MINI-SENTINEL PROSPECTIVE ROUTINE OBSERVATIONAL MONITORING PROGRAM TOOL: COHORT MATCHING

Technical Users' Guide version: 1.0

Prepared by: Joshua Gagne, PharmD, ScD,[1] Shirley Wang, PhD, ScM,[1] Sebastian Schneeweiss, MD, ScD[1]

Author Affiliations: 1. Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham & Women's Hospital and Harvard Medical School, Boston, MA

January 13, 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 Prospective Routine Observational Monitoring Program Tool: Cohort Matching

Technical Users' Guide version: 1.0

Table of Contents

I. OVERVIEW

The Cohort Matching PROMPT (Prospective Routine Observational Monitoring Program Tool) module comprises a suite of flexible, scalable, modular SAS macros that perform effect estimation in a distributed data setting based on propensity-score matched sequential new user cohorts. The base modular program is the Mini-Sentinel Operations Center (MSOC) produced Modular Program 3 (MP3), a program which creates cohorts of new users for one or more drug/procedures of interest and indicators of incident events. MP3 has detailed documentation available on the Mini-Sentinel Website (http://www.mini-sentinel.org/data_activities/modular_programs/details.aspx?ID=111).

The adjustment component of the Cohort Matching PROMPT takes the cohort of new users identified by MP3 and creates indicators for predefined covariate conditions, procedures, and concomitant drug exposures (defined through diagnosis codes, procedure codes and prescription medication claims) and calculates a combined Charlson-Elixhauser comorbidity score. The program leverages macros from the Pharmacoepi Toolbox, including the high dimensional propensity score (hdPS) macro to identify empirical confounders as well as a macro for fixed and variable ratio propensity score matching. These macros are used to calculate three propensity scores using: (1) predefined covariates only; (2) predefined covariates and empirically (i.e., hdPS) selected covariates; and (3) empirically selected covariates only. New users of compared drug/procedures are nearest neighbor matched in 1:1 and 1:100 ratios with 3 calipers (.01, .025, .05 on the propensity score scale) for each propensity score. Detailed documentation for the Pharmacoepi Toolbox macros can be found at: http://www.hdpharmacoepi.org/download/. In sequential monitoring scenarios, matching occurs within each monitoring period and individuals matched in prior monitoring periods remain in their original matches throughout subsequent periods.

The Cohort Matching PROMPT allows users to define “as treated” (AT) or “intention to treat” (ITT) event indicators and follow-up time using the MP3 algorithm for defining events or using a flexible Health Outcome of Interest (HOI) macro with a more complex algorithm for defining events.

There are Cohort Matching PROMPT outputs datasets that remain behind Data Partner firewalls and others that are returned to the MSOC. Both datasets include indicators for drug/procedure exposure, event indicators, follow up time, three propensity scores, and the identifiers of matched sets for each caliper of matching. Output datasets that remain behind Data Partner firewalls have additional variables, including patient identifiers, predefined covariate indicators and empirically defined covariates. The program includes the option to return selected predefined covariate indicators in the MSOC dataset for use in subgroup analyses.

The Table Creator component of the module takes the output data from the adjustment component and uses generates Tables 1 and propensity score distribution figures. These outputs are useful for diagnostic purposes and assessing covariate balance.

Log files from each of the modular components are saved in a folder with the adjustment component output dataset for the MSOC, diagnostic tables, and a dataset with information on run times for each module and returned to the MSOC for review.

The analysis and aggregation component can be used by the MSOC to aggregate results from data returned by multiple Data Partners. This component generates sequential unadjusted and PS-adjusted estimates of rate differences, hazard ratios, number needed to treat, attributable risk, and population attributable risk. PS-adjusted estimates based on 1:1 and 1:100 variable matching ratios are provided as well as decile specific estimates. When subgroup analysis is requested, the macro will output summary and strata specific estimates for predefined strata as well. The results are output to an .xls file.

The Cohort Matching PROMPT also outputs a file containing the necessary data for input into the R code (developed by Martin Kulldorff and colleagues) to perform the maximized sequential probability ratio test based on the 1:1 matched cohort.

II. PARAMETER SPECIFICATIONS

Table 1: 00 _master.sas - Environment Variables Defined by Data Partner		
Variable Name	**Short Description**	**Long Description**
DPID	Data Partner ID	Enter the two character partner ID.
SITEID	Site ID of Data Partner	Enter the two character Site ID.
ENRTABLE	Enrollment Table Name	Enter the name of the MSCDM Enrollment table.
DEMTABLE	Demographics Table Name	Enter the name of the MSCDM Demographics table.
DISTABLE	Dispensing Table Name	Enter the name of the MSCDM Dispensing table.
DIATABLE	Diagnosis Table Name	Enter the name of the MSCDM Diagnosis table.
PROCTABLE	Procedures Table Name	Enter the name of the MSCDM Procedures table.
ENCTABLE	Encounter Table Name	Enter the name of the MSCDM Encounter table.
DEATABLE	Death Table Name	Enter the name of the MSCDM Death table.
VITTABLE	Vitals Table Name	Enter the name of the MSCDM Vitals table.
INFOLDER	Input file folder	Enter the path where the input files will be saved.
MSOC	Output file folder	Enter the path where the shared output tables will be saved.
DPLOCAL	Dataset file folder	Enter the path where the local SAS datasets will be saved.
INDATA	Libname of the MSCDM	Enter the path where the MSCDM data is saved.
tempHDPS	Libname of temporary hdPS files	Enter the path where the temporary hdPS files will be stored. **Note: this path cannot have any spaces in the path name.** **Example:** C:\hdpsprograms\tempHDPSfiles\
TOOLBOX	Libname of pharmacoepi toolbox macros	Enter the path where the pharmacoepi toolbox macros are saved. **Note: this path cannot have any spaces in the path name.** **Example:** C:\hdpsprograms\hdpsSASmacros\
QUERY	Libname for Modular Programs	Enter the path were the Modular Programs are stored.

Table 2: Prospective Macro Inputs (Modify in 00_master.sas)		
Parameter	**Variable Name**	**Description**
Monitoring Period	look	**Details:** look is a macro parameter for the macro prospective. User inputs value for the monitoring period they desire to query (ie. 1 to X). In order to query monitoring period look, there must be start and end dates entered for that monitoring period within the Monitoring

Table 2: Prospective Macro Inputs (Modify in 00_master.sas)		
Parameter	**Variable Name**	**Description**
		Period Dates.sas program. **Defined by:** Requester (Data Partner may modify) **Input type:** Required (cannot be left blank, for at least one monitoring period) **Format:** Numeric (positive whole numbers) **Example:** look=1

Table 3: 01 _monitoring.sas		
Parameter	**Variable Name**	**Description**
Monitoring Period	look	**Details:** Look corresponds to the sequential monitoring period that the user desires to specify within the prospective macro of the Call Modular Programs.sas program. Allows requester the flexibility to query several monitoring periods with different start and end dates. **Defined by:** Requester **Input type:** Required (cannot be left blank, for at least one monitoring period) **Format:** Numeric (positive whole numbers) **Example:** look=1
Follow up Start Date	Startfollowup	**Details:** startfollowup is a global macro variable which defines the start date for a specific monitoring period. **Defined by:** Requester **Input type:** Required (cannot be left blank, for at least one monitoring period) **Format:** mm/dd/yyyy **Example:** startfollowup=08/15/2004
Follow up End Date	ENDDATE	**Details:** ENDDATE is a global macro variable which defines the end date for a specific monitoring period. **Defined by:** Requester **Input type:** Required (cannot be left blank, for at least one monitoring period) **Format:** mm/dd/yyyy **Example:** enddate=08/15/2004

Table 4: 03_modular_adjustment_query Inputs (Modify in 00 _master.sas)		
Parameter	**Variable Name**	**Description**
Request Identifier	REQUESTID	**Details**: REQUESTID is a prefix added to all output files (along with DPID and SITEID) to track the various executions of the program (Refer to Table 2 for definitions of DPID and SITEID). **It is assigned by the MSOC** and **cannot exceed 8 characters**. **Defined by:** Requester **Input type:** Required (cannot be left blank) **Format:** Alphanumeric **Example**: REQUESTID=mpr25_r1
Query File	queryfile	**Details:** Identical to MP3 parameter. Name of the SAS dataset defining the query exposure(s) of interest. It lists the codes of interest and various parameters to specify how each code must be queried by MP3. The file name, including its extension, must be entered. For specific details on the content of this file, please see MP3 documentation. **Named by:** Requester **Input type:** Required **Format:** .sas7bdat or .cport file format **Example:** QUERYFILE=query.sas7bdat or query.cport
Query Event File	queryeventfile	**Details:** name of the SAS dataset defining the query event(s) of interest. It lists the codes of interest and various parameters to specify how each code must be queried by MP3. The file name, including its extension, must be entered. For specific details on the content of this file, MP3 documentation. **Named by:** Request programmer **Input type:** Required **Format: .**sas7bdat or .cport file format **Example:** QUERYEVENTFILE=event.sas7bdat or event.cport
Enrollment Gap	enrolgap	**Details:** Identical to MP3 parameter. Sets the number of days that will be bridged between two consecutive enrollment periods to create a "continuously enrolled" period. For example, if ENROLGAP=30 and a member is eligible for

Table 4: 03_modular_adjustment_query Inputs (Modify in 00 _master.sas)		
Parameter	**Variable Name**	**Description**
		medical and drug coverage in periods 1/1/2007-3/27/2007 and 4/1/2007-12/21/2007 (*i.e.,* a 4-day gap between two consecutive enrollment episodes), the member will be considered continuously enrolled from 1/1/2007 to 12/21/2007. Any gaps in enrollment greater than 30 days will result in a new enrollment period, and all the days in the gap will be considered un-enrolled. **Note1**: A gap of 45-days is recommended for most uses. **Note2**: Enrollment in both drug and medical benefit coverage is always required for MP3. **Note3:** Multiple continuous enrollment periods per member may be assessed. **Defined by:** Requester **Input type:** Required **Format**: Numeric **Example**: ENROLGAP=45 (gaps less than or equal to 45 days will be "bridged" to form one "continuously enrolled" sequence)
Check data availability	datastartcheck	**Details:** Name of the SAS dataset with dates of data availability for Data Partner. **Defined by:** MSOC **Input type:** Required **Format:** Dates **Example:** Dpid Siteid table startdate a. fa ke enrollment 17324
HOI algorithm macro	useHOImacro	**Details:** Use custom health outcome of interest (HOI) macro with complex algorithm for defining HOI instead of MP3 defined outcome. **Defined by:** Requester **Input type:** Optional **Format:** Numeric **Example:** useHOImacro = 1; (default is to use MP3 defined event)
Age Groups	AGESTRAT	**Details:** Output results by custom age groupings. For example, to have results stratified by 20 year increments in members between 40 and 100,

Table 4: 03_modular_adjustment_query Inputs (Modify in 00 _master.sas)		
Parameter	**Variable Name**	**Description**
		the user will input AGESTRAT=40y-59y 60y-79y 80y-99y. Age for patients/new users is determined at the index date for new use. (See MP3 documentation for more detail) **Defined by:** Requester **Input type:** Required (cannot be left blank) **Format:** AA BB ZZ **Example:** AGESTRAT=40y-59y 60y-79y 80y-99y
Covariate Window	covariatewindow	**Details:** covariatewindow is a global macro variable that specifies the length (in days) of the covariate assessment window. User inputs the number of days for the assessment of covariates prior to the index date. **Defined by:** Requester **Input type:** Required (cannot be left blank) **Format:** Numeric (positive whole numbers) **Example:** covariatewindow=180
Covariate Procedures	covariateproc	**Details:** covariateproc is a global macro variable naming the .sas7bdat dataset that identifies the combination of procedure codes that the requester would like to specify as covariates. Covariates will be assessed within the covariate window prior to the index date. Input file requires a "StudyProc" variable which defines the covariate (no spaces are allowed, i.e. percutaneous_coronary), "px" the procedure code(s) for each covariate and "subgroup" an indicator of whether this covariate should be returned to the MSOC. Note that if this macro variable is blank, predefined covariate procedures will not be part of the predefined propensity score. **Defined by:** Requester **Input type:** Optional **Format:** Alphanumeric **Example:** covariatecondition=covariates
Covariate Conditions	covariatecondition	**Details:** covariatecondition is a global macro variable which names the .sas7bdat dataset that identifies the combination of diagnosis codes that the requester would like to specify as covariates. Covariates will be assessed within the

Table 4: 03_modular_adjustment_query Inputs (Modify in 00 _master.sas)		
Parameter	**Variable Name**	**Description**
		covariate window prior to the index date. Input file requires a "StudyDiag" variable which defines the covariate (no spaces are allowed, ie. coronary_artery_disease), "dx" the ICD9 diagnosis code(s) for each covariate and "subgroup" an indicator of whether this covariate should be returned to the MSOC. Note that if this macro variable is blank, predefined covariate conditions will not be part of the predefined propensity score. **Defined by:** Requester **Input type:** Optional **Format:** Alphanumeric **Example:** covariatecondition=covariates
Covariate Drug Exposures	covariaterx	**Details:** covariaterx is a global macro variable which names the .sas7bdat dataset that identifies the combination of NDC codes that the requester would like to specify as covariates. Covariates will be assessed within the covariate window prior to the index date. Input file requires a "drug_cat" variable that describes the drug category (no spaces are allowed, ie. diabetes_drug), "NDC" which contains the 11 digit NDC codes for the desired medication and "subgroup" an indicator of whether this covariate should be returned to the MSOC. Note if this macro variable is blank, predefined covariate drugs will not be included in the predefined propensity score. **Defined by:** Requester **Input type:** Optional **Format:** Alphanumeric **Example:** covariaterx=drugcovariate
Drugclass	drugclass	**Details:** drugclass is a global macro variable naming a .sas7bdat with an NDC-drugclass crosswalk. **Defined by:** Requester **Input type:** Required (cannot be left blank) **Format:** Alphanumeric **Example:** drugclass = classesofdrug
Comorbidity Score	comorbidscore	**Details:** comorbidscore is a global macro

Table 4: 03_modular_adjustment_query Inputs (Modify in 00 _master.sas)		
Parameter	**Variable Name**	**Description**
		variable naming the .sas7bdat dataset that contains the information for creating combined comorbidity scores. **Defined by:** Requester **Input type:** Required (cannot be left blank) **Format:** Alphanumeric **Example:** comorbidscore=comorbidities
ITT Followup	ITT	**Details:** ITT is a global macro variable that specifies the length (in days) for Intention to Treat (ITT). Analysis will be censored for death or disenrollment. If left blank, the default will be to produce event indicator and follow up time for As Treated analyses. NOTE: If ITT analysis is desired, the exposure extension period specified in the QueryFile for MP3 must be greater than or equal to the number of days to follow up for ITT, the EXPEXTPER must be less than or equal to the MP3 specified EPISODEGAP, and EPISODEGAP must be less than or equal to MP3 specified WASHPER. ITT <=EXPEXTPER<=EPISODEGAP<=WASHPER **Defined by:** Requester **Input type:** Optional **Format:** Numeric (positive whole numbers) **Example:** ITT=180
Covariates Considered	covariates_considered	**Details:** covariates_considered is a global macro variable that specifies the number of variables from each data dimension (diagnoses, procedures and drug classes) that is considered for selection as a covariate. If not specified, default value is 100. **Defined by:** Requester **Input type:** Optional (default values will be used if missing) **Format:** Numeric (positive whole numbers) **Example:** covariates_considered=150
Covariates Selected	covariates_selected	**Details:** covariates_selected is a global macro variable that specifies the total number of empirical covariates that are selected from all the data dimensions combined. These selected

Table 4: 03_modular_adjustment_query Inputs (Modify in 00 _master.sas)		
Parameter	**Variable Name**	**Description**
		covariates are used for the propensity score. If not specified, default value is either 200 or number of new users of study drug, whichever value is the smallest. **Defined by:** Requester **Input type:** Optional (default values will be used if missing) **Format:** Numeric (positive whole numbers) **Example:** covariates_selected=150
Covariate Ranking	ranking	**Details:** ranking is a global macro variable that indicates one of three models for selecting variables. Can take the value of "bias", "exp_assoc" and "outcome_assoc". Specifying "bias" will yield a variable list in which the top k variables are selected as ranked by the Bross bias formula (here k is the number entered in the macro parameter covariates_selected). Specifying "exp_assoc" will yield a variable list in which the variables are selected as ranked by the strength of the relationship between confounder and exposure. This is most suitable for cases where there are fewer than 150 exposed outcomes. Specifying "outcome_assoc" will yield a variable list in which the top k variables are selected as ranked by the strength of the relationship between the confounder and the outcome. This is most suitable for disease risk scores. Default is exposure association. **Defined by:** Requester **Input type:** Optional (default values will be used if missing) **Format:** Character **Example:** ranking="bias"
Health Service Utilization	health_service_intensity	**Details:** health_service_intensity is a global macro variable. If value is 1 then macro will infer the intensity of health service utilization by computing quartile of (a) number of codes per patient, and (b) number of unique codes per patient within each dimension. These quartile indicators are then screened like all other hd-PS variables and allowed to enter the final propensity score. This option may be used in

Table 4: 03_modular_adjustment_query Inputs (Modify in 00 _master.sas)		
Parameter	**Variable Name**	**Description**
		place of investigator-defined variables such as "number of unique medications used" or "number of office visits". Default value is 1, yes. **Defined by:** Requester **Input type:** Optional (default values will be used if missing) **Format:** Binary **Example:** health_service_intensity=1
Zero Correction Screen	zero_cell_corr	**Details:** zero_cell_corr is a global macro variable that indicates whether to screen variables with a zero correction added to each cell in the confounder/outcome 2x2 table. Recommended when the number of exposed outcomes is fewer than 150. Default is 1, yes. **Defined by:** Requester **Input type:** Optional (default values will be used if missing) **Format:** Binary **Example:** zero_cell_corr=1
Netezza available?	nz	**Details:** nz is a global macro variable that indicates whether the platform being used is a Netezza. **Defined by:** Requester **Input type:** Optional (default value of nowill be used if missing) **Format:** Binary **Example:** nz = ; nz = 1;

Table 5: 04_table_creator.sas Inputs (Modify in 00 _master.sas)		
Parameter	**Variable Name**	**Description**
Names of compared exposure/treatments	grp1, grp2	**Details**: Grp1 and Grp2 refer to the names of the primary and referent exposure groups. These must appear exactly as in the "group" variable in the input queryfile for MP3 (capitalization matters!) **Defined by:** Requester **Input type:** Required (cannot be left blank) **Format:** Alphanumeric

Table 5: 04_table_creator.sas Inputs (Modify in 00 _master.sas)		
Parameter	**Variable Name**	**Description**
		Example: grp1 = clindamycin; grp2 = cephalexin;
Input template for Table 1's	templatetable	**Details:** Templatetable refers to the name of an input excel file which contains the template for the layout of table 1's for a specific requestid. **Defined by:** Requester **Input type:** Required (cannot be left blank) if Table 1's are desired. Comment out macro call if Table 1's are not desired. **Format:** Alphanumeric **Example**: templatetable = clindamycintable;

Table 6: 07_sequential_analysis.sas		
Parameter	**Variable Name**	**Description**
Library location of data from contributing Data Partner Sites	MSOC1 MSOC2 MSOC3 Etc.	**Details:** Environment variable. Enter the path where the datasets returned from Data Partner to MSOC are located. If program is run at Data Partner site, enter path where output data is stored. **Defined by:** Requestor or Data Partner **Example:** msoc1 = c:/pathtodata/;
Library location for output of analysis results	output	**Details:** Environment variable. Enter the path where the output .xls file and .sas7bdat dataset from sequential analysis macro should be saved. **Defined by:** Requestor or Data Partner **Example:** output= c:/pathtooutput/;
Library location of monitoring period dates sas program	query	**Details:** Environment variable. Enter the path where 01_requestid_monitoring_period_dates is located. **Defined by:** Requestor or Data Partner **Example:** query = c:/pathtomonitoringdates/;
Contributing Data Partner Ids and Site IDs	DPID1 SITEID1 DPID2 SITEID2 DPID3 SITEID3 Etc.	**Details:** DPID(number) and SITEID(number) refer to the DPID and SITEID from each contributing partner site. **Defined by:** Requester **Input type:** Required (cannot be left blank) for DPID(number) or SITEID(number) <= NumAgg

Table 6: 07_sequential_analysis.sas		
Parameter	**Variable Name**	**Description**
		Format: Alphanumeric **Example**: DPID1 = ae; SITEID1 = os; DPID2 = kp; SITEID2 = nw;
Type of analysis	ITT	**Details:** ITT is an indicator of whether the analysis for this REQUESTID is an As Treated or Intention to Treat. The default if this variable is blank is As Treated. When ITT = 1 this indicates the analysis is Intention to Treat. **Defined by:** Requester **Input type:** Required (cannot be left blank) **Format:** Numeric **Example**: ITT = 1;
First monitoring period to analyze	first	**Details:** First refers to the first sequential monitoring period for which analysis results are requested. Note: this number does not have to be the first monitoring period for which data is available. **Defined by:** Requester **Input type:** Required (cannot be left blank) **Format:** Numeric **Example**: first = 1
Last monitoring period to analyze	period	**Details:** Period refers to the last sequential monitoring period for which analysis results are requested. Note: this number does not have to be the last monitoring period for which data is available. **Defined by:** Requester **Input type:** Required (cannot be left blank) **Format:** Numeric **Example**: period = 5
Additional subgroup analyses	subgroup	**Details:** Subgroup refers to covariate indicators for which the requestor would like additional stratified results. **Defined by:** Requester **Input type:** Optional **Format:** Alphanumeric **Example**: subgroup = %quote(covar1, age_cat)
Which Propensity Score to use in analysis	ps	**Details:** Specifies the propensity score to use for the analysis.

Table 6: 07_sequential_analysis.sas		
Parameter	**Variable Name**	**Description**
		Defined by: Requestor **Input type:** Optional (default is hdpspredefined) **Example:** ps = predefined, ps = hdpspredefined, b. ps = hdpsonly,
Which propensity score matching caliper to use	caliper	**Details:** Specifies the propensity score matching caliper to use for the analysis. **Defined by:** Requestor **Input type:** Optional (default is .025) **Example:** caliper = .01, caliper = .025, c. caliper = .05,

III. PROGRAM STEPS

The following provides a summary of the steps performed by the Cohort Matching PROMPT:

Program Steps:

1. Modular Program 3 identifies new users and outcomes for a pre-specified product-outcome pair within a defined window.
    a. Incident drug/procedure, incident outcome, follow-up period, prior enrollment requirements, washout, inclusion and exclusion criteria, exposure extension, etc. are defined through MP3 options (see MP3 documentation for details)
    b. MP3 performs only an "as treated" analysis; if an "intention to treat" analysis is desired, enter intended follow up window to MP3 option: EXPEXTPER (i.e. if ITT analysis is for 180 days, EXPEXTPER = 180)
2. Adjustment component:
    a. User defines: name of SAS dataset containing dates of data availability for relevant Data Partners; covariate assessment window; names of SAS files with relevant code/groupings for predefined covariate conditions, procedures, and drugs; name of SAS file crosswalk between NDC and drug therapeutic class; name of SAS file with codes/groupings to create combined comorbidity score; whether analysis is "as treated" or "intention to treat"; optional hdPS parameters; and which covariates will be used to define subgroups (if any)
    b. Program extracts unique new user patient ID, index exposure dates, and event dates from MP3 output dataset and keeps first treatment episode for individual patients for either of the compared exposures.
    c. Imports user-created files with codes/groupings of predefined diagnosis, procedure and/or drug covariates and codes for combined comorbidity score
        i. Extracts claims matching codes in imported files from MSCDM formatted data
        ii. Keeps claims that belong to identified new users and fall within covariate assessment window
        iii. Creates covariate indicators for predefined covariates for identified new users
        iv. Calculates combined comorbidity score for identified new users
    d. Assess option to use flexible HOI macro with more complex algorithm for defining event:
        i. If HOI macro option is selected, then program runs macro and replaces MP3 event dates with HOI macro event dates.
    e. Assess macro option – "as treated" (AT) or "intention to treat" (ITT) analysis?
        i. If AT, event date is the first date within the treatment episode window
        ii. If ITT, event dates are censored at end of ITT follow up, death, or disenrollment
    f. Creates input datasets for diagnosis, procedure and drug data dimension to feed into high-dimensional propensity score (hdPS) macro
        i. Input datasets include all claims within the data dimension that fall within covariate assessment window for identified cohort of new users (index date is not included)
    g. Runs hdPS macro
        i. Default options are set to include health service intensity (e.g., number of visits to

emergency department, number of ambulatory visits, number of hospitalizations, number of drug dispensed), zero cell correction (which allows hdPS to estimate covariate-exposure associations when a covariate occurs only in one of the two exposure groups), interactions between monitoring period/time and predefined covariates, exposure based selection of empirical covariates, and creation of up to 200 empirically selected covariates (if the number of new users identified for either study group is lower than 200, the number of empirically selected covariates equals the smaller sample size)

h. Matches patients using 1:1 nearest neighbor matching with three calipers (0.010, 0.025, 0.050 on the propensity score scale) for three calculated propensity scores (predefined covariates only, empirically defined covariates only, and predefined + empirically defined covariates)
    i. For sequential surveillance, new propensity scores are calculated during each monitoring period using eligible new users from each prior monitoring period up to and including the current period
    ii. New users of the product of interest are matched to new users of the comparator product who were identified in the same monitoring period
    iii. Only new users identified in the current period are eligible for matching
    iv. Matches identified during prior monitoring periods are not broken – once a patient is matched, s/he stays matched and if a patient is not matched s/he remains unmatched throughout
i. Outputs dataset to DPLOCAL folder which includes variable indicating monitoring period during which patient was identified, variable with start and end dates of monitoring period, *patient IDs,* index dates, event dates, event indicators, 3 propensity scores, 9 matching identifiers (for 3 PS scores using 3 calipers), *predefined and empirically defined covariates*
    i. Dataset is named with REQUESTID, Data Partner ID, and sequential monitoring period number
    ii. This file remains behind the Data Partner's firewall
j. Outputs dataset to MSOC folder with variable indicating monitoring period during which patient was identified, variable with start and end dates of monitoring period, event indicators, 3 propensity scores, 9 matching identifiers (for 3 PS scores using 3 calipers) – note no patient identifiers are returned to MSOC, no covariate indicators are returned unless otherwise specified for subgroup analyses
    i. Dataset is named with REQUESTID, Data Partner ID, and sequential monitoring period number
k. Outputs dataset with time required to run each component of the modular program

3. Table Creator component
    a. Imports Excel files provided by user to create Tables 1 for predefined covariates
    b. Runs table creator macro that has been tailored to specific REQUESTID by user
        i. Outputs unmatched and 9 matched Tables 1 (3 propensity scores with 3 matching calipers each) to show distribution of exposure, outcome, predefined covariates, absolute and standardized differences, and Mahalanobis distance to Excel file
        ii. Outputs 1 unmatched and 9 matched propensity score distribution figures with c-statistic as insert to .rtf file

4. Analysis and aggregation component
    a. MSOC specifies:
        i. User-assigned REQUESTID
        ii. Number of contributing Data Partners and edits DPID and Site ID
        iii. Analysis type
        iv. Which monitoring period(s) to analyze
        v. Subgroup analyses
        vi. Which propensity score and matching caliper to use (if not using default of hdPS + predefined propensity score with 0.025 matching caliper)
    b. Aggregation module puts together data from each site and outputs .xls file with:
        i. Unadjusted rate differences stratified by Data Partner
        ii. Adjusted rate differences stratified by Data Partner and propensity score matchid
        iii. Adjusted rate differences stratified by Data Partner and deciles of propensity score (reports decile-specific estimates and summary estimate)
        iv. Unadjusted hazard ratio from Cox model stratified by Data Partner
        v. Adjusted hazard ratio from Cox model stratified by Data Partner and propensity score matchid
        vi. Adjusted hazard ratio from Cox model stratified by Data Partner and deciles of propensity score (reports decile-specific estimates and summary estimate)
        vii. Number-needed-to-treat, attributable risk, and population attributable risk
        viii. Wald p-value, beta coefficient, and standard error
    c. For each of items i-vii, module provides analysis results for: full cohort and pre-specified subgroups
    d. Module outputs .csv file with input to an R program which runs sequential alerting algorithm

IV. APPENDIX A: DESCRIPTION OF COHORT MATCHING PROMPT FOR ROUTINE SURVEILLANCE OPERATIONS MANUAL

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., NDCs or 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 (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**). The information requested from each Data Partner meets the minimum necessary standard specified in the Mini-Sentinel Principles and Policies (**http://mini-sentinel.org/work_products/About_Us/Mini-Sentinel-Principles-and-Policies.pdf**).

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	*n*	1	0	45	0.051	*j*	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 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-Control Design Tool. Whereas the Cox model provides effect estimates based on a survival analysis, the sequential test 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-Control Design Tool.

V. APPENDIX B: EXAMPLE WORK PLAN FROM BETA TESTING

MINI-SENTINEL WORKPLAN

Date submitted: May 20, 2013

Project Name: 4.10 Module 2: Propensity Score Matching, Diagnostics and Analysis
Mini-Sentinel Request ID: p_wp2_b2

Overall Project Objective: Mini-Sentinel requires the ability to adjust for confounding in rapid assessments. Propensity scores are a particularly useful confounder adjustment technique for between-person comparisons in a distributed data setting as they simultaneously facilitate adjustment for many confounders and preserve data confidentiality. These data requests are for beta testing of the Module 2 suite of modular programs. In this aim we will evaluate 2 sequential monitoring scenarios, clindamycin vs. cephalexin and risk of acute Myocardial Infarction within 30 days (requestid: to08_hdps_wp1_b1); and lisinopril versus beta blockers and risk of angioedema within 30 days (this workplan, with requestid: to08_hdps_wp2_b1). The time window for this data request is January 2009 through December 2010.

Instructions:

Each distributed package will have the same layout of folders and sub-folders in a zip file.
There are many macros and modular programs contained in the distributed package. The program is called by a master sas program that is named: 00_to08_p_wp2_b2_master.sas (in the sasprograms subdirectory).

Regardless of the scenario, to run the package, each Data Partner only needs to open the master sas program and enter site specific information according to the instructions in **sections 1-4**.
These sections are also noted in /* comments */ in the master sas program, as named above.

Please follow the instructions in the header section of the relevant master SAS program. Note that end of path separators are required for all library path names. For example, "C:\user\sas\" and not "C:\user\sas" on Windows based platforms and "/home/user/sas/" and not "/home/user/sas" on UNIX platforms. Additionally, the library pathname to the Toolbox folder must not have spaces in the path.

After entering the site specific information, run the master program. Logs will be printed to the MSOC folder designated by the Data Partner. In the event of problems with running the package, please return the logs to the MSOC for diagnosis and debugging.

When done please:
1) zip/compress the MSOC output folder (file name should be **"&dpid.&sited._to08_hdps_wp2_b2.zip**")
2) upload it to the Mini-Sentinel secure portal (https://portal.mini-sentinel.org) in your site's Private section (Private/**DPID+SiteID**/), and
3) notify the MSOC

The instructions in the header section of the master call program are also documented below:

DATA PARTNERS SECTION 1)

Please Edit DPID and Site ID according to the table below

%let DPID=;
%let SITEID=;

```
/*-------------------------------------------------------------------*\
||                                                                   ||
||  DATA PARTNERS                                    DpID  SiteId    ||
||  ---------------------------------------------------------------- ||
||  Healthcore (one site)                            HC    OS        ||
||  Humana (one site)                                HU    OS        ||
||  Vanderbilt University (one site)                 VB    OS        ||
||  Aetna (one site)                                 AE    OS        ||
||  OptumInsight (one site)                          OP    OS        ||
||  HMORN (7 sites)                                                  ||
||     Group Health Cooperative                      HM    GHC       ||
||     Fallon Community Health Plan                  HM    MPCI      ||
||     Henry Ford Health System                      HM    HFHS      ||
||     Lovelace Health System                        HM    LCF       ||
||     Marshfield Clinic                             HM    MCRF      ||
||     HealthPartners                                HM    HPRF      ||
||     Harvard Pilgrim Health Care                   HM    HPHC      ||
||     Kaiser Permanente (6 sites)                                    ||
||          Kaiser Permanente Colorado               KP    CO        ||
||          Kaiser Permanente Georgia                KP    GA        ||
||          Kaiser Permanente Hawaii                 KP    HI        ||
||          Kaiser Permanente Northern California    KP    NC        ||
||          Kaiser Permanente Northwest              KP    NW        ||
||          Kaiser Permanente Mid Atlantic           KP    MA        ||
\*-------------------------------------------------------------------*/
```

DATA PARTNERS SECTION 2)

Please Edit this section to reflect your name for each Table/File (or View)

%let ENRTABLE=Enrollment;
%let DEMTABLE=Demographic;
%let DISTABLE=Dispensing;
%let DIATABLE=Diagnosis;
%let ENCTABLE=Encounter;
%let PROCTABLE=Procedure;
%let DEATABLE=Death;
%let VITTABLE = Vitals;

DATA PARTNERS SECTION 3)

Please edit paths a-g to reflect library and folder locations for Mini-Sentinel
IMPORTANT NOTE! END OF PATH SEPARATORS ARE REQUIRED!!

a. Data in MSCDM Format libname indata "c:/path/";

b. NDC/ICD9 Codes File Location	%let infolder= c:/path/inputfiles/;
c. MSOC/CSV Output Files	%let msoc= c:/path/msoc/;
d. DPlocal/CSV Output Files	%let dplocal= c:/path/dplocal/;
e. Temporary hdPS files	%let temphdps = c:/path/hdps/temphdps/;
f. Location of toolbox macros	%let toolbox = c:/path/hdps/toolbox/;
g. MSOC modular programs	%let query =c:/path/sasprograms/;

DATA PARTNER SECTION 4)

You may edit %prospective(look =); to reflect the relevant sequential monitoring period.

```
/* To run query for monitoring period 1 */      %PROSPECTIVE(LOOK = 1);
/* To run query for monitoring period 2 */      %PROSPECTIVE(LOOK = 2);
/* To run query for monitoring period 3 */      %PROSPECTIVE(LOOK = 3);
/*etc.*/
```

Workplan Timeline: Please complete by May 25, 2013.

The Module 2 packages contain the following documents:

1) 8 Modular SAS Programs:

00_to08_p_wp2_b2_master.sas
01_to08_p_wp2_b2_monitoring.sas
02_mp3_3.0_beta3
03_modular_adjustment_query
04_table_Creator
06_HOI_myocardial_infarction
07_sequential_analysis

2) Contents of Pharmacoepidemiology Toolbox and MSOC utility macros

Hdmacros.sas
Java_utils.sas
Matching.sas
Ms_agestrat.sas
Ms_cci.sas
Ms_cci_pre.sas
Ms_createepisodes.sas
Ms_envelneeded.sas
Ms_envelope.sas
Ms_episoderec.sas
MS_freezedata.sas
Ms_getdrugs.sas
Ms_getmedical.sas
Ms_lockdata.sas
Ms_makemacvar.sas
Ms_util.sas
Ms_withinelig.sas

Ms_wordcount.sas
Netezza_code_templates.sas
Pharmacoepi.jar
Table_creator.sas
Utils.sas

3) Input files

angioedema.sas7bdat
ccifile.sas7bdat
comorbidscore.sas7bdat
drugclass.sas7bdat
drugcovariate_lisin.sas7bdat
lisin_covariatedx.sas7bdat
lisin_covariatepx.sas7bdat
lisin_studydrug.sas7bdat
med_ufile.sas7bdat
lisinopril.xls

3) This workplan and Module 2 documentation:

mini_sentinel_to08_hdps_wp2_b2_workplan.doc
module2_documentation _2013_04_05.doc

Timeframe for Data to be Included:

January 1, 2009 – December 31, 2010

MSCDM Files Accessed:

All MSCDM files will be accessed

Output Files:

For each monitoring period that is run, this program generates: 1) 2 SAS datasets, 2) 5 log files, 3) 2 .xls files, 4) 1 .rtf file as well as assorted .lst and signature files to the msoc subdirectory. Please return all files in the msoc subdirectory to the MSOC.

This program also creates datasets and .csv files saved into the dplocal subdirectory. These files are to remain with the Data Partner and retained for future runs of to08_hdps_wp2_b2.