SENTINEL REUSABLE PROGRAMS

TreeExtraction Program

Prepared by the Sentinel Operations Center
August 17, 2016
Version: 1.3

Sentinel is sponsored by the U.S. Food and Drug Administration (FDA) to monitor the safety of FDA-regulated medical products. Sentinel is one piece of the Sentinel Initiative, a multi-faceted effort by the FDA to develop a national electronic system that complements previously existing methods of safety surveillance. Sentinel Collaborators include Data and Academic Partners that provide access to health care data and ongoing scientific, technical, methodological, and organizational expertise. The Sentinel Coordinating Center is funded by the FDA through the Department of Health and Human Services (HHS) Contract number HHSF223201400030I.

Modification History

Version	Date	Modification	By
1.1	3/30/2015	• Original Program	Sentinel Operations Center
1.2	6/30/2016	• Changed structure of input files, creating more of them with individual functionality • Added new variable (*exp_wash_up*) and feature – incident exposure based on a set of number of days • Removed variable (*enr_ramp_up*), which represented enrollment washout. Enrollment washout is now determined in the program based on a combination of other variables • Added the ability to collect data for multiple episodes of exposure • Added exposure exclusion feature that creates pre-exposure exclusions relative to the index date	Sentinel Operations Center
1.3	8/17/2016	• Changed variable names to be less specific to vaccine work. • Added age stratification and incident level to results file • Added ability to use ICD-9-CM only or ICD-9-CM and ICD-10-CM together • Added cohort attrition table output • Modified exclusion options • Modified results output so more scenarios can be included within the same execution of the program. • Added *exclusion_distance* variable to Exclusion Exposure Information input file, then used value to create a washout period for exclusions	Sentinel Operations Center

Table of Contents

I. INTRODUCTION

Tree-based scan statistics are a statistical approach for vaccine and drug safety surveillance that detect excess risk in electronic health data that have been grouped into hierarchal tree structures. Using this technique, one evaluates a large number of unsuspected but potential medical product-associated health outcomes of interest (HOIs). It is sometimes referred to as a data-mining technique because it simultaneously evaluates several thousand potential exposure-HOI pairs, adjusting for the multiple testing inherent in the many pairs evaluated.

TreeExtraction is a reusable cohort extraction program designed to create an analytic dataset based on the self-controlled risk interval (SCRI) design. The analytic dataset created from the TreeExtraction program is analyzed using TreeScan™ Software (http://www.treescan.org), which implements tree-based scan statistics. Currently, TreeExtraction will consider the 1:N model of exposure-HOI pairs. That is, there will be 1 defined exposure and N HOIs for a total of N potential exposure-HOI pairs. The upper limit of HOIs (N) will be defined by a requester-supplied hierarchal tree of HOIs (input file: DXTREE_INPUT.sas7bdat).

HOIs will be identified and defined using ICD-9/10-CM diagnosis codes, which are organized into a hierarchical tree structure. A limitation of this version of the TreeExtraction program is that the tree structure is limited to a 5-level tree (i.e., 4 levels plus the terminal leaf level or ICD-9-CM code level). ICD-10-CM codes will be converted into their ICD-9-CM equivalent for use in the tree. Each node in the structure is identified by a string. An example tree is below where a group of ICD-9/10-CM codes are grouped at the 3rd level of the tree (i.e., 06.04.02).

Table 1. Example Hierarchical Tree Structure

Node	Level	ICD-9 / 10	Description
06	1st		Diseases Of The Nervous System And Sense Organs
06.04	2nd		..Epilepsy; convulsions
06.04.02	3rd		……Convulsions
06.04.02.00	4th	780.3 / R56	……..Convulsions
		780.31 / R56.0 or R56.00	……..Febrile convulsions
		780.32 / R56.01	……..Complex febrile convulsions
		780.33 / R56.1	……..Post traumatic seizures
		780.39 / R56.9	……..Other convulsions

Generally, the goal of the program is to count qualifying exposure-HOI pairs per the SCRI design at each node in the tree, and record the time-to-event for each pair.

II. PROGRAM PACKAGE

The Sentinel Operations Center (SOC) uses a uniform folder structure to facilitate communications between SOC and Data Partners and to streamline file management. This section describes the program package structure and requirements for TreeExtraction program package execution.

The program package structure and contents of each folder are listed below.

- *sasprograms*: folder contains the master SAS program that must be edited and then executed by the Data Partner.
- *inputfiles*: folder contains input files and lookup tables needed to execute a request. Input files contain parameter values specific to a particular request (e.g., medical product exposures of interest, continuous enrollment requirements, incidence criteria, etc.). Input files are created for each request by the SOC request programmer; the contents of this folder are not edited by the Data Partner. The folder also contains one subfolder: *macros*, explained next.
    - *macros*: folder contains the macros that comprise the modular program. The contents of this folder are not edited by the Data Partner.
- *msoc*: folder contains output generated by the request that should be sent to SOC.
- *dplocal*: folder contains output generated by the request that should remain with the Data Partner (and may be used to facilitate follow-up queries).

A. COMMON COMPONENTS

Prior to executing the request package, a set of SAS programs known as common components must be initialized. In this context, common components refer to a set of SAS programs that provide appropriate site-specific attributes (e.g., Data Partner description variables, Sentinel Common Data Model (SCDM) table names, folder paths, data completeness dates, etc.) to distributed SAS program packages at the time of code execution. More specifically, when an executing SAS program package accesses the file ms_common_components.sas, global macro variable definitions for key site-specific attributes are made available to the calling program. In this context, common components support two important goals: 1) streamline the setup for the distributed SAS program packages, 2) improve the accuracy of results.

Users must specify the location of their common components file path in the master SAS program in the *sasprograms* folder in order for the package to execute. For more information about common components installation, and to download the SAS programs, visit the common components page on the Sentinel website.

B. NAMING CONVENTIONS

Each request distributed by the SOC is uniquely named using a standard, meaningfully unique request identifier (MSReqID). MSReqID is made up of 5 tokens (i.e., pieces of budget and workplan information) separated by underscores.

MSReqID = [*Requester*]_[*WorkplanType*]_wp*###_[MSDPID]*_[*VersionID*]

Tokens:

- **Requester**: Activity identifier for the TreeScan workgroup, as defined in the Task Order Matrix.
- **Workplan Type**: Workplan Type, as defined at SOC. For TreeScan workgroup, it will be “ahr” for ad hoc request.
- **wp###:** workplan identifier, composed of "wp" concatenated with a 3-digit number (starting at wp001 at the beginning of each sub-activity)
- **MSDPID:** unique Data Partner identifier. For requests that are customized by Data Partner, this is the DPID concatenated with the SiteID (e.g., AEOS, HMHPHC). For requests that are not customized/not specific by Data Partner, this is "nsdp”

Version ID: version identifier, composed of "v" for version or "b" for beta concatenated with a 2-digit number (e.g., v01, b02)

III. MASTER PROGRAM PARAMETERS

There are several master program parameters that must be specified for each request: task order, project, work plan, and Data Partner identifiers, common components file location, and patient exclusion list. Several parameters must be set by the SOC request programmer; two must be set by the Data Partner. Table 2 includes specifications for master program parameters.

Table 2. Master Program Parameters

Parameter	Field Name	Description
Project Identifier	MSPROJID	**Details**: project identifier for internal SOC identification and tracking. **Defined by:** SOC request programmer **Input type:** Required **Format:** Alphanumeric **Example**: *MSPROJID*= tshpv9
Work Plan Type	MSWPTYPE	**Details**: work plan type for internal SOC identification and tracking. **Defined by:** SOC request programmer **Input type:** Required **Format:** Alphanumeric **Example**: *MSWPTYPE*=ahr
Work Plan Identifier	MSWPID	**Details**: work plan identifier for internal SOC identification and tracking. **Note 1:** should follow the format [wp###]. **Note 2:** should be used to uniquely identify a modular program request. **Defined by:** SOC request programmer **Input type:** Required **Format:** Alphanumeric **Example**: *MSWPID*= wp001
Data Partner Identifier	MSDPID	**Details**: Data Partner identifier for internal SOC identification and tracking. **Note 1:** if a package is not Data Partner specific, MSDPID should equal “nsdp”. **Defined by:** SOC request programmer **Input type:** Required

Parameter	Field Name	Description
		Format: Alphanumeric **Example**: *MSDPID* =nsdp
Version Identifier	MSVERID	**Details**: version identifier for internal SOC identification and tracking. Should track each re-distribution of the package (if multiple distributions are required). **Note 1:** should follow the format [v##]. **Defined by:** SOC request programmer **Input type:** Required **Format:** Alphanumeric **Example**: *MSVERID* =v01
Common Components Directory	MSCDMPROG	**Details**: directory that contains ms_common_components.sas. **Defined by:** Data Partner **Input type:** Required **Format:** Alphanumeric **Example**: *MSCDMPROG* =C:\common_components\
Patient Exclusion File Location	ASO_EXCL_LIST	**Details**: name of the dataset containing a list of PatIDs that must be excluded from a particular request. The dataset must contain one variable, PatID, and be included (by the Data Partner) in the inputfiles folder of the request. **Defined by:** Data Partner **Input type:** Optional; leave blank if no PatID exclusion required **Format:** Alphanumeric **Example**: *ASO_EXCL_LIST* = infolder.ASO_EXCL_LIST
Diagnosis Tree Lookup Table File Name	dxtree_input	**Details**: name of the input file containing the diagnosis tree lookup table that will be used in the request. Details in the Lookup Tables section below. **Defined by:** SOC request programmer **Input type:** Required **Format:** Alphanumeric **Example**: *dxtree_input* = infolder.dxtree_input
ICD10-ICD9 Mapping Lookup Table File Name	id10_id9_map	**Details**: name of the input file containing the ICD10-ICD9 lookup table that will be used in the request. Details in the Lookup Tables section below. **Note 1**: This file is only required if *id10_flag* = 1 in the Data Partner Information File. **Defined by:** SOC request programmer **Input type:** Optional **Format:** Alphanumeric **Example**: *id10_id9_map* = infolder.id10_id9_map
Data Partner Information File Name	site_info	**Details**: name of the input file containing the Data Partner Information inputs that will be used in the request. Details in the Input Files section below.

Parameter	Field Name	Description
		Defined by: SOC request programmer **Input type:** Required **Format:** Alphanumeric **Example**: *site_info*= infolder.site_info
General Cohort Information File Name	general_input	**Details**: name of the input file containing the General Cohort Information inputs that will be used in the request. Details in the Input Files section below. **Defined by:** SOC request programmer **Input type:** Required **Format:** Alphanumeric **Example**: *general_input*= infolder.general_input
Concomitant Exposure Information File Name	concomitant_group	**Details**: name of the input file containing the Concomitant Exposure Information inputs that will be used in the request. Details in the Input Files section below. **Defined by:** SOC request programmer **Input type:** Required **Format:** Alphanumeric **Example**: *concomitant_group* = infolder.concomitant_group
Code Information File Name	subgroup	**Details**: name of the input file containing the Code Information inputs that will be used in the request. Details in the Input Files section below. **Defined by:** SOC request programmer **Input type:** Required **Format:** Alphanumeric **Example**: *subgroup*= infolder.subgroup
Adverse Event Encounter Setting Information File Name	AE_enc_setting	**Details**: name of the input file containing the Adverse Event Encounter Setting Information inputs that will be used in the request. Details in the Input Files section below. **Defined by:** SOC request programmer **Input type:** Required **Format:** Alphanumeric **Example**: *AE_enc_setting*= infolder.AE_enc_setting
Diagnosis Priority File Name	dxtree_priority	**Details**: name of the input file containing the Diagnosis Priority Information inputs that will be used in the request. Details in the Input Files section below. **Defined by:** SOC request programmer **Input type:** Required **Format:** Alphanumeric **Example**: *dxtree_priority*= infolder.dxtree_priority

IV. LOOKUP TABLES

There are two lookup tables in the TreeExtraction program – the Diagnosis Tree lookup table and the ICD10-ICD9 Mapping lookup table.

A. DIAGNOSIS TREE

The Diagnosis Tree lookup table (DXTREE_INPUT.sas7bdat) includes a hierarchical tree of codes that are eligible to be health outcomes of interest.

This lookup table should include a unique row for each code that makes up the tree. The requester will identify the diagnosis tree that will be used for a given program, and the SOC request programmer will ensure the diagnosis tree is in the correct format. The current program is limited to a 5-level tree, i.e., 4 aggregated levels plus the terminal leaf or ICD-9-CM level. The diagnosis code category is always DX, and the diagnosis code type is always 09 for ICD-9-CM codes. ICD-10-CM codes will be converted into their equivalent ICD-9-CM using the ICD10-ICD9 Mapping lookup file (ID10_ID9_MAP.sas7bdat).

Table 3 includes specifications for this file.

Table 3. DXTREE_INPUT Specification

Parameter	Field Name	Description
Diagnosis Code	dx	**Details:** Diagnosis codes of interest that are input as strings **Defined by:** SOC request programmer **Input type:** Required **Format:** Char (11) **Example:** *dx*=008.61
Diagnosis Code Type	dx_codetype	**Details:** The code type that is used to populate the Diagnosis Tree. Valid values include: • **09**: ICD-9-CM **Defined by:** SOC request programmer **Input type:** Required **Format:** Char (2) **Example:** *dx_codetype*=09
First level of the tree	mlccs1	**Details:** The code is a string that represents the first level of the tree. **Defined by:** SOC request programmer **Input type:** Required **Format:** Char (11) **Example:** *mlccs1*=09
Second level of the tree	mlccs2	**Details:** The code is a string that represents the second level of the tree. **Defined by:** SOC request programmer **Input type:** Required **Format:** Char (11)

Parameter	Field Name	Description
		Example: *mlcss2*=09.01
Third level of the tree	mlccs3	**Details:** The code is a string that represents the third level of the tree. **Defined by:** SOC request programmer **Input type:** Required **Format:** Char (11) **Example:** *mlccs3*=09.01.00
Fourth level of the tree	mlccs4	**Details:** The code is a string that represents the fourth level of the tree. **Defined by:** SOC request programmer **Input type:** Required **Format:** Char (11) **Example:** *mlccs4*=09.01.00.00
Fifth level of the tree	mlccs5	**Details:** The code is a string that represents the fifth level of the tree. The fifth level of the tree is also often called the leaf level or terminal level. The fifth level of the tree and the ICD-9-CM codes should be identical. **Defined by:** SOC request programmer **Input type:** Required **Format:** Char (11) **Example:** *mlccs5*=008.61
Diagnosis Code Description	ICD9_DES	**Details:** Description of the diagnosis code **Defined by:** SOC request programmer **Input type:** Required **Format:** Char (107) **Example:** *ICD9_DES* = DERMATITIS NOS

B. ICD10-ICD9 MAPPING

When ICD-10-CM codes are expected to be queried during a TreeExtraction request, then the *id10_flag* variable in the Data Partner Information File must be set to 1, and this lookup table is then required. If *id10_flag*=0, then this lookup table can be omitted. With the introduction of ICD-10-CM codes and the absence of an ICD-10-CM tree, ICD-10-CM codes will be converted into their equivalent ICD-9-CM using the ICD10-ICD9 Mapping lookup file (ID10_ID9_MAP.sas7bdat). The ICD-9-CM equivalents must be in the Diagnosis Tree lookup table (DXTREE_INPUT.sas7bdat).

Table 4 includes specifications for this file.

Table 4. ID10_ID9_MAP Specification

Parameter	Field Name	Description
ICD10 Code	orig_dx	**Details:** ICD10 health outcome of interest that has the potential to be included in the analytic dataset. It will be a string. Decimals will be compressed within the program. **Defined by:** SOC request programmer **Input type:** Required **Format:** Char (18) **Example:** *orig_dx* = R56.0
ICD10 Diagnosis Code Type	orig_dx_codetype	**Details:** Code type required by lookup file. Valid values include: • **10**: ICD-10-CM **Defined by:** SOC request programmer **Input type:** Required **Format:** Char (2) **Example:** *orig_dx_codetype*=10
ICD9 Code	dx	**Details:** ICD9 health outcome of interest that has the potential to be included in the analytic dataset. It will be a string. Decimals will be compressed within the program. **Defined by:** SOC request programmer **Input type:** Required **Format:** Char (18) **Example:** *dx* = 780.2
ICD9 Diagnosis Code Type	dx_codetype	**Details:** Code type required by lookup file. Valid values include: • **09**: ICD-9-CM **Defined by:** SOC request programmer **Input type:** Required **Format:** Char (2) **Example:** *dx_codetype*=09

V. INPUT FILES

There are 7 input files, each explained below.

A. DATA PARTNER INFORMATION FILE

The Data Partner Information input file (SITE_INFO.sas7bdat) includes Data Partner identifying information and cohort start and end dates for the request. These dates determine the boundaries for data extracted from the Sentinel Distributed Database (SDD). Dates must be set carefully based on the request. For example, if the exposure of interest becomes available on June 1, 2006, and the requester

requires a six-month period of enrollment prior to exposure, the study cohort date must begin on or before January 1, 2006 (to ensure that exposures of interest on June 1, 2006 can be eligible for inclusion by meeting enrollment requirements).

The dates specified here for cohort extraction apply globally to all the analyses being performed in a singular execution of this program. That is, if there is more than one cohort being extracted (i.e., there is more than one value for *exposure_group_name* or *exposure_group* as described in Table 6 below), then these dates apply globally to all the cohorts. Users must plan accordingly and set the cohort extraction dates based on the earliest and latest required data.

This input file must be created separately for each Data Partner. Table 5 includes specifications for this file.

Table 5. SITE_INFO Specification

Parameter	Field Name	Description
Data Partner ID	DPID	**Details:** standard Sentinel Data Partner identifier. **Defined by:** SOC request programmer **Input type:** Required **Format:** Char (2) **Example:** *DPID* = MS
Site ID	SITEID	**Details:** standard Sentinel site identifier. **Defined by:** SOC request programmer **Input type:** Required **Format:** Char (4) **Example:** *SITEID* = OC
Cohort Start Date	Cohort_start_dt	**Details:** defines the start date for a study period. **Note 1:** The requester has to carefully plan for the required enrollment/washout period requirements when setting these study period dates. For example, if the exposure of interest becomes available on June 1, 2006, and the requester requires a six-month period of enrollment prior to exposure, the study cohort date must begin on or before January 1, 2006 (to ensure that exposures of interest on June 1, 2006 can be eligible for inclusion by meeting enrollment requirements). **Defined by:** Requester / SOC request programmer **Input type:** Required **Format:** SAS Date (Date9.) **Example:** *Cohort_start_dt* = 01JAN2004
Cohort End Date	Cohort_end_dt	**Details:** defines the end date for a study period. **Note 1:** The requester has to carefully plan for the required minimum post-exposure follow-up periods when setting these study period dates. For example, if the minimum post-exposure

Parameter	Field Name	Description
		follow-up is 365 days and the Cohort End Date is Dec 31, 2012, the latest possible eligible exposure can be Dec 31, 2011. **Note 2:** The requester may enter the date in any logical date format. The SOC request programmer will convert that date into a SAS date. **Defined by:** Requester / SOC request programmer **Input type:** Required **Format:** SAS date (Date9.) **Example:** *Cohort_end_dt* =31DEC2009
Administrative Services Only Requirement	ASO_EXCL_FLAG	**Details:** indicator of whether the Administrative Services Only (ASO) population must be excluded from the request. Allowable values: • **1**: Exclude population • **0**: Include population **Note 1**: if the ASO population must be excluded from a request, a dataset including PatID values for exclusion must be included in the *inputfiles* folder and the dataset name should be included in the appropriate section of the master program. **Defined by:** Requester **Input type:** Required (cannot be left blank) **Format:** Num (8) **Example:** *ASO_EXCL_FLAG*=1
ICD10 Requirement	id10_flag	**Details:** indicator of whether ICD-10-CM codes will be included in this request. Allowable values: • **1**: ICD-10-CM included • **0**: ICD-10-CM not included **Note 1**: if ICD-10-CM codes must be included in a request, the lookup table (ID10_ID9_MAP.sas7bdat) is required. If not, then this file may be omitted. **Defined by:** Requester **Input type:** Required (cannot be left blank) **Format:** Num (8) **Example:** *id10_flag*=1

B. GENERAL COHORT INFORMATION FILE

The General Cohort Information input file (GENERAL_INPUT.sas7bdat) includes information on age and enrollment coverage requirements, incidence criteria, and required follow-up duration.

This input file should include a unique row for each cohort (given as either *exposure_group_name* or *exposure_group*). Table 6 includes specifications for this file.

Table 6. GENERAL_INPUT Specification

Parameter	Field Name	Description
Cohort Name	exposure_group_name	**Details:** name of the cohort of interest defined as either a singular exposure of interest or a grouping of same-day exposures. Defining unique *exposure_group_name* values allows for the specification of multiple cohorts in a single execution of the program package. **Note 1**: The ordering of exposures in *exposure_group_name* is purposeful and intended to aid the SOC request programmer and end-user. If the cohort of interest is a grouping of same day exposures, then the exposures are separated by underscores but concatenated in a single name. The first listed exposure of interest is the *primary* exposure of interest, which is explained below. For example, if the *exposure_group_name* = TDaP_HPV4, then the primary exposure of interest is TDaP, but there is also an interest in collecting data on same-day exposures to HPV4. **Defined by:** SOC request programmer **Input type:** Required (cannot be left blank) **Format:** Char (30) **Example:** *exposure_group_name* = TDaP_HPV4
Cohort Identification Number	exposure_group	**Details:** *exposure_group* is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the *exposure_group_name* and *exposure_group*. For example, if there are two values for *exposure_group_name*, then there will be two corresponding values for *exposure_group.* **Defined by:** SOC request programmer **Input type:** Required (cannot be left blank) **Format:** Num (8)

Parameter	Field Name	Description
		Example: *exposure_group*=1
Enrollment Gap	Enr_gap	**Details:** sets the number of days that will be bridged between two consecutive enrollment periods to create a "continuously enrolled" period. For example, if *Enr_gap*=30 and a member has the required insurance 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. **Note 1**: A gap of 45-days is recommended for most uses. **Defined by:** Requester **Input type:** Required (default=45) **Format**: Num (8) **Example**: *Enr_gap*=45 (gaps less than or equal to 45 days will be "bridged" to form one "continuously enrolled" sequence)
Health Outcome of Interest Washout Period	AE_wash_up	**Details:** length of the washout period to determine HOI incidence, given in days. **Note 1:** this period is indexed to the identification date of the health outcome of interest (not the index date/exposure date). Therefore, the length of the period will be the same but the timeline will be different for each identified HOI. **Note 2:** As a general rule, it is undesirable to have multiple instances of the same HOI of interest appear in the same observation period. Therefore, *AE_wash_up* is required to be >= (*F_up_window_end + 1)* (explained below). **Named by:** Requester **Input type:** Required (default=183) **Format:** Num (8) **Example:** *AE_wash_up* = 183
Start of Follow-up Window	F_up_window_st	**Details:** sets the start of the follow-up or observation window during which an incident health outcome of interest could occur in

Parameter	Field Name	Description
		days. Day 0 is always assumed to be the day of exposure and the index date. If the follow-up window was 7-28 days after exposure, then *F_up_window_st*=7. **Note 1:** Valid entries must be greater than or equal to 0. **Defined by:** Requester **Input type:** Required (default=1) **Format**: Num (8) **Example** *F_up_window_st*=7
End of Follow-up Window	F_up_window_end	**Details:** sets the end of the follow-up or observation window during which an incident health outcome of interest could occur. Day 0 is always assumed to be the day of exposure and the index date. If the follow-up window was 7-28 days after exposure, then *F_up_window_end*=28. **Note1**: In this program, a user is required to have continuous enrollment throughout the follow-up period for inclusion in the cohort. **Note2:** In this release of the program, the control window cannot be before the exposure index date. Therefore, valid entries must be greater than or equal to *F_up_window_st*. **Defined by:** Requester **Input type:** Required (default=56) **Format**: Num (8) **Example**: *F_up_window_end*=28
Medical Coverage	Medcov	**Details:** indicates whether continuous enrollment in medical coverage is required. Allowable values: • Y: Yes **Defined by:** Requester **Input type:** Required (default=Y) **Format:** Char (1) **Example:** *Medcov*=Y
Drug Coverage	Drugcov	**Details:** indicates whether continuous enrollment in drug coverage is required. Allowable values: • Y: Yes • <blank>: any coverage is permissible.

Parameter	Field Name	Description
		Defined by: Requester **Input type:** Required **Format:** Char (1) **Example:** *Drugcov*=Y
Pre-Birth Enrollment Allowance	birth_enr_dist_allowed	**Details:** This variable is provides a grace period of enrollment prior to patient birth date. This allowance adjusts patient enrollment. It is intended for queries that look at infant populations. **Defined by:** Requester **Input type:** Required (default=30) **Format:** Num (8) **Example:** *birth_enr_dist_allowed*=30
Post-Birth Enrollment Allowance	enr_birth_grace_period	**Details:** grace period applied to patients who appear "enrolled" after their birth date but who have likely been enrolled since birth. It is given in days. It is intended for queries that look at infant populations. **Defined by:** Requester **Input type:** Required (default=45) **Format:** Num (8) **Example:** *enr_birth_grace_period*=45
Incidence Level for the tree	Incident_level	**Details:** the level of the tree that defines an incident HOI. The tree as defined in Diagnosis Tree lookup file (DXTREE_INPUT.sas7bdat) has 5 levels. This parameter designates that a diagnosis may be incident if there are no diagnoses that share the same node at the *Incident_level* of the tree in the HOI washout period (*AE_wash_up).* Allowable values: • 1:MLCCS Level 1 • 2: MLCCS Level 2 • 3: MLCCS Level 3 • 4: MLCCS Level 4 • 5: MLCCS Level 5 **Defined by:** Requester **Input type:** Required **Format:** Num (8) **Example:** *Incident_level*=3
Start of the Age Group of Interest	Age_begin	**Details:** the earliest age of eligibility to be included in the cohort, or the earliest potential index date/date of exposure. For example, if a

Parameter	Field Name	Description
		cohort was created among 9-26 year olds, *Age_begin* would be 9. Age for the cohort is determined at the time of exposure (which is the index date). **Defined by:** Requester **Input type:** Required (default=0) **Format:** Num (8) **Example:** *Age_begin*=9
End of the Age Group of Interest	Age_finish	**Details:** The latest age of eligibility to be included in the cohort, or the latest potential index date/date of exposure. For example, if a cohort were created among 9-26 year olds, then *Age_finish* would be 26. Age for the cohort is determined at the time of exposure (which is the index date). **Defined by:** Requester **Input type:** Required (default=160) **Format:** Num (8) **Example:** *Age_finish*=26
Time Increment for use with the Age Variables	Age_period	**Details:** This variable indicates the time increment to be associated with the *Age_begin* and *Age_finish*. It can be entered in any recognized standard SAS unit of date and time (e.g., DAY, YEAR, WEEK, HOUR, etc.) **Note 1:** time intervals are not case-sensitive. **Defined by:** Requester **Input type:** Required (default=YEAR) **Format:** Char (12) **Example:** *Age_period*=YEAR
Exposure Washout Period	exp_wash_up	**Details:** length in days of the washout period to determine exposure incidence. Applies only to the exposure of interest that is designated as *primary.* **Note 1:** This period is indexed to the identification of the exposure. The exposure that must be incident is given by the *primary* value in the CONCOMITANT_GROUP.sas7bdat input file for the *exposure_group* (i.e., cohort identification number of interest). The *primary* value corresponds to a *subgroup* index number. The SUBGROUP.sas7bdat file contains the code list for the exposure of interest. Incidence is assessed with respect to the

Parameter	Field Name	Description
		primary exposure only, not the combination of exposures. **Note 2:** If a request wants to ensure that there are not multiple exposures in the same observation window, it is necessary to set *exp_wash_up>=(F_up_window_end + 1).* **Named by:** Requester **Input type:** Required **Format:** Num (8) **Example:** *exp_wash_up* = 42

C. CONCOMITANT EXPOSURE INFORMATION FILE

The Concomitant Exposure Information input file (CONCOMITANT_GROUP.sas7bdat) includes the grouping of exposures that will be considered to be part of a same-day concomitant exposure grouping. The Concomitant Exposure Information input file is linked to the General Cohort Information input file (GENERAL_INPUT.sas7bdat) by *exposure_group* value. It is also linked to the Code Information input file (SUBGROUP.sas7bdat) by the *subgroup* value.

This input file should include a unique row for each *exposure_group and exposure_order*. Table 7 includes specifications for this file.

Table 7. CONCOMITANT_GROUP Specification

Parameter	Field Name	Description
Cohort Identification Number	exposure_group	**Details:** *exposure_group* is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the *exposure_group_name* and *exposure_group*. For example, if there are two values for *exposure_group_name*, then there will be two corresponding values for *exposure_group.* **Defined by:** SOC request programmer **Input type:** Required (cannot be left blank) **Format:** Num (8) **Example:** *exposure_group*=1
Exposure Name	group	**Details:** name of one of the exposures of interest within a cohort as defined by *exposure_group*. For each *exposure_group_name,* there may be a grouping of same-day concomitant exposures of interest connected by underscores. The *group* should correspond to a single exposure defined using codes in the Code Information

		input file (SUBGROUP.sas7bdat) given by the *subgroup* value. **Defined by:** SOC request programmer **Input type:** Required (cannot be left blank) **Format:** Char (30) **Example:** *group* = HPV
Primary Exposure of Interest Indicator	primary	**Details:** binary variable (1/0) indicating whether the *group* exposure is the primary exposure of interest. The exposure of interest designated as primary is the only exposure to which *exp_wash_up* parameters apply. **Defined by:** SOC request programmer **Input type:** Required (cannot be left blank) **Format:** Num (8) **Example:** *primary* = 1
Exposure Order Indicator	exposure_order	**Details:** ordinal variable that identifies the ordering of the exposures in a *exposure_group.* Exposures designated as *primary* will always be assigned *exposure_order*=1. **Defined by:** SOC request programmer **Input type:** Required (cannot be left blank) **Format:** Num (8) **Example:** *exposure_order* = 1
Code List Indicator	subgroup	**Details:** *subgroup* links to the SUBGROUP.sas7bdat input files. The *subgroup* value indicates the appropriate code list to define a particular exposure (i.e., *group*) within a cohort (i.e., *exposure_group).* For example, if *exposure_group*=1, *group*=HPV and *subgroup*=1, the code list to define HPV will be in the SUBGROUP.sas7bdat input file with *subgroup*=1. **Note 1**: The number of subgroup values will depend on the number of independent exposures of interest after all the *exposure_group*(s) are considered along with any exclusion exposures of interest. **Defined by:** SOC request programmer **Input type:** Required (cannot be left blank) **Format:** Num (8) **Example:** *subgroup*=1

D. EXCLUSION EXPOSURE INFORMATION FILE

The Exclusion Exposure Information input file (GROUP_EXCLUSION.sas7bdat) includes the grouping of exposures that will be considered exclusion criteria for a particular cohort, along with the number of days defining the exclusion period prior to the primary exposure of interest. The Exclusion Exposure Information input file is linked to the General Cohort Information input file (GENERAL_INPUT.sas7bdat) by *exposure_group* value. It is also linked to the Code Information input file (SUBGROUP.sas7bdat) by the *subgroup* value.

This input file should include a unique row for each *exposure_group* and *subgroup*. Table 8 includes specifications for this file.

Table 8. GROUP_EXCLUSION Specification

Parameter	Field Name	Description
Cohort Identification Number	exposure_group	**Details:** *exposure_group* is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the *exposure_group_name* and *exposure_group*. For example, if there are two values for *exposure_group_name*, then there will be two corresponding values for *exposure_group.* **Defined by:** SOC request programmer **Input type:** Required (cannot be left blank) **Format:** Num (8) **Example:** *exposure_group*=1
Code List Indicator	subgroup	**Details:** *subgroup* links to the SUBGROUP.sas7bdat input files. The *subgroup* value indicates the appropriate code list to define a particular exposure (i.e., *group*) within a cohort (i.e., *exposure_group)*. For example, if *exposure_group*=1, *group*=HPV and *subgroup*=1, the code list to define HPV will be in the SUBGROUP.sas7bdat input file with *subgroup*=1. **Note 1**: The number of subgroup values will depend on the number of independent exposures of interest after all the *exposure_group*(s) are considered along with any exclusion exposures of interest. **Note 2**: Exclusions will be applied to the number of calendar days (*exclusion_distance*) prior to the *primary* exposure of interest. **Defined by:** SOC request programmer

Parameter	Field Name	Description
		Input type: Required (cannot be left blank) **Format:** Num (8) **Example:** *subgroup*=1
Exclusion Days Prior to Index Date	exclusion_distance	**Details:** number of days that defines the exclusion period prior to the primary exposure of interest **Note 1**: The exclusion period excludes the index date of the *primary* exposure of interest. **Defined by:** SOC request programmer **Input type:** Required (cannot be left blank) **Format:** Num (8) **Example:** *exclusion_distance* = 183

E. CODE INFORMATION FILE

The Code Information input file (SUBGROUP.sas7bdat) includes the comprehensive list of codes used to define cohorts of interest. The Code Information input file is linked to the Concomitant Exposure Information input file (CONCOMITANT_GROUP.sas7bdat) and the Exclusion Exposure Information input file (GROUP_EXCLUSION.sas7bdat) by *subgroup* value.

This input file should include a unique row for each *subgroup, code, code_category*, and *code_type*. Table 9 includes specifications for this file.

Table 9. SUBGROUP Specification

Parameter	Field Name	Description
Code List Indicator	subgroup	**Details:** The *subgroup* value indicates the appropriate code list to define a particular exposure (i.e., *group*) within a cohort (i.e., *exposure_group)*. Other files are linked by this value to obtain the list of exposure codes. **Note 1**: The number of subgroup values will depend on the number of independent exposures of interest after all the *exposure_group*(s) are considered along with any exclusion exposures of interest. **Defined by:** SOC request programmer **Input type:** Required (cannot be left blank) **Format:** Num (8) **Example:** *subgroup*=1
Code	code	**Details:** NDC, procedure and/or diagnosis codes of interest. **Note 1:** There is no ability to use a wildcard in the program. If an end-user submits a request

Parameter	Field Name	Description
		with a wildcard, then an SOC request programmer must translate that request into codes that can be read by the program. **Note 2:** The SOC request programmer may compress the decimal points when creating this input file but there is also coding within the program to compress decimal points if the SOC request programmer does not do so. **Defined by:** Requester **Input type:** Required **Format:** Char (11) **Example:** (*code_category*=RX; *code_type*=11), *code*=12345678911
Code Category	code_category	**Details:** type of each code category value included in the *code_type* field (below) of this file. Valid values include: • **RX:** NDC • **DX:** Diagnosis code • **PX:** Procedure code **Defined by:** Requester **Input type:** Required **Format:** Char (2) **Example:** *code_category*=PX
Code Type	code_type	**Details:** type of each code value included in the *code_category* field (above) of this file. **Note 1:** This variable is optional if all codes in the table have *code_category* = RX. Valid values include: If *code_category*= RX: • **11**: 11-digit NDC If *code_category* = DX: • **09**: ICD-9-CM • **10**: ICD-10-CM • **11:** ICD-11-CM • **OT:** Other If *code_category* = PX: • **09**: ICD-9-CM • **10**: ICD-10-CM • **11**: ICD-11-CM • **ND:** 11-digit NDC Code

Parameter	Field Name	Description
		• **C4**: CPT-4 (*i.e.,* HCPCS Level I) • **HC**: HCPCS (*i.e.,* HCPCS Level II) • **H3**: HCPCS Level III • **C2**: CPT Category II • **C3**: CPT Category III • **RE**: Revenue • **LO**: Local homegrown • **OT**: Other **Defined by:** Requester **Input type:** Optional **Format:** Char (2) **Example:** (*code_category*=DX), *code_type*=09

F. ADVERSE EVENT ENCOUNTER SETTING INFORMATION FILE

The Adverse Event Encounter Setting Information input file (AE_ENC_SETTING.sas7bdat) includes information on the encounter settings that a HOI must occur in to be considered incident for the purposes of generation of the analytic dataset. The Adverse Event Encounter Setting Information input file is linked to the General Cohort Information input file (GENERAL_INPUT.sas7bdat) by *exposure_group* value.

This input file should include a unique row for each *exposure_group* and *incident_enc_setting*. Table 10 includes specifications for this file.

Table 10. AE_ENC_SETTING Specification

Parameter	Field Name	Description
Cohort Identification Number	exposure_group	**Details:** *exposure_group* is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the *exposure_group_name* and *exposure_group*. For example, if there are two values for *exposure_group_name*, then there will be two corresponding values for *exposure_group.* **Defined by:** SOC request programmer **Input type:** Required (cannot be left blank) **Format:** Num (8) **Example:** *exposure_group*=1
Setting to Define Incidence	Incident_enc_setting	**Details:** eligible settings in which an incident health outcome of interest can occur. Allowable values: • **AV:** outpatient • **ED:** emergency department • **IP**: inpatient • **IS:** institutional stay

		• **OA**: other ambulatory encounter **Note 1:** if multiple care settings must be specified, one row per setting per *exposure_group* must be listed. **Named by:** Requester **Input type:** Required **Format:** Char (2) **Example:** *Incident_enc_setting*=IP

G. DIAGNOSIS PRIORITY FILE

The Diagnosis Priority input file (DXTREE_PRIORITY.sas7bdat) provides the priority of the diagnosis to be ascertained as a health outcome of interest if two qualifying incident health outcomes of interest happen to occur on the same day at the same *n*th level of the diagnosis tree, set by the requester with the variable *Incident_level*. The *dx* and *dx_codetype* fields should be identical in the DXTREE_PRIORITY.sas7bdat and DXTREE_INPUT.sas7bdat files; however, these fields are repeated for each *exposure_group* in the DXTREE_PRIORITY.sas7bdat file.

The Diagnosis Priority file is created by the SOC request programmer according to the specifications of the requester. The simplest type of priority file is based on frequency counts in the background population of interest.

In the current version of this program, the diagnosis code category is always DX. ICD-10-CM codes will be converted to their ICD-9-CM equivalent for use with DXTREE_INPUT.sas7bdat and the conversion will be done with the ID10_ID9_MAP.sas7bdat file.

Table 11. DXTREE_PRIORITY Specifications

Parameter	Field Name	Description
Cohort Identification Number	exposure_group	**Details:** *exposure_group* is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the *exposure_group_name* and *exposure_group*. For example, if there are two values for *exposure_group_name*, then there will be two corresponding values for *exposure_group.* **Defined by:** SOC request programmer **Input type:** Required (cannot be left blank) **Format:** Num (8) **Example:** *exposure_group*=1
Diagnosis Code	dx	**Details:** Diagnosis codes of interest that are input as strings. **Defined by:** Requester **Input type:** Required **Format:** Char (11) **Example:** *dx*=242.01

Parameter	Field Name	Description
Diagnosis Code Type	dx_codetype	**Details:** The code type that is used to populate the Diagnosis Tree. Valid values include: • **09**: ICD-9-CM **Defined by:** Requester **Input type:** Required **Format:** Char (2) **Example:** *dx_codetype*=09
Incidence Level for the tree	Incident_level	**Details:** the level of the tree that defines an incident HOI. The tree as defined in Diagnosis Tree lookup file (DXTREE_INPUT.sas7bdat) has 5 levels. This parameter designates that a diagnosis may be incident if there are no diagnoses that share the same node at the *Incident_level* of the tree in the HOI washout period (*AE_wash_up).* Allowable values: • 1:MLCCS Level 1 • 2: MLCCS Level 2 • 3: MLCCS Level 3 • 4: MLCCS Level 4 • 5: MLCCS Level 5 **Defined by:** Requester **Input type:** Required **Format:** Num (8) **Example:** *Incident_level*=3
Priority Level of Diagnosis	priority	**Details:** the priority of the diagnosis to be ascertained as a health outcome of interest if two qualifying incident health outcomes of interest happen to occur on the same day at the same *incident_level* of the diagnosis tree. **Note 1:** Within each incident level node, there is a numeric priority list of all diagnoses from 1 to the maximum number of diagnoses in that node. **Defined by:** Requester **Input type:** Required **Format:** Num (8) **Example:** *priority*=33

VI. OUTPUT FILES

There are seven output datasets, one log file, and one signature file output to the msoc folder and returned to SOC. There is one output dataset and one .LST file output to the dplocal folder and retained at the Data Partner site.

All output files occur on the unit of the valid exposure. That is, none of the information should be interpreted on the unit of the patient since patients are allowed to contribute multiple valid exposures to the same analysis.

A. OUTPUT TO THE SENTINEL OPERATIONS SENTINEL (MSOC FOLDER FILE)

1. Analytic Dataset

The Analytic Dataset output file (SELF_CONTROL_ORIG_DX.sas7bdat) provides information on the site-specific health outcomes of interest and their accompanying time-to-event, which is required for analysis with the TreeScan software.

Table 12. SELF_CONTROL_ORIG_DX Specifications

Parameter	Field Name	Description
Data Partner ID	DPID	**Details:** standard Sentinel Data Partner identifier. **Format:** Char (2) **Example:** *DPID* = MS
Site ID	SITEID	**Details:** standard Sentinel site identifier. **Format:** Char (4) **Example:** *SITEID* = OC
Cohort Identification Number	exposure_group	**Details:** *exposure_group* is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the *exposure_group_name* and *exposure_group*. For example, if there are two values for *exposure_group_name*, then there will be two corresponding values for *exposure_group.* **Format:** Num (8) **Example:** *exposure_group*=1
Cohort Name	exposure_group_name	**Details:** name of the cohort of interest defined as either a singular exposure of interest or a grouping of same-day exposures. Defining unique *exposure_group_name* values allows for the specification of multiple cohorts in a single execution of the program package. **Note 1**: The ordering of exposures in *exposure_group_name* is purposeful and intended to aid the SOC request programmer and end-user. If the cohort of interest is a grouping of same day

Parameter	Field Name	Description
		exposures, then the exposures are separated by underscores but concatenated in a single name. The first listed exposure of interest is the *primary* exposure of interest, which is explained below. For example, if the *exposure_group_name* = TDaP_HPV4, then the primary exposure of interest is TDaP, but there is also an interest in collecting data on same-day exposures to HPV4. **Format:** Char (30) **Example:** *exposure_group_name* = TDaP_HPV4
Incidence Level for the tree	incident_level	**Details:** the level of the tree that defines an incident HOI. The tree as defined in Diagnosis Tree lookup file (DXTREE_INPUT.sas7bdat) has 5 levels. This variable designates that a diagnosis may be incident if there are no diagnoses that share the same node at the *Incident_level* of the tree in the HOI washout period (*AE_wash_up).* Allowable values: • 1: Level 1 • 2: Level 2 • 3: Level 3 • 4: Level 4 • 5: Level 5 **Format:** Num (8) **Example:** *Incident_level*=3
Exposure Combination Identification Number	exp_comb	**Details:** concatenation of binary (0/1) indicators for presence or absence of exposure of interest that corresponds to the *exposure_group_name*. For example, if the *exposure_group_name* is TDaP_HPV4 and the result pertains to the ascertainment of a TDaP vaccination without a same-day concomitant HPV4 vaccination, then the *exp_comb*=10. The number of characters is dictated by the *exposure_group_name* that has the most same-day exposure groupings included. For example, if there is an *exposure_group_name* with 3 exposures, then the *exp_comb* file will have 3 characters. **Note 1:** '.' characters indicate that the program is not recording any information on other exposures. For example, if the *exposure_group_name* is TDaP_HPV4 but another *exposure_group_name* in the execution of the program has 3 exposures, then the potential *exp_comb values* for TDaP_HPV4 will be "10." or "11."

Parameter	Field Name	Description
		Format: Char (30) **Example:** *exp_comb* = 101
Age Grouping at Index Date	age_group	**Details:** stratification of age in years at index date, i.e. exposure. **Format:** Char (5) **Example:** *age_group* = 20-24
Original Incident Health Outcome of Interest	orig_dx	**Details:** original health outcome of interest that must be converted into its ICD-9-CM equivalent using the ICD10-ICD9 Mapping lookup table (ID10_ID9_MAP.sas7bdat) to be included in the analytic dataset. It will be a string. **Note 1**: If the *orig_dx* happens to be an ICD-9-CM code, then no conversion is necessary, and ICD-9-CM code is used. **Format:** Char (18) **Example:** *orig_dx* = 780.2 (if ICD-9-CM) *orig_dx*=R56.0 (if ICD-10-CM)
Original Diagnosis Code Type	orig_dx_codetype	**Details:** Code type required by lookup file. Valid values include: • **09**: ICD-9-CM • **10**: ICD-10-CM **Format:** Char (2) **Example:** *orig_dx_codetype*=09
Ascertained Incident Health Outcome of Interest	dx	**Details:** incident health outcome of interest that is included in the analytic dataset. It will be a string. **Note 1**: If the *orig_dx* happens to be an ICD-10-CM code, then it will be converted into its ICD-9-CM equivalent code using the ICD10-ICD9 Mapping lookup table (ID10_ID9_MAP.sas7bdat). **Format:** Char (18) **Example:** *dx* = 780.2
Ascertained Diagnosis Code Type	dx_codetype	**Details:** code type that is used to populate the Diagnosis Tree (DXTREE_INPUT.sas7bdat). Valid values include: • **09**: ICD-9-CM **Format:** Char (2) **Example:** *dx_codetype*=09
Time to Event for the Ascertained Health	days_from_exp_to_outcome	**Details:** time to event of the incident health outcome of interest that is indexed on the day of exposure where day of exposure=0. *days_from_exp_to_outcome* must fall within the

Parameter	Field Name	Description
Outcome of Interest		allowable follow-up window and is given in days. For example, it is 12 if the diagnosis occurred 12 days after vaccination. **Format:** Num (8) **Example:** *days_from_exp_to_outcome* = 23
Eligible Outcome Counts	exp_dx_dist_cnt	**Details:** Number of distinct patients having the particular outcome of interest following an eligible exposure with a particular time-to-event. That is, if two patients at the Data Partner both have incident codes of 780.2 two days after exposure, then they are aggregated into a singular line that is displayed here. **Format:** Num (8) **Example:** *exp_dx_dist_cnt* = 2

2. Eligible Exposures Descriptive Statistics

The Eligible Exposures output file (ELIG_EXP_DOSES.sas7bdat) is file that provides a count of all valid exposures that meet enrollment and inclusion/exclusion criteria. These data are used to calculate attributable risk.

This output file should include a row for each unique combination of cohort (exposure_group), exposure combination identification code (exp_comb), and age grouping at index date (age_group).

Table 13. ELIG_EXP_DOSES Specifications

Parameter	Field Name	Description
Cohort Identification Number	exposure_group	**Details:** *exposure_group* is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the *exposure_group_name* and *exposure_group*. For example, if there are two values for *exposure_group_name*, then there will be two corresponding values for *exposure_group.* **Format:** Num (8) **Example:** *exposure_group*=1
Cohort Name	exposure_group_name	**Details:** name of the cohort of interest defined as either a singular exposure of interest or a grouping of same-day exposures. Defining unique *exposure_group_name* values allows for the specification of multiple cohorts in a single execution of the program package.

Parameter	Field Name	Description
		Note 1: The ordering of exposures in *exposure_group_name* is purposeful and intended to aid the SOC request programmer and end-user. If the cohort of interest is a grouping of same day exposures, then the exposures are separated by underscores but concatenated in a single name. The first listed exposure of interest is the *primary* exposure of interest, which is explained below. For example, if the *exposure_group_name* = TDaP_HPV4, then the primary exposure of interest is TDaP, but there is also an interest in collecting data on same-day exposures to HPV4. **Format:** Char (30) **Example:** *exposure_group_name* = TDaP_HPV4
Exposure Combination Identification Number	exp_comb	**Details:** concatenation of binary (0/1) indicators for presence or absence of exposure of interest that corresponds to the *exposure_group_name*. For example, if the *exposure_group_name* is TDaP_HPV4 and the result pertains to the ascertainment of a TDaP vaccination without a same-day concomitant HPV4 vaccination, then the *exp_comb*=10. The number of characters is dictated by the *exposure_group_name* that has the most same-day exposure groupings included. For example, if there is a *exposure_group_name* with 3 exposures, then the *exp_comb* file will have 3 characters. **Note 1:** '.' characters indicate that the program is not recording any information on other exposures. For example, if the *exposure_group_name* is TDaP_HPV4 but another *exposure_group_name* in the execution of the program has 3 exposures, then the potential *exp_comb values* for TDaP_HPV4 will be "10." or "11."

Parameter	Field Name	Description
		Format: Char (30) **Example:** *exp_comb* = 101
Age Grouping at Index Date	age_group	**Details:** stratification of age in years at index date, i.e. exposure. Valid values: • 15-19 • 20-24 • 25-29 • 30-34 • 35-39 • 40-44 • 45-49 **Format:** Char (5) **Example:** *age_group* = 20-24
Eligible Exposure Dose Count	elig_exp_dose_cnt	**Details:** number of unique episodes of exposure that have met all enrollment and inclusion/exclusion criteria. **Format:** Num (8) **Example:** *elig_exp_dose_cnt* = 45000

3. Age at Exposure Distribution

The Age at Exposure Distribution output file (EXP_AGE.sas7bdat) includes the distribution of age at index date (i.e., date of exposure).

This output file should include a row for each unique combination of cohort (*exposure_group*), exposure combination identification code (*exp_comb*), and integer value between *Age_begin* and *Age_finish* with the unit of time defined by *Age_period*.

Table 14. EXP_AGE Specification

Parameter	Field Name	Description
Cohort Identification Number	exposure_group	**Details:** *exposure_group* is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the *exposure_group_name* and *exposure_group*. For example, if there are two values for *exposure_group_name*, then there will be two corresponding values for *exposure_group.* **Format:** Num (8) **Example:** *exposure_group*=1
Exposure Combination	exp_comb	**Details:** concatenation of binary (0/1) indicators for presence or absence of exposure of interest that corresponds to the *exposure_group_name*.

Parameter	Field Name	Description
Identification Number		For example, if the *exposure_group_name* is TDaP_HPV4 and the result pertains to the ascertainment of a TDaP vaccination without a same-day concomitant HPV4 vaccination, then the *exp_comb*=10. The number of characters is dictated by the *exposure_group_name* that has the most same-day exposure groupings included. For example, if there is a *exposure_group_name* with 3 exposures, then the *exp_comb* file will have 3 characters. **Note 1:** '.' characters indicate that the program is not recording any information on other exposures. For example, if the *exposure_group_name* is TDaP_HPV4 but another *exposure_group_name* in the execution of the program has 3 exposures, then the potential *exp_comb values* for TDaP_HPV4 will be "10." or "11." **Format:** Char (30) **Example:** *exp_comb* = 101
Time Increment for use with the Age Variables	Age_period	**Details:** This variable indicates the time increment to be associated with the *Age_begin* and *Age_finish*. **Format:** Char (12) **Example:** *Age_period*=YEAR
Age at Date of Exposure	age_at_exp_dt	**Details:** This variable indicates age at exposure in the units specified by *Age_period*. **Format:** Num (8) **Example:** *age_at_exp_dt* =330
Eligible Exposures Count	COUNT	**Details:** Counts of eligible exposures that occur at each *age_at_exp_dt*. **Format:** Num (8) **Example:** *COUNT* =8
Eligible Exposures Percentage	PERCENT	**Details:** Percentage of eligible exposures that occur at each *age_at_exp_dt*. **Note 1:** Percentages are calculated within *exposure_group* such that all percentages associated with a particular *exposure_group* add up to 100. **Note 2:** Valid values will be between 0 and 100. **Format:** Num (8) **Example:** *PERCENT* =2.5

4. Exposure Group Incident Level Outcome Summary

The Exposure Group Incident Level Outcome Summary output file (EXPOSURE_GROUP_LEVEL_OUTC.sas7bdat) includes counts of unique patients and incident adverse events by cohort and incidence level.

This output file should include a row for each unique combination of cohort (*exposure_group*) and incidence level for the tree (*incident_level*).

Table 15. ORIG_DX_WKDAY Specification

Parameter	Field Name	Description
Cohort Name	exposure_group_name	**Details:** name of the cohort of interest defined as either a singular exposure of interest or a grouping of same-day exposures. **Format:** Char (30) **Example:** *exposure_group_name* = TDaP_HPV4
Cohort Identification Number	exposure_group	**Details:** *exposure_group* is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the *exposure_group_name* and *exposure_group*. For example, if there are two values for *exposure_group_name*, then there will be two corresponding values for *exposure_group.* **Format:** Num (8) **Example:** *exposure_group*=1
Incidence Level for the tree	incident_level	**Details:** the level of the tree that defines an incident HOI. The tree as defined in Diagnosis Tree lookup file (DXTREE_INPUT.sas7bdat) has 5 levels. This variable designates that a diagnosis may be incident if there are no diagnoses that share the same node at the *Incident_level* of the tree in the HOI washout period (*AE_wash_up).* Allowable values: • 1:MLCCS Level 1 • 2: MLCCS Level 2 • 3: MLCCS Level 3 • 4: MLCCS Level 4 • 5: MLCCS Level 5 **Format:** Num (8) **Example:** *Incident_level*=3
Patients with Exposed Adverse Events Count	exp_diag_pnt_cnt	**Details:** Counts of unique patients having an incident adverse event in the observation window following exposure.

Parameter	Field Name	Description
		Format: Num (8) **Example:** *exp_diag_pnt_cnt* =8
Exposed Adverse Events Count	exp_diag_event_cnt	**Details:** Counts of unique incident adverse events in the observation window following exposure. **Format:** Num (8) **Example:** *exp_diag_event_cnt* =2.5

5. Original Health Outcome of Interest Day of the Week Distribution

The Original Health Outcomes of Interest Day of the Week Distribution output file (ORIG_DX_WKDAY.sas7bdat) includes the distribution of the original health outcome of interest by the day of the week on which it occurs.

This output file should include a row for each unique combination of cohort (*exposure_group*), incidence level for the tree (*incident_level*), exposure combination identification code (*exp_comb*), original diagnosis code (*orig_dx*), and HOI day of the week (*dx_wkday*).

Table 16. ORIG_DX_WKDAY Specification

Parameter	Field Name	Description
Cohort Identification Number	exposure_group	**Details:** *exposure_group* is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the *exposure_group_name* and *exposure_group*. For example, if there are two values for *exposure_group_name*, then there will be two corresponding values for *exposure_group.* **Format:** Num (8) **Example:** *exposure_group*=1
Incidence Level for the tree	incident_level	**Details:** the level of the tree that defines an incident HOI. The tree as defined in Diagnosis Tree lookup file (DXTREE_INPUT.sas7bdat) has 5 levels. This variable designates that a diagnosis may be incident if there are no diagnoses that share the same node at the *Incident_level* of the tree in the HOI washout period (*AE_wash_up).* Allowable values: • 1:MLCCS Level 1 • 2: MLCCS Level 2 • 3: MLCCS Level 3 • 4: MLCCS Level 4 • 5: MLCCS Level 5

Parameter	Field Name	Description
		Format: Num (8) **Example:** *Incident_level*=3
Exposure Combination Identification Number	exp_comb	**Details:** concatenation of binary (0/1) indicators for presence or absence of exposure of interest that corresponds to the *exposure_group_name*. For example, if the *exposure_group_name* is TDaP_HPV4 and the result pertains to the ascertainment of a TDaP vaccination without a same-day concomitant HPV4 vaccination, then the *exp_comb*=10. The number of characters is dictated by the *exposure_group_name* that has the most same-day exposure groupings included. For example, if there is a *exposure_group_name* with 3 exposures, then the *exp_comb* file will have 3 characters. **Note 1:** '.' characters indicate that the program is not recording any information on other exposures. For example, if the *exposure_group_name* is TDaP_HPV4 but another *exposure_group_name* in the execution of the program has 3 exposures, then the potential *exp_comb values* for TDaP_HPV4 will be "10." or "11." **Format:** Char (30) **Example:** *exp_comb* = 101
Original Incident Health Outcome of Interest	orig_dx	**Details:** original health outcome of interest. It will be a string. **Format:** Char (18) **Example:** *orig_dx* = 780.2 (if ICD-9-CM) *orig_dx*=R56.0 (if ICD-10-CM)
Day of the Week of the Original Health Outcome of Interest	dx_wkday	**Details:** stratification of all eligible health outcomes of interest by the day of the week of their occurrence. **Note 1:** 1=Sunday **Format:** Num (8) **Example:** *dx_wkday* =3
Eligible Outcomes Count	COUNT	**Details:** Counts of original health outcomes of interest that fall on a particular weekday. **Format:** Num (8) **Example:** *COUNT* =8

Parameter	Field Name	Description
Eligible Outcomes Percentage	PERCENT	**Details:** Percentage of original health outcomes of interest that fall on a particular weekday. **Note 1:** Percentages are calculated within *exposure_group* such that all percentages associated with a particular *exposure_group* add up to 100. **Note 2:** Valid values will be between 0 and 100. **Format:** Num (8) **Example:** *PERCENT*=2.5

6. Ascertained Health Outcome of Interst Day of the Distribution

The Ascertained Health Outcome of Interest Day of the Week Distribution output file (DX_WKDAY.sas7bdat) includes the distribution of the ascertained health outcome of interest by the day of the week on which it occurs.

This output file should include a row for each unique combination of cohort (*exposure_group*), incidence level for the tree (*incident_level*), exposure combination identification code (*exp_comb*), HOI (*dx*), and HOI day of the week (*dx_wkday*).

Table 17. DX_WKDAY Specification

Parameter	Field Name	Description
Cohort Identification Number	exposure_group	**Details:** *exposure_group* is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the *exposure_group_name* and *exposure_group*. For example, if there are two values for *exposure_group_name*, then there will be two corresponding values for *exposure_group.* **Format:** Num (8) **Example:** *exposure_group*=1
Incidence Level for the tree	incident_level	**Details:** the level of the tree that defines an incident HOI. The tree as defined in Diagnosis Tree lookup file (DXTREE_INPUT.sas7bdat) has 5 levels. This variable designates that a diagnosis may be incident if there are no diagnoses that share the same node at the *Incident_level* of the tree in the HOI washout period (*AE_wash_up).* Allowable values: • 1: Level 1

Parameter	Field Name	Description
		• 2: Level 2 • 3: Level 3 • 4: Level 4 • 5: Level 5 **Format:** Num (8) **Example:** *Incident_level*=3
Exposure Combination Identification Number	exp_comb	**Details:** concatenation of binary (0/1) indicators for presence or absence of exposure of interest that corresponds to the *exposure_group_name*. For example, if the *exposure_group_name* is TDaP_HPV4 and the result pertains to the ascertainment of a TDaP vaccination without a same-day concomitant HPV4 vaccination, then the *exp_comb*=10. The number of characters is dictated by the *exposure_group_name* that has the most same-day exposure groupings included. For example, if there is a *exposure_group_name* with 3 exposures, then the *exp_comb* file will have 3 characters. **Note 1:** '.' characters indicate that the program is not recording any information on other exposures. For example, if the *exposure_group_name* is TDaP_HPV4 but another *exposure_group_name* in the execution of the program has 3 exposures, then the potential *exp_comb values* for TDaP_HPV4 will be "10." or "11." **Format:** Char (30) **Example:** *exp_comb* = 101
Ascertained Incident Health Outcome of Interest	dx	**Details:** ascertained incident health outcome of interest that is included in the analytic dataset. It will be a string, and decimals will not be compressed. **Format:** Char (18) **Example:** *dx* = 780.2
Day of the Week of Ascertained Health Outcome of Interest	dx_wkday	**Details:** stratification of all ascertained health outcomes of interest by the day of the week of their occurrence. **Note 1:** 1=Sunday **Format:** Num (8) **Example:** *dx_wkday* =3

Parameter	Field Name	Description
Eligible Outcomes Count	COUNT	**Details:** Counts of ascertained health outcomes of interest that fall on a particular weekday. **Format:** Num (8) **Example:** *COUNT*=8
Eligible Outcomes Percentage	PERCENT	**Details:** Percentage of ascertained health outcomes of interest that fall on a particular weekday. **Note 1:** Percentages are calculated within *exposure_group* such that all percentages associated with a particular *exposure_group* add up to 100. **Note 2:** Valid values will be between 0 and 100. **Format:** Num (8) **Example:** *PERCENT*=2.5

7. Exposure Day of the Week Distribution

The Exposure Day of the Week Distribution output file (EXP_WKDAY.sas7bdat) includes the distribution of exposure day of the week for each valid exposure.

This output file should include a row for each unique combination of cohort *(exposure_group)*, exposure combination identification code (*exp_comb*), and exposure day of the week (*exp_wkday*).

Table 18. EXP_WKDAY Specification

Parameter	Field Name	Description
Cohort Identification Number	exposure_group	**Details:** *exposure_group* is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the *exposure_group_name* and *exposure_group*. For example, if there are two values for *exposure_group_name*, then there will be two corresponding values for *exposure_group.* **Format:** Num (8) **Example:** *exposure_group*=1
Exposure Combination Identification Number	exp_comb	**Details:** concatenation of binary (0/1) indicators for presence or absence of exposure of interest that corresponds to the *exposure_group_name*. For example, if the *exposure_group_name* is TDaP_HPV4 and the result pertains to the ascertainment of a TDaP vaccination without a same-day concomitant HPV4 vaccination, then the *exp_comb*=10.

Parameter	Field Name	Description
		The number of characters is dictated by the *exposure_group_name* that has the most same-day exposure groupings included. For example, if there is a *exposure_group_name* with 3 exposures, then the *exp_comb* file will have 3 characters. **Note 1:** '.' characters indicate that the program is not recording any information on other exposures. For example, if the *exposure_group_name* is TDaP_HPV4 but another *exposure_group_name* in the execution of the program has 3 exposures, then the potential *exp_comb values* for TDaP_HPV4 will be "10." or "11." **Format:** Char (30) **Example:** *exp_comb* = 101
Day of the Week of Eligible Exposure	exp_wkday	**Details:** stratification of all eligible exposures by the day of the week of their administration. **Note 1:** 1=Sunday **Format:** Num (8) **Example:** *exp_wkday* =3
Eligible Exposures Count	COUNT	**Details:** Counts of eligible exposures that fall on a particular day of the week. **Format:** Num (8) **Example:** *COUNT* =8
Eligible Exposures Percentage	PERCENT	**Details:** Percentage of eligible exposures that fall on a particular day of the week **Note 1:** Percentages are calculated within *exposure_group* such that all percentages associated with a particular *exposure_group* add up to 100. **Note 2:** Valid values will be between 0 and 100. **Format:** Num (8) **Example:** *PERCENT* =2.5

8. Cohort Exposure Attrition

The Cohort Exposure Attrition output file (EXPOSURE_GROUP_PROCESS_FLOW.sas7bdat) includes counts of unique patients and exposures at key points in the hierarchical process having:

1. any exposure
2. primary exposure
3. incident primary exposure
4. incident primary exposure meeting drug enrollment and demographic eligibility
5. eligible exposure meeting medical enrollment eligibility
6. eligible exposures after applying exclusions

These counts will allow investigators to determine the number of patients and exposures excluded when applying each additional restriction.

This output file should include a row for each unique cohort *(exposure_group)*.

Table 19. EXPOSURE_GROUP_PROCESS_FLOW Specification

Parameter	Field Name	Description
Cohort Identification Number	exposure_group	**Details:** *exposure_group* is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the *exposure_group_name* and *exposure_group*. For example, if there are two values for *exposure_group_name*, then there will be two corresponding values for *exposure_group.* **Format:** Num (8) **Example:** *exposure_group*=1
Cohort Name	exposure_group_name	**Details:** name of the cohort of interest defined as either a singular exposure of interest or a grouping of same-day exposures. Defining unique *exposure_group_name* values allows for the specification of multiple cohorts in a single execution of the program package. **Note 1**: The ordering of exposures in *exposure_group_name* is purposeful and intended to aid the SOC request programmer and end-user. If the cohort of interest is a grouping of same day exposures, then the exposures are separated by underscores but concatenated in a single name. The first listed exposure of interest is the *primary* exposure of interest, which is explained below. For example, if the *exposure_group_name* = TDaP_HPV4, then the primary exposure of interest is TDaP, but there is also an interest

Parameter	Field Name	Description
		in collecting data on same-day exposures to HPV4. **Format:** Char (30) **Example:** *exposure_group_name* = TDaP_HPV4
Count of Exposed Patients	exp_pnt_cnt	**Details:** Counts of unique patients having at least one exposure. **Format:** Num (8) **Example:** *exp_pnt_cnt* =8
Count of Exposures	exp_event_cnt	**Details:** Counts of unique exposures. **Format:** Num (8) **Example:** *exp_event_cnt* =8
Count of Patients with Primary Exposure	prim_exp_pnt_cnt	**Details:** Counts of unique patients having at least one *primary* exposure. **Format:** Num (8) **Example:** *prim_exp_pnt_cnt* =8
Count of Primary Exposures	prim_exp_event_cnt	**Details:** Counts of unique *primary* exposures. **Format:** Num (8) **Example:** *prim_exp_event_cnt* =8
Count of Patients with Incident Exposures	inc_exp_pnt_cnt	**Details:** Counts of unique patients having at least one *incident* primary exposure. **Format:** Num (8) **Example:** *COUNT* =8
Count of Incident Exposures	inc_exp_event_cnt	**Details:** Counts of unique *incident* primary exposures. **Format:** Num (8) **Example:** *inc_exp_event_cnt* =8
Count of Patients having Exposures with Eligible Drug Enrollment and Demographics	drug_enr_demo_elig_pnt_cnt	**Details:** Counts of unique patients having at least one incident primary exposure that meets *drug enrollment and demographic eligibility* criteria. **Format:** Num (8) **Example:** *drug_enr_demo_elig_pnt_cnt* =8
Count of Exposures with Eligible Drug Enrollment and Demographics	drug_enr_demo_exp_event_elig	**Details:** Counts of unique incident primary exposures that meet *drug enrollment and demographic eligibility* criteria. **Format:** Num (8) **Example:** *drug_enr_demo_exp_event_elig* =8
Count of Patients having Exposures with Eligible	med_enr_demo_elig_pnt_cnt	**Details:** Counts of unique patients having at least one eligible exposure that also meets *medical enrollment eligibility* criteria.

Parameter	Field Name	Description
Medical Enrollment		**Format:** Num (8) **Example:** *med_enr_demo_elig_pnt_cnt* =8
Count of Exposures with Eligible Medical Enrollment	med_enr_demo_exp_event_elig	**Details:** Counts of unique eligible exposures that also meet *medical enrollment eligibility* criteria. **Format:** Num (8) **Example:** *med_enr_demo_exp_event_elig* =8
Count of Patients having Eligible Exposures after Exclusions	elig_pnt_cnt_after_excl	**Details:** Counts of unique patients having at least one fully-eligible exposure *after applying exclusions*. **Format:** Num (8) **Example:** *elig_pnt_cnt_after_excl* =8
Count of Eligible Exposures after Exclusions	exp_event_elig_after_excl	**Details:** Counts of unique fully-eligible exposures *after applying exclusions*. **Format:** Num (8) **Example:** *exp_event_elig_after_excl* =8

9. Log File

The log file (TREE_EXTRACTION_SCRI.log) includes a record of the commands in the tree_extraction_SCRI.sas main program. Additionally, the log file records warnings and errors generated by SAS.

10. Signature File

The signature file (MS_TIME_SIGNATURE.sas7bdat) contains metadata associated with the request, including request identifiers, program identifiers, database version, and run time metrics.

B. DPLOCAL OUTPUT

1. Analytic Dataset Crosswalk

The Analytic Dataset Crosswalk file (CRSWLK_FOR_FREEZE.sas7bdat) identifies all exposures that have contributing outcomes to the TreeScan analytic dataset. It is preserved at the Data Partner site to allow re-identification of these exposures should more extensive follow-up be required.

Table 20. CRSWLK_FOR_FREEZE Specification

Parameter	Field Name	Description
Cohort Name	exposure_group	**Details:** *exposure_group* is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the *exposure_group_name* and *exposure_group*. For example, if there are two values for *exposure_group_name*, then there will be two corresponding values for *exposure_group.* **Format:** Num (8) **Example:** *exposure_group*=1
Incidence Level for the tree	incident_level	**Details:** the level of the tree that defines an incident HOI. The tree as defined in Diagnosis Tree lookup file (DXTREE_INPUT.sas7bdat) has 5 levels. This variable designates that a diagnosis may be incident if there are no diagnoses that share the same node at the *Incident_level* of the tree in the HOI washout period (*AE_wash_up).* Allowable values: • 1: MLCCS Level 1 • 2: MLCCS Level 2 • 3: MLCCS Level 3 • 4: MLCCS Level 4 • 5: MLCCS Level 5 **Format:** Num (8) **Example:** *Incident_level*=3
Exposure Combination Identification Number	exp_comb	**Details:** concatenation of binary (0/1) indicators for presence or absence of exposure of interest that corresponds to the *exposure_group_name*. For example, if the *exposure_group_name* is TDaP_HPV4 and the result pertains to the ascertainment of a TDaP vaccination without a same-day concomitant HPV4 vaccination, then the *exp_comb*=10. The number of characters is dictated by the *exposure_group_name* that has the most same-day exposure groupings included. For example, if there is a *exposure_group_name* with 3 exposures, then the *exp_comb* file will have 3 characters. **Note 1:** '.' characters indicate that the program is not recording any information on

Parameter	Field Name	Description
		other exposures. For example, if the *exposure_group_name* is TDaP_HPV4 but another *exposure_group_name* in the execution of the program has 3 exposures, then the potential *exp_comb values* for TDaP_HPV4 will be "10." or "11." **Format:** Char (30) **Example:** *exp_comb* = 101
Unique Patient Identifier	PATID	**Details:** unique patient identifier. **Format:** Char (30) **Example:** *PATID* = APEIufyq39845
Incident node	mlccs_incident	**Details:** the incident node in the Multi-Level Clinical Classification System (MLCCS) that contains the ascertained health outcome of interest **Format:** Char (30) **Example:** *mlccs_incident* = 10.01.08
Original Incident Health Outcome of Interest	orig_dx	**Details:** original health outcome of interest that must be converted into its ICD-9-CM equivalent using the ICD10-ICD9 Mapping lookup table (ID10_ID9_MAP.sas7bdat) to be included in the analytic dataset. It will be a string. **Note 1**: If the *orig_dx* happens to be an ICD-9-CM code, then no conversion is necessary. **Format:** Char (18) **Example:** *orig_dx* = 780.2 (if ICD-9-CM) *orig_dx*=R56.0 (if ICD-10-CM)
Original Diagnosis Code Type	orig_dx_codetype	**Details:** Code type required by lookup file. Valid values include: • **09**: ICD-9-CM • **10**: ICD-10-CM **Format:** Char (2) **Example:** *orig_dx_codetype*=09
Ascertained Incident Health Outcome of Interest	dx	**Details:** incident health outcome of interest that is included in the analytic dataset. It will be a string. **Note 1**: If the *orig_dx* happens to be an ICD-10-CM code, then it will be converted into its ICD-9-CM equivalent code using the ICD10-

Parameter	Field Name	Description
		ICD9 Mapping lookup table (ID10_ID9_MAP.sas7bdat). **Format:** Char (18) **Example:** *dx* = 780.2
Ascertained Diagnosis Code Type	dx_codetype	**Details:** code type that is used to populate the Diagnosis Tree (DXTREE_INPUT.sas7bdat). Valid values include: • **09**: ICD-9-CM **Format:** Char (2) **Example:** *dx_codetype*=09
Date for Incident Health Outcome of Interest	dx_adate	**Details:** date of the qualifying incident health outcome of interest. **Format:** SAS date (date9.) **Example:** *dx_adate*=08/15/2004
Date for Exposure of Interest	exp_adate	**Details:** date for the exposure of interest. **Format:** SAS date (date9.) **Example:** *exp_adate*=08/15/2004

2. LST File

This is a free text SAS output that is created when "PROC FREQ" or other procedures print an automatic SAS output during program execution.

VII. PROGRAM STEPS

1. Check site-specific information.
2. Read input files and set default parameters for any missing information.
3. Create unique combinations of parameters for the various *exposure_group*(s) (i.e., cohorts) for later use in the program.
4. Process the DXTREE_PRIORITY.sas7bdat file with the *Incident_level* parameters to identify the correct priority list for each *exposure_group* for tie-breaker procedures.
5. Combine ICD-9-CM diagnosis tree codes with ICD-10-CM diagnosis tree codes according to ID10-ID9_MAPPING.sas7bdat file.
6. Gather relevant data from SCDM tables. To save computation time, TreeExtraction first identifies all relevant diagnosis (DX), procedure (PX), and dispensing (RX) codes based on input files to identify any exposures or outcomes being considered for all *exposure_group*(s) being covered. From these inputs, the TreeExtraction program subsets the SCDM utilization tables to the data that are necessary for the entire subsequent TreeExtraction execution.
7. Gather all records with relevant codes as determined by prior step.

8. Delete patients according to the ASO_EXCL_FLAG input file if *ASO_excl_flag*=1.
9. Create table based on exposure records linked to *exposure_group* and exclusion *subgroup* using EXCLUSION_GROUP.sas7bdat parameters.
10. Subset pool of patients to those that have *primary* exposure event per *exposure_group* and attach *exposure_group* parameters to each record. Given that this is a SCRI design, only members with the exposure of interest will be included in the study population. The exposure of interest will be identified using a combination of diagnosis, procedure, and dispensing codes. Version 1.3 of the program can only identify an exposure using "OR" Boolean logic. It does not have the capability to identify exposures with "AND" logic.
11. Limit *primary* exposure of interest to only incident exposures as specified by *exp_wash_up*.
12. Add enrollment and demographics tables to the patients previously identified.
13. Create enrollment episodes for each potential member that could be included in the cohort. Bridge gaps less than or equal to *enr_gap* days.
14. Make birth adjustments to enrollment criteria.
15. Determine that member meets enrollment criteria for study population. For members with enrollment within the study period dates, members will have a required pre-exposure enrollment period. The minimum pre-exposure enrollment is calculated as the maximum of exposure washout (*exp_wash_up)* OR the health outcome of interest washout (*AE_wash_up)* – observation window startup (*F_up_win_start*). They also have a required post-exposure enrollment period (*F_up_window_end*) to ensure complete follow-up. During this time, members must have the required coverage (*medcov, drugcov*) continuously with an allowance for membership coverage gaps (*Enr_gap*). Because of the requirements for continuous coverage, all members will have a minimum of MAX(*exp_wash_up* , *AE_wash_up* - *F_up_win_start*) + *F_up_window_end* + 1 (for day zero or day of exposure) days of coverage (i.e., with allowances for gaps).
16. Determine that member meets age criteria for *exposure_group* on the date of exposure. Members of the cohort will be required to be in an identified age group of interest using the *age_begin*, *age_finish*, and *age_period* parameters.
17. Exclude records with prior exposures based on exclusion criteria. Add non-primary exposures to the selected *primary* exposures for each *exposure_group.*
18. Calculate descriptive statistics for each *exposure_group*. Note: All members with eligible exposures that meet all enrollment criteria and have had the exposure of interest in the appropriate age range will be included in the calculation of any descriptive statistics.
    a. Create Cohort Exposure Attrition file (EXPOSURE_GROUP_PROCESS_FLOW.sas7bdat) that includes counts of unique patients and exposures at key points in the hierarchical process.
    b. Calculate number of eligible exposures that are "at risk" to experience any health outcome of interest and output to ELIG_EXP_DOSES.sas7bdat.
    c. Calculate the patient age at date of exposure and output to EXP_AGE.sas7bdat.
    d. Determine the day of the week of the exposure (e.g., Thursday) and output to EXP_WKDAY.sas7bdat.

19. Find eligible incident health outcomes of interest. For an outcome to "count" or make it into the analytic dataset, it has to occur to a member that has passed all enrollment criteria and had a qualifying exposure of interest. Given these requirements, the HOI further has to occur in the follow-up window defined by *F_up_window_st, F_up_window_end* in the user-defined *Incident_enc_setting*. The follow-up windows should always be anchored by the assumption that the exposure date is day 0. If a qualifying outcome occurs in this window in the appropriate setting, it further has to meet incidence criteria. To be considered incident, a diagnosis must be not have occurred in the *AE_wash_up* days in any setting(s) at the *Incident_level* using the DXTREE_INPUT.sas7bdat file. Break any ties for same day incident diagnoses that occur using the *priority* parameter.
20. Create EXPOSURE_GROUP_LEVEL_OUTC.sas7bdat file with patient and event counts for exposure/AE pairs by *exposure _group* and specified *incident_level*.
21. Calculated the number of HOI (*orig_dx and dx*) per day of the week (e.g., Thursday) and output to ORIG_DX_WKDAY.sas7bdat. and DX_WKDAY.sas7bdat.
22. Create CRSWLK_FOR_FREEZE.sas7bdat file. This is the dataset kept locally at Data Partner that might be used at a later date for further alert follow-up
23. Create SELF_CONTROL_ORIG_DX.sas7bdat file. This is the analytic dataset file that will be returned to SOC. The counts of events are summarized on the level of the original dx – ICD-9-CM or ICD-10-CM.