

MINI-SENTINEL METHODS

PROSPECTIVE ROUTINE OBSERVATIONAL MONITORING PROGRAM TOOLS (PROMPT)

USERS' GUIDE (DRAFT)

Prepared by: Richard Platt, MD, MSc,¹ Patrick Archdeacon, MD, MS,² Jeffrey Brown, PhD,¹ Ryan Carnahan, PharmD, MS, BCPP,³ Janelle Charles, PhD,⁴ Elizabeth Chrischilles, PhD, MS³ Andrea Cook, PhD,⁵ Lesley Curtis, PhD,⁶ Joshua Gagne, PharmD, ScD,⁷ David Graham, MD, MPH,⁴ Sean Hennessy, PhD, PharmD,⁸ Martin Kuldorff, PhD,¹ Grace Lee, MD, MPH,¹ Charles Leonard, PharmD, MSCE,⁸ Mark Levenson, PhD,⁴ Jennifer Nelson, PhD,⁵ Michael Nguyen, MD,⁹ Antonio Paredes, PhD,⁹ Marsha Reichman, PhD,⁴ Mary Beth Ritchey, PhD,¹⁰ Sebastian Schneeweiss, MD, ScD,⁷ Azadeh Shoaibi, MS, MHS,² Robert Temple, MD,⁴ Ram Tiwari, PhD,⁴ Darren Toh, ScD,¹ Katherine Yih, PhD, MPH,¹ Rongmei Zhang, PhD,¹¹

Author Affiliations: 1. Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA 2. Office of Medical Policy, Center for Drug Evaluation and Research, FDA, Silver Spring, MD 3. Department of Epidemiology, University of Iowa College of Public Health, Iowa City, IA 4. Center for Drug Evaluation and Research, FDA, Silver Spring, MD 5. Biostatistics Unit, Group Health Research Institute and Department of Biostatistics, University of Washington, Seattle, WA 6. Duke Clinical Research Institute and Duke University School of Medicine, Durham, NC 7. Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Boston, MA 8. Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 9. Center for Biologics Evaluation and Research, FDA, Silver Spring, MD 10. Center for Devices and Radiological Health, FDA, Silver Spring, MD 11. Division of Biometrics VII, Office of Biostatistics, Office of Translation Sciences, FDA, Silver Springs, MD

**Version 1.0
January 10, 2014**

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

Prospective Routine Observational Monitoring Program Tools (PROMPT) Users' Guide

Table of Contents

I.	INTRODUCTION	- 1 -
II.	STEPS FOR IMPLEMENTING TARGETED PROSPECTIVE SURVEILLANCE	- 1 -
A.	SELECT MEDICAL PRODUCT(S) FOR ASSESSMENT	- 1 -
B.	ASSEMBLE SURVEILLANCE TEAM	- 1 -
C.	DEVELOP SURVEILLANCE PLAN	- 2 -
D.	IMPLEMENT SURVEILLANCE	- 5 -
E.	REPORT RESULTS OF SURVEILLANCE & MODIFY SURVEILLANCE PLAN AS NEEDED	- 5 -
F.	FOLLOW-UP SURVEILLANCE FINDINGS	- 6 -
III.	APPENDICES	- 7 -
A.	APPENDIX 1: TAXONOMY MODULE SELECTION TOOL	- 7 -
B.	APPENDIX 2: CONTINUOUS AND GROUP SEQUENTIAL ANALYSIS WITH SELF-CONTROLLED DESIGN TOOL.....	- 9 -
C.	APPENDIX 3: PROPENSITY SCORE MATCHING TOOL.....	- 15 -
D.	APPENDIX 4: GROUP SEQUENTIAL GEE (GS GEE) COHORT REGRESSION TOOL	- 20 -
E.	APPENDIX 5: GROUP SEQUENTIAL INVERSE PROBABILITY OF TREATMENT WEIGHTING (GS IPTW) REGRESSION TOOL.....	- 26 -
F.	APPENDIX 6: LIBRARY OF CORE CONFOUNDERS	- 30 -

I. INTRODUCTION

The Mini-Sentinel pilot has developed the capacity to perform prospective routine observational monitoring of medical products. This Prospective Routine Observational Monitoring Program Tools (PROMPT) system will allow FDA to assess the occurrence of a fixed number of pre-specified health outcomes of interest (HOIs) that may occur in association with use of newly approved medical products. This surveillance will occur prospectively and sequentially as new products begin to be taken up by the health care system. The surveillance system will complement other approaches to safety monitoring, including 1) surveillance based on spontaneous reports, 2) untargeted surveillance (data mining) of healthcare data, 3) customized surveillance programs requiring custom protocols, and 4) formal epidemiologic studies intended to support causal inference, examine dose-response relationships, identify high-risk subgroups, etc.

The PROMPT system is ultimately expected to allow FDA to monitor multiple products simultaneously. It employs a variety of surveillance methods that will be implemented through a library of method modules, called *PROMPTs* that can be customized to accommodate specific agents, outcomes, populations, and periods for assessment. The four methods for which PROMPTs have been developed are self-controlled design, propensity score matching (including the option for matching on high dimensional propensity score), outcome regression, and inverse probability of treatment weighting. PROMPT's emphasis is on identifying potential risks of a limited number of HOIs that have been selected because 1) that HOI occurs in association with several other medical products of that type (e.g., acute liver injury for drugs; febrile seizures for vaccines) or another product in the class and is thus of general interest; or 2) there is a reason to suspect that product in particular might increase the risk of that HOI, for example because of a signal identified in pre-approval animal studies or clinical trials or because of the mechanism of action of the product. PROMPT was not designed for untargeted surveillance of HOIs for which there is little or no reason for concern regarding the product being monitored.

II. STEPS FOR IMPLEMENTING TARGETED PROSPECTIVE SURVEILLANCE

A. SELECT MEDICAL PRODUCT(S) FOR ASSESSMENT

The first step will be to identify product(s) to be monitored. This will be done by FDA in consultation with the MS PROMPT Implementation Team for determining the suitability of a product for PROMPT surveillance. Planning for prospective surveillance should begin approximately six months before the Mini-Sentinel Distributed Database (MSDD) is expected to show uptake.

B. ASSEMBLE SURVEILLANCE TEAM

The Surveillance Team for a particular product-HOI pair should include one or more FDA content experts, one or more FDA safety evaluation experts, one or more FDA biostatisticians, other FDA experts, and Mini-Sentinel investigators including, at least initially, one or more Mini-Sentinel Protocol Core co-leaders and a lead Mini-Sentinel biostatistician. At least initially, the relevant PROMPT module leader(s) and the Taxonomy Workgroup leader will serve as consultant(s) to the Surveillance Team. Once the Surveillance Team selects a specific PROMPT module, technical experts for those tools will be assigned to support the Surveillance Team. The Mini-Sentinel Operations Center (MSOC) will assign other individuals to support the Surveillance Team as needed.

C. DEVELOP SURVEILLANCE PLAN

1. Step 1: Select and define health outcomes of interest (HOIs)

FDA will recommend HOIs for each product to the Surveillance Team. This selection will be designed to fill gaps in the existing safety evidence base and to complement other surveillance activities conducted at FDA (e.g., spontaneous reporting). Considerations for selection include:

- Pre-marketing pre-clinical and clinical safety data
- Post-marketing spontaneous event reports (including those from outside the U.S.)
- Biologic plausibility/mechanism of action
- Event severity
- Availability and known or expected performance characteristics of algorithms to identify that HOI

The Surveillance Team will examine suitability of HOIs (i.e., considering likely excess risk, latency period, and other features of the exposure-HOI pair) for routine surveillance, and may recommend additional or alternative HOIs. HOIs will typically be defined using the Mini-Sentinel HOI Algorithms Library (under development) or other sources. Algorithms may need to be developed and added to the Library in response to emerging regulatory needs. In addition, for each HOI, the Surveillance Team will specify the beginning and duration of the at-risk window after exposure. Additional steps may be appropriate to further refine outcome definitions. Some examples include: 1) it may be desirable to include only incident events, defined as the first documented code in the MSDD in a pre-specified period or to define a specific period prior to exposure date during which the HOI did not occur (e.g. 30 days, 180 days); 2) for some HOIs, it may be desirable to exclude events occurring on the first day of exposure in order to exclude events with onset prior to receipt of the medical product; and 3) depending on the HOI, it may be desirable to include events defined by a subset of encounter types (e.g., outpatient, inpatient, and emergency department) to more narrowly define outcomes to reduce noise or restrict to the most severe outcomes, which may be less prone to misclassification. Although a number of recommended algorithms specify inclusion of only primary diagnoses, sensitivity analysis should be considered that use alternate algorithms that include various HOI specifications. The MSOC Data Core is continually investigating the variability and completeness of capture of key data elements such as diagnoses due to the potential to introduce bias.

Some HOIs have multiple potential coding algorithms. For example, one algorithm may maximize the positive predictive value (PPV), whereas another may maximize sensitivity while maintaining an acceptable PPV. Some algorithms include requirements for laboratory data, which is not yet a fully-developed capability in the MSDD, whereas others do not. Some algorithms may function better in a certain subpopulation of interest (e.g., children), whereas others might be best applied to a general population. These considerations should be taken into account when selecting HOI algorithms. The Mini-Sentinel HOI Algorithm Library includes details for selected priority outcomes. It should also be noted that current algorithms have been developed based on the International Classification of Diseases, Ninth Revision (ICD-9) and other coding systems currently in use. When ICD coding shifts to the Tenth Revision (ICD-10) on October 1, 2014, this will require the specification of new algorithms based on ICD-10 codes in order to utilize newly available data.

2. Step 2: Define medical product exposure and cohort eligibility

For each medical product of interest, the definition of exposure to that product based on the MSDD will be determined. Eligibility criteria for inclusion in the surveillance cohort will also be specified and may vary by the particular HOI being monitored. For instance, when monitoring for the occurrence of acute myocardial infarction (AMI), those with an AMI before first receipt of the monitored medical product may be excluded to restrict the surveillance to incident events. Cohorts may be selected based on indications for the product being monitored, and those with contraindications to receiving the product may be excluded. In addition, separate cohorts may be defined based on a variety of factors. In the event that a product has multiple indications, it may be appropriate to identify separate cohorts based on indication. Since product safety may vary depending on risk factors for a HOI in exposed patients, separate cohorts may also be identified based on absence or presence of major risk factors for the event. Subgroups of interest may also be defined. The cohort and exposure definition selected will have to be consistent with the inclusion/exclusion and exposure parameters available within each PROMPT module; initially, this restriction may limit some implementation options.

3. Step 3: Select appropriate epidemiologic design and comparator for each product-HOI pair

The Taxonomy Framework (Appendix 1) provides guidance on the selection of an appropriate design, with the primary decision being whether to use a within-person (i.e., self-controlled) or between-person (i.e., cohort) design. *PROMPT: Group Sequential Analysis with Self-Controlled Design* makes within-person comparisons while *PROMPT: Cohort Matching*, *PROMPT: Group Sequential GEE (GS GEE) Cohort Regression* and *PROMPT: Group Sequential Inverse Probability of Treatment Weighting (GS IPTW) Regression* compare independent exposed and comparator groups. Each PROMPT module package contains: a) overview summaries (please see Appendices 2, 3, 4, and 5); b) technical users' guides; and c) input programming specifications and programming code.

Each design requires different decisions about comparators. In the case of a self-controlled design, an appropriate control window must be selected, either prior to receipt of the medical product or after a washout period following the at-risk window of interest or both. In the case of a cohort design, users of an alternative product (i.e., active comparator) must be selected (or a medical product-unexposed comparator can be used with *PROMPT GS GEE*). In the cohort designs, typically, an active comparator referent will be preferred to minimize concerns of confounding by indication and other biases. The active comparator may be a product thought not to pose a risk of the HOI or, alternatively, one with a well-characterized risk of the HOI. A given medical product might have multiple comparators. If multiple comparators are desired, this may require multiple runs of the selected PROMPT module in parallel. Further, different comparators may be selected for different HOIs. When a product has multiple indications, different comparators may be selected for different indications. In many instances, new users rather than prevalent users of an alternative product may be preferred. Additional guidance for selection of comparators in cohort designs is provided in Chapter 5 (Comparator Selection) of

Developing an Observational CER Protocol: A User's Guide, published by the Agency for Healthcare Research and Quality.¹

The Taxonomy PROMPT Selection Tool (Appendix 1) provides a detailed set of considerations for choosing a study design (self-controlled vs. cohort) for a particular medical product-HOI pair. Key considerations include: onset of the exposure-risk window, exposure use pattern, strength of within- and between-person confounding, and suitability of available comparators (active vs. medical product unexposed). When a cohort design is selected, the team will need to specify whether the primary measure of association is a ratio measure or a difference measure. Ratio measures may be less sensitive than difference measures to the imperfect performance of algorithms to identify HOIs, although this will depend on the validation metrics of the given HOI algorithms implemented. While it would be important to select one measure as that of principal interest, it may be desirable to generate both measures as they provide complimentary information.

For between-person designs, the Surveillance Team will specify which potential confounders will be adjusted for. A list of core confounders that deserve consideration is presented in Appendix 6; not all of the factors included on this list necessarily need to be adjusted for in every surveillance activity. Further, additional factors not found on this list may be considered. However, specification of additional confounding factors will reduce the degree to which PROMPT surveillance can be automated and if the inclusion of numerous such factors is deemed critical, it may suggest that a protocol-based assessment is preferable to PROMPT surveillance. Naturally, each factor selected will have to be consistent with the ability of the PROMPT module to define and implement the factor; initially, this restriction may limit the use of some factors. Mini-Sentinel intends to create and curate a Confounder Algorithms Library based on experiences from PROMPT surveillance and other Mini-Sentinel activities. The Taxonomy PROMPT Selection Tool provides guidance for choosing a confounder adjustment strategy (e.g., matching, regression, or weighting). Other factors may also be important to consider. For instance, *PROMPT: Cohort Matching* has the ability to further adjust for empirically-identified variables that were not pre-specified.

4. Step 4: Specify the sequential analysis plan for each product-HOI pair

A key feature of PROMPT is that the surveillance methods are designed to be implemented routinely over time as a new product is introduced. Descriptive monitoring can occur each time any Mini-Sentinel Data Partner refreshes the information in the database at their site. For instance, a description of the rate and the population composition of new product uptake as well as the overall incidence of each HOI across the entire surveillance cohort (i.e., exposed and unexposed groups combined) can be assessed any time that data are refreshed at any Mini-Sentinel Data Partner. Logistical and resource considerations will be taken into account when deciding whether all Data Partners should re-run a given PROMPT module each time data are refreshed, or whether a given increment of exposure should be required to re-run the PROMPT module.

¹ Developing an Observational CER Protocol: A User's Guide, published by the Agency for Healthcare Research and Quality, Chapter 5 (Comparator Selection). Available at: http://www.effectivehealthcare.ahrq.gov/ehc/assets/File/Ch_5-User-Guide-to-OCER_130129.pdf.

Formal statistical testing to assess pre-specified product-HOI signals will occur at pre-specified time points based on input parameters that reflect sequential design choices made by the Surveillance Team. In the absence of a signal for a particular medical product-HOI pair, surveillance will end when the maximum planned sample size or number of events is reached. Each PROMPT module allows for a variety of potential sequential designs to be implemented, and these are represented by the following general inputs:

- The frequency of testing, which may be based on calendar time to accommodate required reporting (e.g., an 18-month report), the schedule with which the Data Partners update data, or on the number of new users/doses of the product of interest in order to ensure that an adequate amount of new information has been observed to warrant formal testing.
- The general shape of the signaling boundary over time (e.g., a constant or flat signaling threshold over time versus a signaling boundary that employs early conservatism with a higher signaling threshold at early tests and lower thresholds at later tests).
- The maximum planned sample size or number of events for surveillance.
- The desired overall Type 1 error across all tests that are performed.

As is done in sequential monitoring planning activities for randomized clinical trials, these quantities will be determined based on scientific, practical, and statistical criteria relevant to the surveillance questions of interest, including the desired statistical power for detecting differences between exposure groups that are of interest to FDA.

Documentation of the Surveillance Team members, products, comparators, HOIs, HOI algorithms, other design choices (e.g., cohort selection, any special approaches to confounder adjustment), a specific PROMPT module, input parameters for that PROMPT module, and a general plan for follow-up of positive signals will constitute the Surveillance Plan. The Surveillance Plan will be reviewed by FDA and posted for comment on the Mini-Sentinel public website. After the public comment period and final FDA approval, subsequent changes to the Surveillance Plan will require FDA approval, and then they will be dated and posted on the Mini-Sentinel public website.

D. IMPLEMENT SURVEILLANCE

From the perspective of the Surveillance Team, surveillance will begin when the Surveillance Plan approved by FDA is submitted to the MSOC for implementation. Upon receipt, the MSOC will seek clarification from the Surveillance Team on any necessary points, customize the necessary PROMPT modules, test them as appropriate, and initiate surveillance by distributing the surveillance package to the appropriate Data Partners. Surveillance results will be returned from the Mini-Sentinel Data Partners and aggregated by the MSOC for Surveillance Team review. This process will iterate (based on the specifications of the surveillance plan) until the end of surveillance.

E. REPORT RESULTS OF SURVEILLANCE & MODIFY SURVEILLANCE PLAN AS NEEDED

Interim findings and final surveillance results will be presented to the Surveillance Team based on the output of each PROMPT module and the specifications in the surveillance plan. Interim findings and final results will be communicated in accordance with Mini-Sentinel's Principles and Policies.

F. FOLLOW-UP SURVEILLANCE FINDINGS

Timely, pre-planned product-specific assessment of positive signals is part of the active surveillance framework. A general plan for follow-up of positive signals will be outlined during the planning phase, although this may be modified according to the nature of the positive signal. The Mini-Sentinel final report on a *Framework for Assessment of Signal Refinement Positive Results* provides guidance on the types of follow-up activities that might be considered.² Signals that are detected will be further investigated by first evaluating utilization patterns, cohort characteristics, and analytic assumptions. Adjustment for additional confounders and testing against additional comparators, as well as, a variety of pattern evaluation descriptive analyses would next be considered. Where applicable, a temporal scan will be conducted to assess whether the observed adverse outcomes cluster within the specified at-risk time window. Subgroup analyses (e.g., by site, age group, risk factor, and/or specific diagnosis or procedure code within a given outcome group) will be performed to assess the robustness of the signal. If a signal persists, quantitative bias analysis and medical record review will be considered, depending on the nature of the adverse event and the results of the above analyses. Decisions regarding the follow-up of a particular positive or negative signal will be made by FDA in consultation with the Surveillance Team.

² McClure, DL et al. Framework for Assessment of Signal Refinement Positive Results. Mini-Sentinel 2013. Available at: http://mini-sentinel.org/methods/methods_development/details.aspx?ID=1042.

III. APPENDICES

A. APPENDIX 1: TAXONOMY MODULE SELECTION TOOL

1. Purpose

The Mini-Sentinel Taxonomy Module Selection Tool is intended to help guide the decision-making process when selecting a PROMPT module for a routine active surveillance activity. It is not designed to fully replace this process but rather to expedite key decisions related to surveillance design and analysis.

2. Descriptions

Mini-Sentinel has developed four PROMPT modules to conduct routine active surveillance, each with a different confounder adjustment strategy: (1) self-controlled design; (2) propensity score exposure matching; (3) outcome regression; (4) inverse probability weighting. Details about the programs can be found in Appendices 2-5.

The Taxonomy Module Selection Tool is a Microsoft Excel-based program that guides users through the Mini-Sentinel Taxonomy Framework and provides recommendations for design and analytic strategies for active monitoring of a pre-specified health outcomes of interest (HOI) for an FDA-selected exposure based on user-specified parameters related to that exposure-HOI pair (or scenario). The tool requires users to specify 11 scenario characteristics, which are categorized as those determined by the stakeholder or investigator, those pertaining to exposure, those pertaining to the HOI and those pertaining to the (potential) link between the exposure and the HOI. For each characteristic, users select from among up to four options.

3. Instructions for Use

1. Open the [Taxonomy Module Selection Tool](#).³ An example scenario comes preloaded.
2. Type the name of the medical product into the “Medical product” column (column B). The tool can accommodate up to 35 scenarios. See below for instructions on how to expand to more than 35 scenarios.
3. Type the name of the health outcome of interest in the “Health outcome of interest (HOI)” column (column C).
4. Select the appropriate option for each of the 11 characteristics in columns D through N. Brief explanations and options available for each characteristic are provided immediately below the table. To select an option for a given characteristic, click on the cell in which the column for that characteristic intersects with the row for the particular scenario (e.g., cell D6 for “Effect measure of interest” for scenario 1). Either click on the arrow to make a

³ Gagne, et al. Taxonomy of Observational Study Designs – Phase 3. Mini-Sentinel 2013. Available at: http://mini-sentinel.org/methods/methods_development/details.aspx?ID=1044.

- selection or type in one of the available options. Typing will cause the option to appear. To select it, simply click on the name of the option. The tool requires that the options in each cell exactly match one of the options below the table; else an error will occur. Options can be changed at any time. A selection for each of 11 characteristics is required in order to have the recommended design and analysis appear.
5. Once an option has been selected for each of the 11 characteristics, the tool will provide a recommended design and analysis approach. The tool will also provide important warnings where appropriate.
 6. Save the file if you would like to return to the specified inputs. Be sure to save the file with a new name to prevent overwriting files. An original blank version can be downloaded from the Mini-Sentinel website.⁴

4. Additional Information

- The tool accommodates up to 35 scenarios at a time. If a user wishes to include more than 35 scenarios, multiple files can be used.
- Users may want to estimate both a difference measure (e.g., risk difference) and a ratio measure (e.g., rate ratio) of association for a particular scenario. To accomplish this, the medical product and HOI pair should be inserted on two separate rows, with *difference measure* selected as the “Effect measure of interest” in one row and *ratio measure* selected in the other row.
- Cells outside of those used to select the scenario characteristics (i.e., non-gray cells) are locked to prevent accidental changes to the formulas in the tool. A password is required to unlock these cells.

⁴ Gagne, et al. Taxonomy of Observational Study Designs – Phase 3. Mini-Sentinel 2013. Available at: http://mini-sentinel.org/methods/methods_development/details.aspx?ID=1044.

B. APPENDIX 2: CONTINUOUS AND GROUP SEQUENTIAL ANALYSIS WITH SELF-CONTROLLED DESIGN TOOL

1. Overview of Design

Self-controlled methods are a powerful and often used design for drug and vaccine safety assessments, when looking for acute outcomes that occur soon after initial exposure. By using exposed and unexposed time from the same study subjects, all non-time varying confounders are automatically adjusted for, including gender, insurance coverage, education, income, geography, etc. Self-controlled methods use a pre-defined risk window, such as 1-21 days after initial exposure, and can make comparisons to a pre-defined control window of the same or different length. Depending on the product-outcome pair under investigation, the control window may be either before exposure or after the risk window. If occurrence of the outcome influences or could influence whether the medical product is given to the patient, it is important to use a control window after the risk window rather than a pre-exposure control window, in order to avoid bias by contraindication. Technically, it is possible to have a control window right after the initial exposure, if no increased risk is plausible then, and a risk window after the control window, but this is not commonly done.

There are several different self-controlled designs. The one employed by this PROMPT module: Continuous and Group Sequential Analysis with Self-Controlled Design Tool has sometimes been referred to as the “self-controlled risk interval” design.^{5,6,7,8}

The current version of the tool has some limitations that should be noted. For example, in the current version of the tool, the ratio of the durations of the risk and control windows must be the same for all patients. The tool could be modified to allow for different ratios for different patients. In addition, the self-controlled tool does not currently adjust for time-varying confounders unless explicitly modified to do so. This capability has not been built into the tool as yet. Hence, it should be used with caution for very young children, whose baseline risk of certain health outcomes can change quickly within small increments of age, such as at 3 versus 5 weeks of age. In such cases, it would be important to adjust for age using an offset term.

⁵ Kramarz P, DeStefano F, Gargiullo PM, et al. Does influenza vaccination exacerbate asthma? Analysis of a large cohort of children with asthma. Vaccine Safety Datalink Team. *Archives of family medicine*. Jul 2000;9(7):617-623.

⁶ Klein NP, Hansen J, Lewis E, et al. Post-marketing safety evaluation of a tetanus toxoid, reduced diphtheria toxoid and 3-component acellular pertussis vaccine administered to a cohort of adolescents in a United States health maintenance organization. *The Pediatric Infectious Disease Journal*. Jul 2010;29(7):613-617.

⁷ Greene SK, Kulldorff M, Lewis EM, et al. Near real-time surveillance for influenza vaccine safety: proof-of-concept in the Vaccine Safety Datalink Project. *American Journal of Epidemiology*. Jan 15 2010;171(2):177-188.

⁸ Lee GM, Greene SK, Weintraub ES, et al. H1N1 and seasonal influenza vaccine safety in the vaccine safety datalink project. *American Journal of Preventive Medicine*. Aug 2011;41(2):121-128.

The self-controlled design should be used with caution if there is seasonality in both medical product initiation and the outcome under study, in which case seasonality must be adjusted for. If season influences only one of these two (i.e., either the drug/vaccine exposure or the outcome), the method will still be unbiased, and no adjustment is needed. Another potential source of bias is if there is something that triggers both the exposure and the outcome. For example, patients are often given the pneumococcal polysaccharide vaccine (PPSV) just before an organ transplant, and organ transplants may cause adverse events. Hence, in a self-controlled analysis, there may be more adverse events just after PPSV vaccination even though the vaccine is not responsible. However, this potential bias would not be unique to self-controlled approaches.

2. Data

For both the extraction of data (in SAS) and the analysis of data (in R), the user must make some decisions in advance. For data extraction, these include inclusion/exclusion criteria, exposure definition, health outcome of interest definition, risk and control intervals, etc. The data extraction SAS program is parameterized such that the user specifies these criteria in look-up and mapping tables rather than in the program itself. For convenience, the investigators may create, populate, and/or revise these tables in Access, but they must be imported into SAS in order for the extraction program to call them.

We present the items to be specified and the look-up tables by means of four example studies. These are 1) lisinopril and angioedema, 2) clindamycin and acute myocardial infarction (AMI), 3) Measles, Mumps, Rubella, and Varicella (MMRV) vaccine and seizures, with a 1-6 day post-vaccination control interval (or “window”), and 4) MMRV vaccine and seizures, with a 15-28 day post-vaccination control interval. To extract appropriate data for analysis using this self-controlled design, a number of features must be specified, based on clinical and/or epidemiologic judgment. These are listed and illustrated via the four example studies in Table 1.

Table 1. Items to Be Specified and Look-up Tables for Four Example Studies

	Lisinopril	Clindamycin	MMRV-1	MMRV-2	Notes
Minimum period of continuous enrollment with pharmacy and medical benefits required prior to dispensing date	183 days	183 days	183 days	All fields same as for MMRV-1 except for control window	An apparent gap in enrollment can be specified such that the gap is bridged (ignored).
Age range at exposure	≥18 yrs.	≥45 yrs.	12-23 mo.		
Exposure	Lisinopril	Clindamycin	MMRV		
Exposure exclusions	Other ACEIs, ARBs, direct renin	Other antibiotics	None		Specify if patients with certain exposures on the

	Lisinopril	Clindamycin	MMRV-1	MMRV-2	Notes
	inhibitors				same day as or in the x days prior to the exposure of interest are to be excluded.
Exposure look-back period from dispensing date*	0-183 days	0-183 days	0-183 days		
Health outcome of interest (HOI)	Angioedema	AMI	Seizures		
HOI settings	All	Inpatient	Inpatient, ED (inpatient has higher priority within a day)		Possible settings in which to look for potential cases include inpatient, ED, and clinic (outpatient).
HOI exclusions	Prior angioedema	Prior AMI	Prior seizures		
HOI exclusion settings	All	Inpatient	All		For MMRV, cases of interest are those in inpatient and ED settings but only if there were no cases in the prior 183 days in <i>any</i> setting, including clinic.
HOI look-back period from diagnosis date**	1-183 days	1-28 days	1-183 days		The case must be the patient's first one to occur in this period of time in order to count.

	Lisinopril	Clindamycin	MMRV-1	MMRV-2	Notes
Risk window	1-28 days	1-14 days	7-10 days		This is the post-exposure (Day 0) period during which an increased risk of the HOI is considered possible and will be examined.
Comparison window	-35 to -8 days	15 to 28 days	1 to 6 days	15 to 28 days	This is the period, either pre- or post-exposure (Day 0), during which the risk is considered to be at baseline.
Maximum period to collect to include both windows	-35 to +28 days	1 to 28 days	1 to 28 days	1 to 28 days	

*Include only those with no dispensing of any drug in the drug look-up table (for the cohort in question) during the 0-183 days prior to the dispensing date. Individuals who initiated both the drug of interest and an excluded drug on the dispensing date (Day 0) are to be excluded.

**Two considerations go into this look-back period. One is clinical—for example, a single episode of illness might lead to multiple visits, so we specify that we want to count cases only if they represent the first occurrence of the condition in a certain number of days. The lisinopril-angioedema and MMRV-seizures assessments are examples of this, where we set the look-back period to 6 months (183 days), because we did not want to include repeat visits for the same episode of illness. The other consideration is method-related—with the self-controlled design, the only informative patients are ones that have the outcome of interest in either the risk or control interval; patients who have a case of the outcome in the risk interval and another case in the control interval would not be used. Therefore, we specify that the case must be the first in the maximum span between risk and control intervals, as in the clindamycin-AMI study, in which we specified a look-back period of 28 days.

3. Descriptive Analyses

The descriptive outputs from each look at the data will include:

- Test number
- Number of the most recent batch of data included
- Number of new events in the risk interval (“cases”)
- Number of new events in the control interval (“controls”)
- Cumulative number of exposed patients
- Cumulative number of cases
- Cumulative number of controls

- Expected number of cases

In addition, there will be a data tracking table, which will include:

- Data partner
- Number of exposed patients
- Number of cases
- Number of controls
- Earliest possible dispensing date in dataset
- Latest possible dispensing date in dataset
- Date of dataset creation

4. Sequential Monitoring

In sequential monitoring using this tool, the hypothesis being tested is that the risk of the HOI in the risk (exposed) interval is equal to the risk in the control (unexposed) interval, adjusting any difference in length between the two intervals. The test statistic is the log likelihood ratio. The default shape of the stopping boundary is flat with respect to the log likelihood ratio. Both continuous and group sequential analysis can be done, and any testing frequency can legitimately be followed.

Before conducting sequential tests using the data, the user must specify and enter directly into the R code several parameters such as the overall false positive rate (or alpha level) and maximum number of cases to accrue before stopping surveillance: These parameters are described in Table 2.

Table 2. Parameter Specification for Sequential Monitoring

Parameter	Explanation	Considerations
alpha	Desired alpha level overall	0.05 is the default, if no other value is chosen
M	Minimum number of cases in risk + control intervals to signal	1 is the default, if no other value is chosen; using 1 does not mean the scenario of 1 case in the risk interval and 0 cases in the control interval will produce a signal
N	Cumulative number of cases in risk + control intervals at which surveillance is to stop even if H_0 has not been rejected (sometimes called the “upper limit”)	No default; the decision should depend in part on power to detect the lowest risk of concern for public health

The associated technical users' guide for programmers describes how self-controlled sequential analysis can be conducted within the Mini-Sentinel project, utilizing SAS code and an R package that has been developed. It does not provide details about sequential statistical analysis or self-controlled designs but rather focuses on implementation, using SAS code for data extraction and R functions for sequential analysis. Detailed step-by-step guidance and multiple sample screen shots are provided in the associated technical users' guide for programmers.

C. APPENDIX 3: PROPENSITY SCORE MATCHING TOOL

1. Overview

This program performs effect estimation and sequential testing in a distributed data setting based on exposure propensity-score matched sequential, parallel new user cohorts. Cohort approaches are particularly useful when comparing outcomes between patients exposed to different medical products. The approach focuses on new users, which ensures the accurate assessment of temporality among exposures, outcomes, and other study variables, and ensures that outcomes that occur shortly after initiation are captured. The use of an active comparator reduces confounding to the extent that the outcome risk factors similarly determine exposure to the product of interest and the comparator. Use of active comparators also ensures that both exposure groups have progressed to the point of requiring treatment, and can also prevent immortal time bias. Propensity score methods are used to further minimize confounding by balancing a potentially large number of possible confounders.

2. Design and data

This program uses a standard active-comparator new user cohort design. The program automatically identifies new users of the product of interest and new users of a user-specified comparator product within each monitoring period. New use is defined by no prior use of the product (or potentially of other pre-specified products) in a pre-specified period preceding each patient's product initiation (i.e., index) date. Outcomes are identified over a pre-specified risk window following product initiation. The program can accommodate any outcome definition that can be coded in a SAS macro. Patients can be followed for as long as they are exposed to the product ("as treated") or using an intention-to-treat approach, in which patients continue to contribute person-time to the index product category over an interval of pre-specified length.

Potential confounders are identified in a baseline period of pre-specified length preceding each patient's index date. Importantly, all confounders are measured before exposure to the medical product. Pre-defined potential confounders are forced into a site (or Data Partner)-specific propensity score model with options for including empirically identified potential confounders using the high dimensional propensity score algorithm, a comorbidity score, and health service utilization variables, such as number of drugs used, number of physician visits, and number of hospitalizations. A separate propensity score is estimated in each Data Partner and in each monitoring period. Patients are matched by propensity score within each Data Partner and monitoring period. The program permits subgroup analyses on any pre-defined variable.

The following list summarizes the key program inputs. A more detailed and comprehensive list of inputs can be found in the technical documentation for this program.

ELIGIBILITY INFORMATION

- Enrollment gap: specifies the number of days bridged between two consecutive enrollment periods to create a single continuous enrollment period.
- Inclusion/exclusion conditions: defined by creating a SAS dataset with codes defining the inclusion or exclusion of conditions(s) of interest.

EXPOSURE INFORMATION

- Medical product of interest: defined by creating a SAS dataset with codes (i.e., National Drug Codes [NDCs] or Current Procedural Terminology [CPTs]) to identify the product of interest.
- Comparator of interest: defined by creating a SAS dataset with codes (i.e., NDCs or CPTs) to identify the comparator product of interest.
- New user definition:
 - Duration: specifies length of washout period (in days) to determine new user status.
 - Products to define new use: defined by creating a SAS dataset with codes (i.e., NDCs or CPTs) for products to which patients must not have had exposure during the washout period in order for them to be considered new users of the product and comparator product of interest.
- Exposure definition during follow-up: specifies whether to use an “as treated” or “intention to treat” (ITT) approach to defining exposure status following the index date.
 - Induction period: specifies when, with respect to the index date, follow-up begins for both the “as treated” and ITT approaches.
 - Treatment episode gap: specifies the number of days allowed between two consecutive claims to consider them as part of the same treatment episode.
- Duration of ITT follow-up: if the ITT approach is selected, this specifies the maximum duration of follow-up for each patient starting at the end of the induction period.

COVARIATE INFORMATION

- Length of covariate assessment period: specifies the length of the period preceding the index date (i.e., the medical product initiation date) over which potential confounders are measured.
- Pre-specified covariates:
 - Procedures: defined by creating a SAS dataset with procedure codes that will be used as covariates.
 - Conditions: defined by creating a SAS dataset with ICD9 codes that will be used as covariates.
 - Medications: defined by creating a SAS dataset with NDC codes that will be used as covariates.
 - Comorbidity score: indicates whether the user would like to include the Combined Comorbidity Score.
 - Health service utilization variables: indicates whether the user would like to include the health service utilization variables.
- Subgroups: indicates which (if any) of the pre-specified covariates will be used as subgroup indicators. (Note: these are specified in the input SAS datasets above)
- Age groups: specifies cut points for age strata

- High-dimensional propensity score options: (Note: standard defaults are used if no options are specified)
 - Ranking algorithm: indicates whether empirically-identified variables are ranked based on associations with exposure only, with outcome only, or with both exposure and outcome. (Note: default is based on exposure only)
 - Covariates considered: specifies the number of empirically-identified variables to consider from each data dimension (i.e., diagnosis codes, procedure codes, drug codes). (Note: default is 100)
 - Covariates selected: specifies how many empirical covariates are included in the propensity score. (Note: default is smaller of 200 or number of initiators of the product of interest)

OUTCOME INFORMATION

- Outcome of interest: defined by a SAS-algorithm to identify the outcome.
- Outcome washout: specifies whether patients are allowed to have the outcome of interest prior to the index date. In the case that patients are not allowed to have the outcome of interest before the index date, the duration of washout preceding the index date must be specified. This is also used to specify minimum duration of pre-index date enrollment.

When analyses are conducted prospectively, new Data Partner-specific propensity scores are estimated in each monitoring period. Propensity score models include all eligible patients up to and including each new monitoring period. However, only new initiators in the most recent monitoring period are matched over time. Matches identified during prior monitoring periods are not broken; once a patient is matched, that patient remains matched throughout the sequential analysis.

The program generates a de-identified, individual-level data set for each Data Partner in each monitoring period. The data set contains the minimum information required for central aggregation and analysis by the Mini-Sentinel Operations Center, including a de-identified Data Partner indicator, the monitoring period in which each patient was identified, a variable indicating each patient's person-time of follow-up, propensity score values, propensity score matched set numbers, subgroup indicators, and other subgroup variables (age, sex, and race). The propensity score summarizes the necessary information for confounding adjustment while obscuring detailed patient-level information. The table below provides an example of the file, with hypothetical data, that each Data Partner will create. The data set contains one row for each patient included in the analysis. The only individual covariate data are those required for the subgroup indicators as all other covariate information is summarized by the propensity score. This approach has been reviewed by a legal expert who confirmed that it complies

with HIPAA.⁹ The information requested from each Data Partner meets the minimum necessary standard specified in the HPIAA Privacy Rule.¹⁰

Table 3. Example of Aggregated Data That Each Data Partner Will Transmit to MSOC

Data Partner ID	Random patient ID	Exposure status	Outcome status	Person-time	Propensity score	Match set ID	Age	Sex	Race	Subgroup indicator
1	1	1	0	320	0.131	1	55	M	1	1
1	2	0	0	309	0.131	1	55	M	1	1
1	3	1	1	193	0.085	2	63	F	3	0
1	4	0	0	246	0.084	2	62	F	3	0
1
1	<i>n</i>	1	0	45	0.051	<i>j</i>	64	F	1	0

By including a person-time variable, the de-identified individual-level data permit an aggregate time-to-event analysis at the Mini-Sentinel Operations Center, including estimation of hazard ratios and incidence rate differences. Data sets from each Data Partner are appended and used for subsequent analysis

3. Descriptive analyses and effect estimation

The program automatically generates tables of patient characteristics, stratified by exposure group, for the unmatched cohort and for each matched cohort, separately from each Data Partner and each monitoring period. Tables include measures of covariate balance, including absolute and standardized differences, which indicate balance in specific variables, and the Mahalanobis distance, which provides a measure of balance across all variables while accounting for their correlation. The tables also include the number of patients in each exposure group, the number matched from each group (where appropriate), the number that experienced outcomes, and the mean person-time of follow-up.

The program also automatically generates figures depicting the propensity score distributions for each exposure groups, separately from each Data Partner and each monitoring period. Figures include c-statistics for each propensity score model.

Using summarized data generated in the data extraction step, the program can estimate both hazard ratios (with 95% confidence intervals) and incidence rate differences (with 95% confidence intervals). Note that the confidence intervals do not account for repeated looks or correlation in the data across looks, but are provided for descriptive purposes. Data are aggregated across Data Partners in a stratified

⁹ Rassen JA et al. Evaluating strategies for data sharing and analyses in distributed data settings. Mini-Sentinel 2013. Available at: http://www.mini-sentinel.org/methods/methods_development/details.aspx?ID=1041.

¹⁰ U.S. Department of Health and Human Services. Office for Civil Rights. Minimum Necessary Requirement. 45 CFR 164.502(b), 164.514(d). Available at: <http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/minimumnecessary.html>.

Cox regression model to estimate hazard ratios. The time scale in the Cox model is time since medical product initiation. The program estimates rate differences with a Mantel-Haenszel difference estimator for stratified person-time data. Both approaches stratify by both Data Partner and matched set within each Data Partner. The program also calculates unadjusted hazard ratios and risk differences, stratified by Data Partner as well as the number needed to treat/harm (NNT/NNH), the attributable risk, and the population attributable risk.

4. Sequential monitoring

To perform sequential testing on the propensity score matched aggregated data, the program uses the same likelihood ratio test for continuous and group sequential analysis, as described above for the Self-Controlled Design Tool, which is based on counts of outcomes among the exposure groups. For inputs, the program requires a date range for each monitoring period, which specifies the time period for the assessment. For prospective analyses, start date and end date for each monitoring period are required. The program also requires the same inputs as for the Self-Controlled Design Tool.

D. APPENDIX 4: GROUP SEQUENTIAL GEE (GS GEE) COHORT REGRESSION TOOL

1. Overview

Group Sequential regression using generalized estimating equations (GS GEE) is an approach that can flexibly and robustly accommodate a variety of different exposure and outcome types. Confounding is taken into account through regression adjustment for characteristics measured at baseline and thought to be associated with both the exposure of interest and the health outcome of interest (HOI).

Specifically, like the PROMPT propensity score matching approach, this method is suitable for use in a cohort design where users of an exposure of interest are compared to those who receive an alternative exposure, e.g. an active-comparator new user cohort design, or are unexposed during the same (or concurrent) period of time. Instead of exposure matching on a propensity score, however, regression estimation is used to control for baseline confounders. While the GS GEE methodology can accommodate continuous confounders, data sharing and privacy issues with data partners may require the use of aggregated data, and thus the use of categorical confounding variables. This could involve the inclusion of several categorized individual confounders like age, gender, and comorbidity status or it could adjust for a categorized summary score like a propensity score. Since the method uses a robust GEE framework, it can accommodate a variety of safety outcomes, including outcomes that are binary (i.e., yes or no), count (i.e., number of events), or continuous in nature (e.g., level of a laboratory value). Below we describe two common types of exposure-HOI scenarios that are likely to arise in Mini-Sentinel and discuss how the GS GEE method can be utilized in each instance.

For short term exposures (e.g., a one-time injection or an antibiotic that is used for short period of time), the method involves defining a binary indicator of being exposed or unexposed and a binary occurrence of an HOI within a pre-specified risk window following product initiation. Subjects are not included in a given sequential analysis until the completion of their HOI risk window (i.e., until that subject's risk window has been fully observed) so that all patients have the same follow-up time. A binomial error structure is assumed and combined with a logit link function to estimate an adjusted odds ratio (OR) as the measure of risk in the exposed group relative to the comparator group.

For longer term exposures (e.g., a chronically used drug taken over several months or years), the method defines binary exposure status based on the first record of receipt of either the exposure of interest or the comparator. The observation time, or time at-risk, is calculated as the time from exposure initiation to the first of either exposure discontinuation (with a lag if desired), occurrence of the outcome of interest, or disenrollment. This definition yields a binary outcome indicating whether an HOI occurred during the at-risk period, i.e., the period in which the person was observed to be exposed. For an outcome to be included in a given sequential analysis, it must have occurred before the date of that analysis. Poisson regression is used to estimate an adjusted relative rate (RR) that takes into account duration of time at-risk through an offset term included in the model.

GS GEE can be implemented as a one-time analysis or in a formal sequential monitoring framework with multiple testing and early stopping. If implemented sequentially, the user can specify the total number and frequency of sequential tests as well as the desired level of the signaling threshold over time according to scientific, practical, and statistical preferences. Regardless of the outcome type, exposure type, and sequential design specifications, the GS GEE method calculates a score test statistic and uses this standardized test statistic to inform signaling for elevated risk or continuation of monitoring. This

group sequential test also uses a non-parametric permutation approach that is particularly suited for rare outcomes (i.e., does not rely on large sample theoretical assumptions).

2. Design and data

In order to extract the needed cohort data to implement GS GEE, 12 main program parameters must be specified. These include a request identifier, scenario label, study start date and current analysis day for the query period, age stratifications, an indicator for production of individual level data set, and six input files (Analysis Plan, Exposure Codes, Exposure Categories, Outcome Codes, Outcome Windows, and Age Groups). Three of these parameters are specified by the requester; nine parameters are specified by the request programmer, based on information provided by the requestor. Table 4 contains detailed specifications for each of these required parameters.

Table 4. Main Program Parameters

Parameter	Description
Request Identifier	A prefix added to output log and list files to track the various executions of the program.
Scenario Label	A prefix added to output SAS data sets to track the various executions of the program.
Query Start Date	Date for the start of the query identification period. If Query Start Date is 06Sep2005, only treatment episodes initiated on or after this date will be considered.
Analysis Day	Last day of the query identification period, measured in study days. If Analysis Day=14, only treatment episodes initiated in the 14 days beginning with the Query Start Date will be considered.
Individual Level Creation	Used to specify the optional storage of an individual level data set.
Age Group Definitions	Age group categories for reporting.
Analysis Plan File	Name of the SAS dataset defining the look times of previous through current analyses.
Exposure Codes File	Name of the SAS dataset defining the query exposures of interest. It lists the codes of interest and their corresponding code type (e.g. NDC or PX-C4).

Parameter	Description
Exposure Category File	Name of the SAS dataset mapping exposure categories to a numeric value. For simple analyses the exposure class of interest will be equated with 1 and the comparison group with 0.
Outcome Codes File	Name of the SAS dataset defining the outcomes of interest. It lists the codes of interest, corresponding code type (e.g. DX-09) and an outcome number to enable pulling data for the analysis of multiple outcomes at once.
Outcome Windows File	Name of the SAS dataset defining, for each outcome of interest, the risk and exclusion windows as well as care settings for each outcome definition and exclusion criteria.
Age Groups of Interest	Name of the SAS dataset listing the age strata (defined via the Age Group Definitions parameter) of interest, all strata not listed in this table will be excluded from the final analytic tables.

3. Descriptive analyses

Contents of Final Report with Appendices

- Methods summary
 1. Method: e.g. GS GEE Logistic Regression
 2. Brief Description: e.g. Adjusted logistic regression model applied using GEE framework
 3. Estimate: e.g. Adjusted Odds Ratio
 4. Exposure of Interest: e.g. MMRV
 5. Control Comparator: e.g. MMR+V
 6. Outcome: e.g. Seizure
 7. Confounders: e.g. Age, Sex and Site
 8. Adjusted for Look Time: e.g. No
 9. Look Times(Days): e.g. 364, 455, 546, 637, 728, 819, 910, 1001, 1092, 1183, 1274
 10. Boundary: e.g. Pocock using Unifying Boundary approach

- Demographics and Exposure Uptake
 - Table 1: Demographics and confounders stratified by exposed and unexposed groups at Current Analysis Time
 - Appendix Table A.3: Demographics over each analysis time for exposed and control combined
 - Appendix Table A.4: Demographics over each analysis time within the exposed group
 - Figure 1: Uptake of Exposed and Unexposed groups over time
 - Appendix Figure A.1.X Uptake of Exposure and Unexposed Groups over time for Site X
- Statistical Analysis: Tables will contain different information depending on the type of exposure, e.g., single time (GS GEE Binomial) versus variable length (GS GEE Poisson). Example GS GEE output shown below.
 - Table 2 Results of adjusted analysis by analysis time with the following columns
 1. Look: indicating analysis number
 2. Days: indicating study day when analysis occurred
 3. Exposed total sample size (N) at analysis
 4. Exposed total number of outcomes and unadjusted proportion with outcome at current look
 5. Control total sample size (N) at analysis
 6. Control total number of outcomes and unadjusted proportion with outcome at each analysis
 7. Adjusted OR from GS GEE Binomial approach at each analysis
 8. Score Test Statistic at each analysis
 9. Critical Boundary from GS GEE Binomial approach using simulation at each analysis
 10. Signal (Yes/No) indicating if the Score Test Statistic is greater than the Critical Boundary
 - Appendix Table A.1: Outcome counts and incidence rates by look number, overall and within confounder strata

- Appendix Table A.2: Outcome counts and incidence rates in the exposed group by look number, overall and within confounder strata
- Internal Diagnostics (SAS output files):
 - SAS output of regression with and without including exposure status. Not to obtain adjusted OR or RR one uses the regression model including exposure status, but to derive the score test statistic one only uses the model without including exposure status.
 - Analysis look plan by site. If a site does not have adequate exposure early on or we are not able to obtain their data until a later analysis (e.g. the first time site X is included in the analysis due to inadequate exposure uptake was at analysis 3) we need to keep this both internally to create the correct sequential monitoring boundaries, but also may be informative if the risk estimate has changed strangely at a given analysis time

4. Sequential monitoring

GS GEE can be implemented as a one-time analysis or in a formal sequential monitoring framework where routine testing is desired and is accounted for statistically. The hypothesis being tested depends upon the type of exposure and HOI and the corresponding regression model used. For example, for a binary outcome and logistic regression model, the null hypothesis (H_0) based on the estimated OR comparing the independent exposure and comparator groups is $H_0: OR=1$. At each analysis, the null hypothesis is evaluated via the computation of a standardized score test statistic which is then compared to a pre-specified threshold to determine whether there is a signal for elevated risk, or whether monitoring should continue.

To incorporate group sequential monitoring, GS GEE uses a non-parametric permutation approach that is particularly suited for rare outcomes. Specifically, it flexibly simulates data under the null hypothesis of no difference between exposure groups (e.g., $H_0: OR=1$ for logistic regression, $H_0: RR=1$ for Poisson regression). It uses the unifying boundary approach and defines the boundary based on the permuted data, incorporating the probability of stopping at earlier analysis times and type I error inflation due to repeated testing. The user can select a pre-specified number of analysis times, timing of analyses (based on observed, or expected sample size, at each analysis time), and a total expected maximum sample size by the end of the assessment based on scientific, practical, and statistical preferences. Boundary shape is also user-specified and so can flexibly handle a number of scenarios. For example, a flatter boundary (e.g., a boundary with a constant signaling threshold over time) will, on average, signal earlier for lower elevated risk than a boundary that is more conservative; that is, it requires a stronger effect to signal at earlier analysis times. However, given the same sample size, a flatter boundary will have less power to signal later on during the surveillance period compared with boundaries that employ early conservatism by having a higher signaling threshold at earlier tests. The boundary values for GS GEE are based on the standardized test statistic (as opposed to an error-spending or alpha-spending scale). Therefore, signal decision rules can be planned directly on the standardized scale of the risk quantity of interest rather than the alpha scale, and thus readily facilitate straightforward sequential design decision-making. In all, the following sequential analysis parameters must be specified: shape of the boundary (Pocock, O'Brien Fleming), planned testing frequency (e.g., 12 looks with the first look after 10,000 observations and then evenly spaced looks after that point or 12 looks with the first look after 1 year and quarterly looks after that), and total maximum sample size at end of study. Once specified, the signaling boundary at

each analysis time point can be computed based both on these input parameters as well as the permuted score test statistic under the null.

E. APPENDIX 5: GROUP SEQUENTIAL INVERSE PROBABILITY OF TREATMENT WEIGHTING (GS IPTW) REGRESSION TOOL

1. Overview

Group Sequential regression using IPTW (GS IPTW) is a flexible approach for new user cohorts with short term exposures. The method is designed to perform site-stratified, inverse probability of treatment weighted (IPTW) estimation and group sequential testing in a distributed data setting where the quantity of interest is the overall adjusted risk difference (RD). Adjustment for confounding is achieved through the use of site-specific propensity score models that use pre-specified confounders to predict the probability of exposure to either the exposure of interest or the comparator. An adjusted RD is then estimated where balance in baseline confounder distributions between exposure groups is achieved through the weighting of each observation (or strata) by the inverse of the predicted probability of exposure. The variance of the risk difference is also calculated and corrected to account for variability in the estimation of the IPT weights. The site-specific adjusted RD and variance estimates are sent to a central location where they are combined to provide a single overall adjusted risk difference estimate and variance. Given the overall risk difference estimate and variance a standardized test statistic is calculated ($RD/\sqrt{\text{var}(RD)}$) and used to inform signaling for elevated risk or continuation of monitoring. The advantage of this approach is that it strongly controls for site confounding and has been shown to be as efficient as a non-stratified estimate when no site-level confounding exists. Additionally, only one event is required to be able to estimate a risk difference, making it well suited to a rare event scenario.

2. Design and data

There are 12 main program parameters that must be specified. These include a request identifier, scenario label, study start date and current analysis day for the query period, age stratifications, an indicator for production of individual level data set, and six input files (Analysis Plan, Exposure Codes, Exposure Categories, Outcome Codes, Outcome Windows, and Age Groups). Three of these parameters are specified by the requester; nine parameters are specified by the request programmer, based on information provided by the requestor. Table 5 contains detailed specifications for each of these required parameters.

Table 5. Main Program Parameter Specifications

Parameter	Description
Request Identifier	A prefix added to output log and list files to track the various executions of the program.
Scenario Label	A prefix added to output SAS data sets to track the various executions of the program.
Query Start Date	Date for the start of the query identification period. If Query Start Date =06Sep2005, only treatment episodes initiated on or after this date will be considered.

Parameter	Description
Analysis Day	Last day of the query identification period, measured in study days. If Analysis Day=14, only treatment episodes initiated in the 14 days beginning with Query Start Date will be considered.
Individual Level Creation	Used to specify the optional storage of an individual level data set.
Age Group Definitions	Age group categories for reporting.
Analysis Plan File	Name of the SAS dataset defining the look times of previous through current analyses.
Exposure Codes File	Name of the SAS dataset defining the query exposures of interest. It lists the codes of interest and their corresponding code type (e.g. NDC or PX-C4).
Exposure Category File	Name of the SAS dataset mapping exposure categories to a numeric value. For simple analyses the exposure class of interest will be equated with 1 and the comparison group with 0.
Outcome Codes File	Name of the SAS dataset defining the outcomes of interest. It lists the codes of interest, corresponding code type (e.g. DX-09) and an outcome number to enable pulling data for the analysis of multiple outcomes at once.
Outcome Windows File	Name of the SAS dataset defining, for each outcome of interest, the risk and exclusion windows as well as care settings for each outcome definition and exclusion criteria
Age Groups of Interest	Name of the SAS dataset listing the age strata (defined via the Age Group Definitions parameter) of interest, all strata not listed in this table will be excluded from the final analytic tables.

INPUTFILE SPECIFICATIONS

In addition to the twelve main program parameters, several required parameters must be specified in the Analysis Plan, Exposure Codes, Exposure Categories, Outcome Codes, Outcome Windows, and Age Groups files.

3. Descriptive and statistical analyses

- Demographics and Exposure Uptake
 - Table 1: Demographics and confounders stratified by exposed and unexposed groups at current analysis time
 - Figure 1: Uptake of Exposed and Unexposed groups over time
 - Appendix Table A.3: Demographics over each analysis time for both exposed and control combined
 - Appendix Table A.4: Demographics over each analysis time within the exposed group
 - Appendix Figure A.1.X Uptake of Exposure and Unexposed Groups over time for Site X

STATISTICAL ANALYSIS:

- Table 2: Results of adjusted analysis by analysis time with the following columns
 1. Look: indicating analysis number
 2. Days: indicating study day when analysis occurred
 3. Exposed total sample size (N) at analysis
 4. Exposed total number of outcomes and unadjusted proportion with outcome at analysis
 5. Exposed adjusted risk difference estimate using IPTW
 6. Control total sample size (N) at analysis
 7. Control total number of outcomes and unadjusted proportion with outcome at each analysis
 8. Control adjusted risk estimate using IPTW
 9. Adjusted Risk Difference stratified IPTW results at each analysis
 10. Stratified Test Statistic at each analysis
 11. Critical Boundary from GS IPTW approach using simulation at each analysis
 12. Signal (Yes/No) indicating if the Score Test Statistic is greater than the Critical Boundary
- Appendix Table A1: Current analysis by site (Similar to Analysis Table 2 except only stratify analysis by site instead of analysis time)

- Appendix Table A2: Current analysis by subgroup (Similar to Analysis Table 1 except only stratify analysis by subgroup instead of analysis time)
- Appendix Tables A3(a-c): Repeat Analysis Table 1, but trim weights at the 99%-tile, 95%-tile, and 90%-tile
- Appendix Figure A.2.X: Histogram of propensity scores stratified by exposure group for each site 1 through X
- Appendix Figure A.3.X: Histogram of IPT weights stratified by exposure group for each site 1 through X
- Internal Diagnostics (SAS output files)
 - SAS output of site specific logistic regression models to form the propensity scores to assess any model fit issues or different predictors of exposure by site
 - Analysis look plan by site. If a site does not have adequate exposure early on or we are not able to obtain their data until a later analysis (e.g. the first time site X is included in the analysis due to inadequate exposure uptake was at analysis 3) we need to keep this both internally to create the correct sequential monitoring boundaries, but also may be informative if the risk estimate has changed strangely at a given analysis time.

4. Sequential monitoring

To incorporate group sequential monitoring, GS IPTW uses a non-parametric permutation approach, that is particularly suited for rare outcomes, by flexibly simulating data under the null hypothesis of no elevated risk in the exposure group (i.e. $H_0: RD=0$). It uses an error spending approach to define the boundary based on the permuted data, thus incorporating the concepts of both stopping at earlier analysis times and repeated testing. It requires a specified number of analysis times, timing of analyses (based on observed, or expected sample size, at each analysis time), and a total expected sample size by the end of the assessment. Specifically, the method assumes a cumulative alpha or type I error, $\alpha(t)$, defined as the cumulative amount of type I error spent at analysis t and all previous analyses, $1, \dots, t-1$. It assumes that $0 < \alpha(1) \leq \dots \leq \alpha(T) = \alpha$, where α is the overall type I error to be spent across the evaluation period. The function $\alpha(t)$ can be any increasing monotonic function that preserves family-wise error, but for this program it can be an approximately flat boundary on the standardized test statistic scale (Pocock boundary function) or a higher boundary at the beginning that decreases over time on the order of $\sqrt{n(t)}$, which is a common boundary for efficacy in clinical trials (an O'Brien-Fleming boundary function). Boundary shape is dependent on desired signaling thresholds occurring earlier versus later in an assessment. Flatter boundaries will signal earlier for lower elevated risk, but will have less power to signal later in the assessment relative to boundaries that have a higher threshold to signal early in the assessment yielding more power to signal later. Boundary shape is chosen based on what decision rules are desired (i.e. low boundaries early are only desirable if an action is to be taken for a given signal) in combination with statistical criteria. Given the error spending function and permuted standardized stratified risk difference test statistics under the null, we use this error spending boundary to yield a standardized test statistic boundary that will be the stopping boundary for indicating whether there is an elevated risk or whether the assessment is to continue.

F. APPENDIX 6: LIBRARY OF CORE CONFOUNDERS

Potential core confounders are shown in Table 6. These are confounders that one would likely control for in each evaluation. Yet, the inclusion of each covariate in a multivariate model should be carefully considered given the medical product-health outcome of interest (HOI) pair being examined, as not all core confounders may be relevant for a particular evaluation.

Table 6. Potential Core Confounders

Demographics
Age ⁱ
Sex ⁱ
Calendar time*
Data Partner*
Healthcare utilization in baseline period
of visits to emergency departments ⁱ
of ambulatory visits ⁱ
of hospitalizations ⁱ
of distinct drugs ordered/dispensed ⁱ
of prescriptions ordered/dispensed ⁱ
Lifestyle Factors
Smoking ^{**i} , per algorithm developed by the Mini-Sentinel “15 Cohorts” workgroup
Body mass index ^{**i} , if available in the common data model; otherwise, per algorithm developed by the Mini-Sentinel “15 Cohorts” workgroup
Combined Charlson-Elixhauser comorbidity index ⁱⁱ

*Covariate requiring special consideration given the sequential nature of planned analyses

**Discreetly-captured data field not currently in the Mini Sentinel Common Data Model, therefore alternate diagnosis-based algorithm suggested

ⁱToh S, Garcia Rodriguez LA, Hernan MA. *Pharmacoepidemiol Drug Saf.* 2011 Aug;20(8):849-57.

ⁱⁱGagne JJ et al. *Journal of Clinical Epidemiology.* 2011;64:749-759.

Notes:

- The combined comorbidity index is expected to be applied as a total score with the weights that have been validated as a predictor of mortality. If sample size is sufficient, it may be helpful to include the individual comorbidity flags rather than the overall score, particularly since the weights may vary depending on the outcome.
 - Alcohol use disorder is included in the combined comorbidity index, so was removed from the original list of core confounders since it will not be considered separately.
 - Race/ethnicity was considered but excluded since it is not available for most Data Partners.
 - Nursing home residence status was considered, but thought not to be available in the common data model; yet, consider referencing the final report from the “15 Cohorts” workgroup for discussion of codes that may be able to identify persons ever having been in a nursing home.
 - Zip code removed. Discussed region or other higher level category, but overall felt this would not necessarily be helpful, and would add variance to models.
 - Some Data Partners will have BMI, while others will not.
-