is important to emphasize that the split-sample approach using modular programs is always interpreted as a single process, even though the analysis involves two datasets. The final interpretation of the results remain the same, i.e., as an association that persisted despite best efforts to reduce biases in an expedited way. If a full protocol-based assessment focused on signal evaluation follows a modular program analysis used for signal refinement, it can use the full Mini-Sentinel dataset. However, this second assessment should *not* be interpreted as independent confirmation of the association, such as would be established via replication of the same product-outcome association in two different populations. Instead, the protocol-based assessment should be interpreted as an analysis that has reduced insofar as possible systematic errors that may have been present or residual in the original modular program analysis. Although it does not involve new data, it is nevertheless new information about the data already in hand.

Some committee members noted that from a frequentist perspective, one could perform an analysis with all Mini-Sentinel data using a modified threshold for statistical significance, e.g., $(0.05)^2=0.0025$. This would be an acceptable approach to preserving statistical efficiency and maintaining the same degree of protection against false positives,^{21,22} and indeed some members of the committee preferred this approach. Others preferred not to rely on statistical significance testing. Regardless, one disadvantage of this approach is the potential political risk it generates for FDA if, for example, it does not act on a safety problem with a p-value of 0.01. Statistically, however, it is a completely acceptable alternative approach.

Another alternative using all Mini-Sentinel data is an empirical Bayesian approach to shrink the estimates toward the null.^{21,22} This approach was preferred by some committee members.

A third alternative approach simply calculates measures of association and variability (e.g., relative risks and confidence intervals) using all Mini-Sentinel data, leaving the interpretation and use of that information to FDA or other stakeholders, in the context of all other information they have at the time about the possible association. However, its relatively subjective nature, i.e., absence of clear framework for decision-making for FDA or other stakeholders to follow, may make some uncomfortable.

It is likely that these alternative approaches will evolve over time, and be strengthened by experiences in other fields (e.g., genomics). It could be envisaged that when laying down the initial plans for a signal refinement, FDA investigators may choose to use one of these alternative approaches. Regardless, it is recommended that the choice of approach to be used be made a priori.

F. SIGNAL EVALUATION (ANALYSIS WITH A STRONG PRIOR)

If existing evidence from external sources suggested a strong prior, i.e., a possibly true relationship between a medical product and a health outcome of interest, the committee recommends that a protocol-based assessment be used to investigate the association more rigorously.

The committee believes rapid signal evaluation using modular approaches may be worthwhile only uncommonly and under certain limited circumstances. The rapid analysis should and can be done using the entire Mini-Sentinel Distributed Database, if and only if: 1) the goal is to obtain some useful information quickly for urgent regulatory decisions, 2) an a priori determination can be made that a modular program analysis could be sufficient to enable such decisions, and 3) the decision to initiate or continue with the more rigorous protocol-based assessment does not depend on the results of the rapid analysis. In this case, the modular program analysis can be considered the first stage in a more robust analysis, to respond to or identify a regulatory emergency. The Mini-Sentinel data will then be re-analyzed subsequently in the more rigorous protocol-based assessment, employing more sophisticated approaches to reduce systematic errors (e.g., adjusting for more confounders, validating the outcome events through chart review), or assess specific subgroups (e.g., female) or specific characteristics of the outcome (e.g., specific endpoints in a composite outcome).

G. RELEASING RESULTS FROM RAPID QUERIES USING MODULAR PROGRAMS

The committee recommends that results of all Mini-Sentinel analyses, modular program or protocol driven, and regardless of the strength of the prior, *always* be released publicly, along with appropriate caveats and limitations, when the analysis is considered complete by FDA. If the FDA would be uncomfortable releasing the results of a rapid analysis when complete, they should not embark on it in the first place, and perhaps should instead do a full protocol-based analysis. This recommendation is consistent with the current Mini-Sentinel policies: the results of completed protocol-based assessments will be published on the Mini-Sentinel public website upon acceptance of the final report by the FDA, while the results of modular program or summary table assessments will be published within 60 days after the end of the calendar quarter in which the query is considered completed, unless the Agency determines that delaying posting of results is in the public interest.²³

H. ANALYSIS FOR WHICH RAPID ASSESSMENTS USING MODULAR PROGRAMS ARE NOT SUITABLE

The committee recommends against using modular programs that allow comparisons of event rates between different exposure groups in the following three situations. First, these programs should not be

used solely to obtain feasibility counts purely for study size determination; the needed feasibility information can be obtained without making a comparison between groups.

Second, these programs should not be used when systematic errors (e.g., uncontrolled or residual confounding, outcome misclassification) are expected *a priori* to be too severe for the analysis to provide actionable information. An example would be the relation between celecoxib and upper gastrointestinal bleeding using ibuprofen as a comparator. Celecoxib is more likely to be given to patients who are at a higher risk of upper gastrointestinal bleeding in clinical practice, so an analysis with limited confounding adjustment may observe a higher risk of bleeding among celecoxib users compared to ibuprofen users, even though celecoxib does not increase the risk of bleeding more than ibuprofen. Again, this may become less of an issue when routine, sophisticated confounding adjustment modular programs become available.

Third, use of rapid assessment programs should not be considered when the precision is expected to be too low to provide important useful information; it would be better to seek alternative approaches to address such questions. Even though the Mini-Sentinel Distributed Database currently has 350 million person-years of observation time, there will always be situations where exposure or outcome is sufficiently rare that Mini-Sentinel will not be able to provide meaningful information. In these situations, FDA will need to rely on other sources such as evidence from animal data, related drugs, spontaneous reports, other databases, etc.

I. REPLICATION IN NON MINI-SENTINEL DATA SOURCE

Even though part or all of Mini-Sentinel data can be re-analyzed under certain scenarios, as described above, the committee recommends that, as with all epidemiological studies, analyses conducted in Mini-Sentinel should ideally be replicated in non Mini-Sentinel data sources whenever possible. Non Mini-Sentinel data sources in the U.S. include, for example, data from the Veterans Health Administration, the Department of Defense, and the Centers for Medicare and Medicaid Services (Medicare and Medicaid), all of which are part of the Federal Partners' Collaboration, another component of the Sentinel Initiative. There are also many non-US electronic healthcare databases.²⁴⁻²⁷

J. WORKED EXAMPLES

This section describes two actual, completed Mini-Sentinel modular program analyses, and how the committee's recommendations would be applied retrospectively.

1. Olmesartan and celiac disease

Olmesartan is an angiotensin receptor blocker approved for treatment of hypertension. In 2011, an unexpectedly high number of cases of celiac disease was observed in olmesartan users in FDA's Adverse Event Report Systems (AERS), but there was limited information from other sources to support such an association at that time. FDA requested that Mini-Sentinel perform a rapid assessment using one of its modular programs to compare the incidence of celiac disease between users of olmesartan and users of other angiotensin receptor blockers. Results from the rapid assessment have been released on the Mini-Sentinel website, and showed that, with limited adjustment for systematic errors, the risk of celiac disease was not substantially higher than the risk with other angiotensin receptor blockers.²⁸

Looking retrospectively at this example, the analysis would likely be classified as a signal refinement analysis with a weak prior. If FDA followed the committee's recommendations, the analysis might be performed using a split-sample approach, and FDA would determine at the outset what it might do upon reviewing the results. For example, if the results suggest an association and other sources of information also support the association, FDA might decide to perform a more robust assessment of the signal. If the results are null, FDA might decide against conducting additional assessments, presumably continuing to collect information through its traditional channels (e.g., AERS).

2. Dabigatran and hemorrhage and ulcer

Dabigatran is an oral direct thrombin inhibitor approved by FDA in 2010 for prevention of stroke in patients with non-valvular atrial fibrillation. The Randomized Evaluation of Long-Term Anticoagulation Therapy (RE-LY) trial found that the rate of major hemorrhage was lower in the dabigatran arm compared with the warfarin arm.²⁹ As part of its safety assessment of new molecular entities, FDA requested that Mini-Sentinel design an active surveillance protocol to monitor the safety of dabigatran in May 2011. Even though pre-market data did not suggest a higher risk of major hemorrhage among dabigatran users than warfarin users, the protocol was to include major hemorrhage as one of the primary outcomes to assess in patients receiving this drug in routine clinical settings. Following reports of dabigatran-associated hemorrhage and ulcer in AERS, the Agency further requested Mini-Sentinel to use one of its modular programs to conduct a rapid assessment of the incidence of hemorrhage (intracerebral and gastrointestinal) and ulcer between dabigatran and warfarin users with a prior diagnosis of atrial fibrillation.

Like the olmesartan example above, this rapid assessment would also likely be classified as a signal refinement analysis with a weak prior (with respect to increased risks of hemorrhage and ulcer with dabigatran use). It is not surprising that both examples are classified in the same category because signal refinement is the primary activity envisioned for Mini-Sentinel modular programs at this stage of the pilot. The same split-sample approach used in the olmesartan rapid assessment would also be applied here. Unlike most signal refinement analyses with a weak prior, the rapid assessment was performed in parallel to a planned protocol-based assessment that started prior to the convening of this committee. However, the purpose of the rapid query would remain the same – to quickly obtain some useful information (i.e., the risk of these outcomes among patients receiving dabigatran outside of randomized controlled trials). The decision to continue with the more rigorous assessment would not be affected by the results of the rapid query. On the other hand, FDA might decide not to perform the rapid assessment if it suspected that the apparent signal observed in AERS arose from severe confounding, and that the rapid query would not be able to adequately adjust for these biases sufficiently to provide actionable information.

VIII. CONCLUSION

In summary, the Mini-Sentinel modular analytic programs can be used to perform the following activities: signal generation (analysis with no prior), signal refinement (analysis with a weak or moderate prior), and signal evaluation (analysis with a strong prior). Before an analysis begins, FDA must determine to which category each analysis belongs, and its planned alternative courses of actions following review of the results. Public disclosure of the results is always needed when the set of analyses planned a priori is completed, even if additional analyses will then be forthcoming. Modular programs are most useful for

signal refinement and the majority of the analyses will likely fall under this category. Modular programs for assessment of an association should not be used when the goal is pure study size determination, when systematic errors are expected to be too severe for the analysis to provide any actionable information, or when the precision is expected to be too low to provide important useful information.

Under this general framework, Mini-Sentinel data may be re-used to refine suspected signals if subsequent analyses are designed to reduce systematic errors or to investigate further the signal (e.g., examine the signal in specific subgroups). If a full protocol-based assessment follows a modular program analysis for signal refinement, however, this latter assessment should not be interpreted as independent confirmation, such as established via replication of the same product-outcome association in two different populations. It should rather be interpreted as an analysis that has reduced, insofar as possible, systematic errors that may have been present or residual in the original modular program analysis. As the Mini-Sentinel program evolves, the sophistication of rapid assessments through modular programs will improve over time, but these rapid queries should only be used to complement, but not replace, formal protocol-based assessments for signal evaluation.

IX. ACKNOWLEDGEMENTS

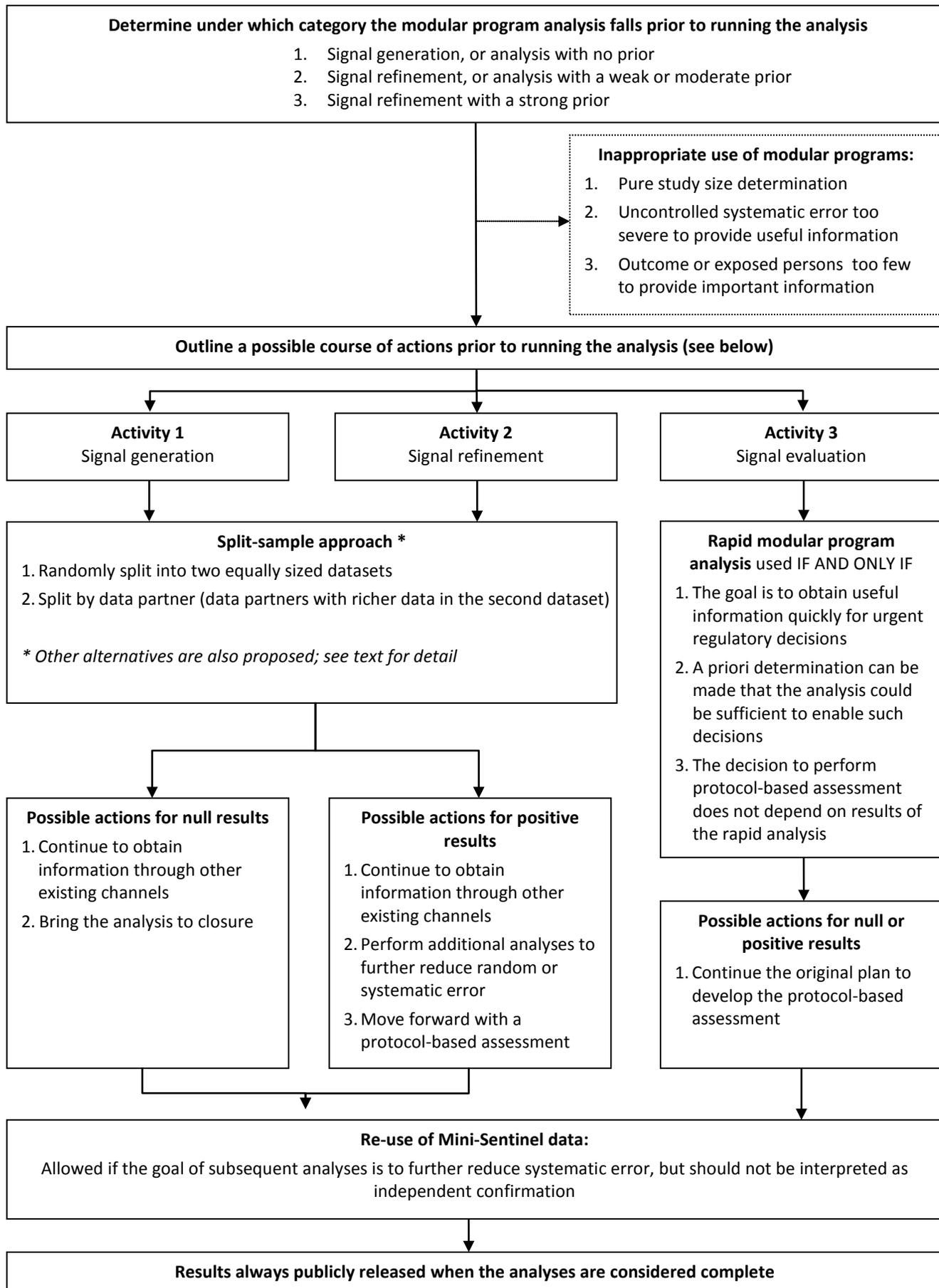
The Mini-Sentinel program is funded by the U.S. Food and Drug Administration through the Department of Health and Human Services Contract number HHSF223200910006I. The views expressed in this paper are those of the authors and are not intended to convey official U.S. Food and Drug Administration policy or guidance. The authors thank Meghan A. Baker, MD, ScD, of Harvard Pilgrim Health Care Institute, and attendees of the Brookings Institution expert workshop on September 5, 2012 for their participation in the discussion; Dr. Steven Goodman, MD, PhD, of Stanford University School of Medicine, for his input on the Bayesian approach; and Madhavi Vajani, MPH, and Kara Coughlin, BA, of Harvard Pilgrim Health Care Institute, for their administrative support.

X. TABLES AND FIGURES

Table. List of the committee members and observers

NAME	AFFILIATION
Committee members	
Brian Strom, MD, MPH (chair)	University of Pennsylvania School of Medicine
Jerry Avorn, MD	Harvard Medical School
Ralph D'Agostino, Sr., PhD	Boston University
Jerry Gurwitz, MD	Meyers Primary Care Institute
Bruce Psaty, MD, PhD	Group Health Research Institute
Kenneth Rothman, DrPH	RTI Health Solutions, RTI International
Kenneth Saag, MD, MSc	University of Alabama at Birmingham
Miriam Sturkenboom, PhD	Erasmus University
Jan Vandenbroucke, MD, PhD	Leiden University
Almut Winterstein, RPh, PhD	University of Florida
Observers	
Meghan Baker, MD, ScD	Harvard Pilgrim Health Care Institute
David Graham, MD, MPH	Food and Drug Administration
Tarek Hammad, MD, PhD	Food and Drug Administration
Hector Izurieta, MD, MPH	Food and Drug Administration
Mark Levenson, PhD	Food and Drug Administration
Marsha Reichman, PhD	Food and Drug Administration
Melissa Robb, RN	Food and Drug Administration
Azadeh Shoaibi, MS	Food and Drug Administration
Marian Strazzeri, MS	Food and Drug Administration
Darren Toh, ScD	Harvard Pilgrim Health Care Institute

Figure. General framework for rapid assessments using customizable modular programs within the Mini-Sentinel program



XI. REFERENCES

1. Food and Drug Administration Amendments Act (FDAAA) of 2007. 2007. (Accessed April 10, 2012, at <http://www.gpo.gov/fdsys/pkg/PLAW-110publ85/html/PLAW-110publ85.htm>)
2. Behrman RE, Benner JS, Brown JS, McClellan M, Woodcock J, Platt R. Developing the Sentinel System—a national resource for evidence development. *N Engl J Med* 2011;364:498-9.
3. 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.
4. Robb MA, Racoosin JA, Sherman RE, et al. The US Food and Drug Administration's Sentinel Initiative: expanding the horizons of medical product safety. *Pharmacoepidemiol Drug Saf* 2012;21 Suppl 1:9-11.
5. Nguyen M, Ball R, Midthun K, Lieu TA. The Food and Drug Administration's Post-Licensure Rapid Immunization Safety Monitoring program: strengthening the federal vaccine safety enterprise. *Pharmacoepidemiol Drug Saf* 2012;21 Suppl 1:291-7.
6. Curtis LH, Weiner MG, Boudreau DM, et al. Design considerations, architecture, and use of the Mini-Sentinel distributed data system. *Pharmacoepidemiol Drug Saf* 2012;21 Suppl 1:23-31.
7. Forrow S, Champion DM, Herrinton LJ, et al. The organizational structure and governing principles of the Food and Drug Administration's Mini-Sentinel pilot program. *Pharmacoepidemiol Drug Saf* 2012;21 Suppl 1:12-7.
8. Mini-Sentinel. Modular program 3: Frequency of select incident events during exposure to a drug/drug group of interest. 2011. (Accessed August 28, 2011, at http://www.mini-sentinel.org/work_products/Data_Activities/Mini-Sentinel_Modular-Program-3_v1.0.pdf)
9. 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.
10. Toh S, Reichman ME, Houstoun M, et al. Comparative risk for angioedema associated with the use of drugs that target the renin-angiotensin-aldosterone system. *Arch Intern Med* 2012;172:1582-9.
11. Yih K, Lieu T, Kulldorff M, et al. Monitoring for intussusception after two rotavirus vaccines by the PRISM program. 2012. (Accessed August 12, 2012, at http://www.mini-sentinel.org/work_products/PRISM/Mini-Sentinel_PRISM_Rotavirus-Protocol.pdf)
12. Nguyen MD, Greene SK, Yih WK, et al. Monitoring for venous thromboembolism after Gardasil vaccination. 2012. (Accessed August 12, 2012, at http://www.mini-sentinel.org/work_products/PRISM/Mini-Sentinel_PRISM_Gardasil-and-Venous-Thromboembolism-Protocol.pdf)
13. Butler MG, Zhou EH, Zhang F, et al. Protocol for the assessment of FDA regulatory policies for long acting beta agonists. 2012. (Accessed August 12, 2012, at http://www.mini-sentinel.org/work_products/PRISM/Mini-Sentinel_PRISM_Long-Acting-Beta-Agonists-Protocol.pdf)

sentinel.org/work_products/Assessments/Mini-Sentinel_Impact-FDA-Regulatory-Policies-LABAs-Protocol.pdf)

14. Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.
15. Rothman KJ. Curbing type I and type II errors. *Eur J Epidemiol* 2010;25:223-4.
16. Greenland S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. *Int J Epidemiol* 2006;35:765-75.
17. Storey JD. The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann Stat* 2003;31:2013-35.
18. Picard RR, Berk KN. Data splitting. *Am Stat* 1990;44:140-7.
19. Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. *Pharmacoepidemiol Drug Saf* 2010;19:858-68.
20. Arbogast PG, Ray WA. Use of disease risk scores in pharmacoepidemiologic studies. *Stat Methods Med Res* 2009;18:67-80.
21. Thomas DC, Siemiatycki J, Dewar R, Robins J, Goldberg M, Armstrong BG. The problem of multiple inference in studies designed to generate hypotheses. *Am J Epidemiol* 1985;122:1080-95.
22. Kallberg H, Alfredsson L, Feychting M, Ahlbom A. Don't split your data. *Eur J Epidemiol* 2010;25:283-4.
23. Mini-Sentinel. Principles and policies. 2012. (Accessed November 11, 2012, at http://mini-sentinel.org/work_products/About_Us/Mini-Sentinel-Principles-and-Policies-August-2012.pdf)
24. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005;58:323-37.
25. Hennessy S. Use of health care databases in pharmacoepidemiology. *Basic Clin Pharmacol Toxicol* 2006;98:311-3.
26. Miller E, Blatman B, Einarson TR. A survey of population-based drug databases in Canada. *CMAJ* 1996;154:1855-64.
27. Coloma PM, Schuemie MJ, Trifiro G, et al. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. *Pharmacoepidemiol Drug Saf* 2011;20:1-11.
28. Mini-Sentinel. Modular program report: Angiotensin II receptor blockers & celiac disease. 2012. (Accessed April 5, 2012, at http://www.mini-sentinel.org/work_products/Assessments/Mini-Sentinel_Angiotensin-II-Receptor-Blockers-and-Celiac-Disease.pdf)

29. Connolly SJ, Ezekowitz MD, Yusuf S, et al. Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2009;361:1139-51.