

# SENTINEL REUSABLE PROGRAMS

## TreeExtraction Program

**Prepared by the Sentinel Operations Center**

**June 30, 2018**

**Version: 1.4.0**

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
1.1.0	3/30/2015	<ul style="list-style-type: none"> <li>Original Program</li> </ul>	Sentinel Operations Center
1.2.0	6/30/2016	<ul style="list-style-type: none"> <li>Changed structure of input files, creating more of them with individual functionality</li> <li>Added new variable (<i>exp_wash_up</i>) and feature – incident exposure based on a set of number of days</li> <li>Removed variable (<i>enr_ramp_up</i>), which represented enrollment washout. Enrollment washout is now determined in the program based on a combination of other variables</li> <li>Added the ability to collect data for multiple episodes of exposure</li> <li>Added exposure exclusion feature that creates pre-exposure exclusions relative to the index date</li> </ul>	Sentinel Operations Center
1.3.0	8/17/2016	<ul style="list-style-type: none"> <li>Changed variable names to be less specific to vaccine work</li> <li>Added age stratification and incident level to results file</li> <li>Added ability to use ICD-9-CM only or ICD-9-CM and ICD-10-CM together</li> <li>Added cohort attrition table output</li> <li>Modified exclusion options</li> <li>Modified results output so more scenarios can be included within the same execution of the program.</li> <li>Added <i>exclusion_distance</i> variable to Exclusion Exposure Information input file, then used value to create a washout period for exclusions</li> </ul>	Sentinel Operations Center
1.4.0	6/30/2018	<ul style="list-style-type: none"> <li>Change the tree input file to a two-column, parent-child format regardless of actual number of tree levels</li> <li>Add option to allow user to supply the code priority file or have the program create the priority file from the underlying data</li> <li>Add primary diagnosis (<i>PDX</i>) option to define outcomes and exclusions</li> <li>Add Adverse Event Wash-Up Encounter Settings (<i>WASH_UP_ENC_SETTINGS</i>) input file to include</li> </ul>	Sentinel Operations Center

		<p>encounter type and primary diagnosis in definition of events in the wash-up period</p> <ul style="list-style-type: none"> <li>• Add <i>time_to_censor</i> variable to main output table and use new parameter <i>post_exp_min_enr</i> to ensure minimum enrollment in post-exposure period</li> <li>• Exclude exposures for patients having death date &lt; exposure date</li> <li>• In the Exclusion Exposure Information input table, replace exclusion distance with exclusion distance start and end</li> <li>• Add optional parameter to allow selection by patient sex</li> <li>• Add option to link to Cohort Identification and Descriptive Analysis (CIDA) Propensity Score Matching (PSM) results to define cohort and exposure</li> </ul>	
--	--	---	--

## Table of Contents

<b>I.</b>	<b>INTRODUCTION</b> .....	<b>1</b>
<b>II.</b>	<b>PROGRAM PACKAGE</b> .....	<b>1</b>
A.	COMMON COMPONENTS .....	2
B.	NAMING CONVENTIONS .....	2
<b>III.</b>	<b>MASTER PROGRAM PARAMETERS</b> .....	<b>3</b>
<b>IV.</b>	<b>LOOKUP TABLES</b> .....	<b>6</b>
A.	DIAGNOSIS TREE .....	7
B.	ICD10-ICD9 MAPPING.....	7
<b>V.</b>	<b>INPUT FILES</b> .....	<b>8</b>
A.	DATA PARTNER INFORMATION FILE.....	8
B.	GENERAL COHORT INFORMATION FILE .....	10
	NOTE 1: NUMERICALLY LOWER LEVELS ARE CLOSER TO THE ROOT. NUMERICALLY HIGHER LEVELS ARE CLOSER TO THE LEAF.....	15
C.	GENERAL PARAMETERS FILE .....	16
D.	CONCOMITANT EXPOSURE INFORMATION FILE .....	19
E.	EXCLUSION EXPOSURE INFORMATION FILE .....	20
F.	CODE INFORMATION FILE.....	22
G.	ADVERSE EVENT ENCOUNTER SETTING INFORMATION FILE.....	25
H.	ADVERSE EVENT WASH-UP ENCOUNTER SETTING INFORMATION FILE .....	26
I.	DIAGNOSIS PRIORITY FILE.....	27
<b>VI.</b>	<b>OUTPUT FILES</b> .....	<b>29</b>
A.	OUTPUT TO THE SENTINEL OPERATIONS CENTER (MSOC FOLDER FILE).....	29
1.	<i>Analytic Dataset</i> .....	29
2.	<i>Eligible Exposures Descriptive Statistics</i> .....	33
3.	<i>Age at Exposure Distribution</i> .....	35
4.	<i>Exposure Group Incident Level Outcome Summary</i> .....	37
5.	<i>Original Health Outcome of Interest Day of the Week Distribution</i> .....	39
6.	<i>Ascertained Health Outcome of Interest Day of the Week Distribution</i> .....	41
7.	<i>Exposure Day of the Week Distribution</i> .....	43
8.	<i>Cohort Exposure Attrition</i> .....	45
9.	<i>Log File</i> .....	47
10.	<i>Signature File</i> .....	48
B.	DPLOCAL OUTPUT .....	48
1.	<i>Analytic Dataset Crosswalk</i> .....	48
2.	<i>PSM Analytic Dataset Crosswalk</i> .....	51
3.	<i>LST File</i> .....	54
<b>VII.</b>	<b>PROGRAM STEPS</b> .....	<b>54</b>

## 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 hierarchical 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, or it can also be used in conjunction with Sentinel’s Cohort Identification and Descriptive Analysis (CIDA) Propensity Score Matching (PSM) tool (<https://www.sentinelinitiative.org/sentinel/surveillance-tools/routine-querying-tools/routine-querying-system>) to create a 1:1 propensity score matched cohort. The analytic dataset created from the TreeExtraction program will be 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 hierarchical tree of HOIs (input file: CHILD\_PARENT.sas7bdat).

HOIs will be identified and defined using ICD-9/10-CM diagnosis codes, which are organized into a hierarchical tree structure. 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 3<sup>rd</sup> level of the tree (i.e., 06.04.02).

**Table 1: Example Hierarchical Tree Structure**

Node	Level	ICD-9 / 10	Description
06	1 <sup>st</sup>		Diseases of The Nervous System And Sense Organs
06.04	2 <sup>nd</sup>		..Epilepsy; convulsions
06.04.02	3 <sup>rd</sup>		.....Convulsions
06.04.02.00	4 <sup>th</sup>	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 required design (i.e., either SCRI or PSM) 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	<p><b>Details:</b> project identifier for internal SOC identification and tracking.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> Alphanumeric  <b>Example:</b> <i>MSPROJID= tshpv9</i></p>
Work Plan Type	MSWPTYPE	<p><b>Details:</b> work plan type for internal SOC identification and tracking.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> Alphanumeric  <b>Example:</b> <i>MSWPTYPE=ahr</i></p>
Work Plan Identifier	MSWPID	<p><b>Details:</b> work plan identifier for internal SOC identification and tracking.</p> <p><b>Note 1:</b> should follow the format [wp###].</p> <p><b>Note 2:</b> should be used to uniquely identify a modular program request.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> Alphanumeric  <b>Example:</b> <i>MSWPID= wp001</i></p>
Data Partner Identifier	MSDPID	<p><b>Details:</b> Data Partner identifier for internal SOC identification and tracking.</p> <p><b>Note 1:</b> if a package is not Data Partner specific, MSDPID should equal “nsdp”.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> Alphanumeric  <b>Example:</b> <i>MSDPID =nsdp</i></p>

Parameter	Field Name	Description
Version Identifier	MSVERID	<p><b>Details:</b> version identifier for internal SOC identification and tracking. Should track each re-distribution of the package (if multiple distributions are required).</p> <p><b>Note 1:</b> should follow the format [v##].</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> Alphanumeric  <b>Example:</b> <i>MSVERID =v01</i></p>
Common Components Directory	MSCDMPROG	<p><b>Details:</b> directory that contains ms_common_components.sas.</p> <p><b>Defined by:</b> Data Partner  <b>Input type:</b> Required  <b>Format:</b> Alphanumeric  <b>Example:</b> <i>MSCDMPROG =C:\common_components\</i></p>
Patient Exclusion File Location	ASO_EXCL_LIST	<p><b>Details:</b> 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.</p> <p><b>Defined by:</b> Data Partner  <b>Input type:</b> Optional; leave blank if no PatID exclusion required  <b>Format:</b> Alphanumeric  <b>Example:</b> <i>ASO_EXCL_LIST = infolder.ASO_EXCL_LIST</i></p>
CIDA DP Local Directory request identifier	CIDA_MSReqID	<p><b>Details:</b> request identifier for the CIDA request containing the PSM results to be linked.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Optional; leave blank if not linking to PSM results  <b>Format:</b> Alphanumeric  <b>Example:</b> <i>CIDA_MSReqID = cder_mpl2p_wp001_nsdv_v01</i></p>
Diagnosis Tree Lookup Table File Name	child_parent	<p><b>Details:</b> 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.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> Alphanumeric  <b>Example:</b> <i>CHILD_PARENT = infolder.child_parent</i></p>



Parameter	Field Name	Description
ICD10-ICD9 Mapping Lookup Table File Name	id10_id9_map	<p><b>Details:</b> 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.</p> <p><b>Note1:</b> This file is only required if <i>id10_flag</i> = 1 in the General Parameters file.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Optional  <b>Format:</b> Alphanumeric  <b>Example:</b> <i>id10_id9_map</i> = infolder.id10_id9_map</p>
Data Partner Information File Name	site_info	<p><b>Details:</b> 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.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> Alphanumeric  <b>Example:</b> <i>site_info</i>= infolder.site_info</p>
General Cohort Information File Name	general_input	<p><b>Details:</b> 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.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> Alphanumeric  <b>Example:</b> <i>general_input</i>= infolder.general_input</p>
General Parameters File Name	general_parameters	<p><b>Details:</b> name of the input file containing the General Parameters inputs that will be used in the request. Details in the Input Files section below.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> Alphanumeric  <b>Example:</b> <i>general_parameters</i>= infolder.general_parameters</p>
Concomitant Exposure Information File Name	concomitant_group	<p><b>Details:</b> 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.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> Alphanumeric  <b>Example:</b> <i>concomitant_group</i> = infolder.concomitant_group</p>

Parameter	Field Name	Description
Code Information File Name	subgroup	<p><b>Details:</b> name of the input file containing the Code Information inputs that will be used in the request. Details in the Input Files section below.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> Alphanumeric  <b>Example:</b> <i>subgroup</i>= infolder.subgroup</p>
Adverse Event Encounter Setting Information File Name	AE_enc_setting	<p><b>Details:</b> 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.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> Alphanumeric  <b>Example:</b> <i>AE_enc_setting</i>= infolder.AE_enc_setting</p>
Adverse Event Wash-Up Encounter Setting Information File Name	wash_up_enc_setting	<p><b>Details:</b> name of the input file containing the Adverse Event Wash-Up Encounter Setting Information inputs that will be used in the request. Details in the Input Files section below.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> Alphanumeric  <b>Example:</b> <i>wash_up_enc_setting</i> = infolder.AE_enc_setting</p>
Diagnosis Priority File Name	dxtree_priority	<p><b>Details:</b> 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.</p> <p><b>Note:</b> This file will only be used if <i>priority_flag</i>=1 in the General Parameters file.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> Alphanumeric  <b>Example:</b> <i>dxtree_priority</i>= infolder.dxtree_priority</p>

#### IV. LOOKUP TABLES

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

## A. Diagnosis Tree

The Child-Parent Diagnosis Tree lookup table (CHILD\_PARENT.sas7bdat) includes a hierarchical tree of codes that are eligible to be health outcomes of interest. Instead of a horizontal file with one variable per hierarchical level, this vertical file structure contains one record per unique child-parent relationship in the tree. That is, the *parent* variable contains a code representing one numerically lower level (or more aggregated level) on 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. ICD-10-CM codes will be converted into their equivalent ICD-9-CM using the ICD10-ICD9 Mapping lookup file (ID10\_ID9\_MAP.sas7bdat).

This file format is identical in structure to the file required for TreeScan™ Software.

Table 3 includes specifications for this file.

**Table 3: CHILD\_PARENT Specification**

Parameter	Field Name	Description
Child	Child	<p><b>Details:</b> This code represents a level of the tree that is one level numerically higher than the code in the <i>parent</i> variable. Numerically higher values are further away from the root.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> Char (11)  <b>Example:</b> <i>child</i> = 008.61</p>
Parent	parent	<p><b>Details:</b> This code represents a level of the tree that is one level lower than the code in the <i>child</i> variable.</p> <p><b>Note 1:</b> When the numerically lowest level of the tree, i.e. level 1, which is known as the root is in the <i>child</i> variable, then the <i>parent</i> variable will be left blank.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> Char (11)  <b>Example:</b> <i>parent</i> = 09.01.00.00</p>

## 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 General Parameters 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 (CHILD\_PARENT.sas7bdat).

Table 4 includes specifications for this file.

**Table 4: ID10\_ID9\_MAP Specification**

Parameter	Field Name	Description
ICD10 Diagnosis Code Type	orig_dx_codetype	<p><b>Details:</b> Code type required by lookup file.</p> <p>Valid values include:</p> <ul style="list-style-type: none"> <li>• <b>10:</b> ICD-10-CM</li> </ul> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> Char (2)  <b>Example:</b> <i>orig_dx_codetype=10</i></p>
ICD9 Code	dx	<p><b>Details:</b> 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.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> Char (18)  <b>Example:</b> <i>dx = 780.2</i></p>
ICD9 Diagnosis Code Type	dx_codetype	<p><b>Details:</b> Code type required by lookup file.</p> <p>Valid values include:</p> <ul style="list-style-type: none"> <li>• <b>09:</b> ICD-9-CM</li> </ul> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> Char (2)  <b>Example:</b> <i>dx_codetype=09</i></p>

## V. INPUT FILES

There are nine 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	<p><b>Details:</b> standard Sentinel Data Partner identifier.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> Char (2)  <b>Example:</b> <i>DPID</i> = MS</p>
Site ID	SITEID	<p><b>Details:</b> standard Sentinel site identifier.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> Char (4)  <b>Example:</b> <i>SITEID</i> = OC</p>
Cohort Start Date	Cohort_start_dt	<p><b>Details:</b> defines the start date for a study period.</p> <p><b>Note 1:</b> 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).</p> <p><b>Defined by:</b> Requester / SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> SAS Date (Date9.)  <b>Example:</b> <i>Cohort_start_dt</i> = 01JAN2004</p>
Cohort End Date	Cohort_end_dt	<p><b>Details:</b> defines the end date for a study period.</p>

Parameter	Field Name	Description
		<p><b>Note1:</b> 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 follow-up is 365 days and the Cohort End Date is Dec 31, 2012, the latest possible eligible exposure can be Dec 31, 2011.</p> <p><b>Note2:</b> The requester may enter the date in any logical date format. The SOC request programmer will convert that date into a SAS date.</p> <p><b>Defined by:</b> Requester / SOC request programmer  <b>Input type:</b> Required  <b>Format:</b> SAS date (Date9.)  <b>Example:</b> <i>Cohort_end_dt</i> =31DEC2009</p>
Administrative Services Only Requirement	ASO_EXCL_FLAG	<p><b>Details:</b> indicator of whether the Administrative Services Only (ASO) population must be excluded from the request.</p> <p>Allowable values:</p> <ul style="list-style-type: none"> <li>• <b>1:</b> Exclude population</li> <li>• <b>0:</b> Include population</li> </ul> <p><b>Note 1:</b> if the ASO population must be excluded from a request, a dataset including PatID values for exclusion must be included in the <i>inputfiles</i> folder and the dataset name should be included in the appropriate section of the master program.</p> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required (cannot be left blank)  <b>Format:</b> Num (8)  <b>Example:</b> <i>ASO_EXCL_FLAG</i>=1</p>

## B. General Cohort Information File

The General Cohort Information input file (GENERAL\_INPUT.sas7bdat) includes information that defines each unique cohort associated with the program run.

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	<p><b>Details:</b> name of the cohort of interest defined as either a singular exposure of interest or a grouping of same-day exposures. Defining unique <i>exposure_group_name</i> values allows for the specification of multiple cohorts in a single execution of the program package.</p> <p><b>Note 1:</b> The ordering of exposures in <i>exposure_group_name</i> 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 <i>primary</i> exposure of interest, which is explained below. For example, if the <i>exposure_group_name</i> = TDaP_HP4, then the primary exposure of interest is TDaP, but there is also an interest in collecting data on same-day exposures to HPV4.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required (cannot be left blank)  <b>Format:</b> Char (30)  <b>Example:</b> <i>exposure_group_name</i> = TDaP_HP4</p>
Cohort Identification Number	exposure_group	<p><b>Details:</b> <i>exposure_group</i> is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the <i>exposure_group_name</i> and <i>exposure_group</i>. For example, if there are two values for <i>exposure_group_name</i>, then there will be two corresponding values for <i>exposure_group</i>.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required (cannot be left blank)  <b>Format:</b> Num (8)  <b>Example:</b> <i>exposure_group</i>=1</p>

Parameter	Field Name	Description
Enrollment Gap	Enr_gap	<p><b>Details:</b> sets the number of days that will be bridged between two consecutive enrollment periods to create a “continuously enrolled” period. For example, if <i>Enr_gap</i>=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>i.e.</i>, 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.</p> <p><b>Note 1:</b> A gap of 45-days is recommended for most uses.</p> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required (default=45)  <b>Format:</b> Num (8)  <b>Example:</b> <i>Enr_gap</i>=45 (gaps less than or equal to 45 days will be “bridged” to form one “continuously enrolled” sequence)</p>
Health Outcome of Interest Washout Period	AE_wash_up	<p><b>Details:</b> length of the washout period to determine HOI incidence, given in days.</p> <p><b>Note 1:</b> 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.</p> <p><b>Note 2:</b> As a general rule, it is undesirable to have multiple instances of the same HOI of interest appear in the same observation period, which most conservatively can span from Day 0 through <i>F_up_window_end</i>. Therefore, <i>AE_wash_up</i> is recommended to be <math>\geq (F\_up\_window\_end + 1)</math>.</p> <p><b>Named by:</b> Requester  <b>Input type:</b> Required (default=183)  <b>Format:</b> Num (8)  <b>Example:</b> <i>AE_wash_up</i> = 183</p>



Parameter	Field Name	Description
Start of Follow-up Window	F_up_window_st	<p><b>Details:</b> sets the start of the follow-up or observation window during which an incident health outcome of interest could occur in 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 <i>F_up_window_st=7</i>.</p> <p><b>Note1:</b> Valid entries must be greater than or equal to 0.</p> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required (default=1)  <b>Format:</b> Num (8)  <b>Example</b> <i>F_up_window_st=7</i></p>
End of Follow-up Window	F_up_window_end	<p><b>Details:</b> 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 <i>F_up_window_end=28</i>.</p> <p><b>Note 1:</b> 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 <i>F_up_window_st</i>.</p> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required (default=56)  <b>Format:</b> Num (8)  <b>Example:</b> <i>F_up_window_end=28</i></p>
Medical Coverage	Medcov	<p><b>Details:</b> indicates whether continuous enrollment in medical coverage is required.</p> <p>Allowable values:</p> <ul style="list-style-type: none"> <li>• Y: Yes</li> </ul> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required (default=Y)  <b>Format:</b> Char (1)  <b>Example:</b> <i>Medcov=Y</i></p>

Parameter	Field Name	Description
Drug Coverage	Drugcov	<p><b>Details:</b> indicates whether continuous enrollment in drug coverage is required.</p> <p>Allowable values:</p> <ul style="list-style-type: none"> <li>• Y: Yes</li> <li>• &lt;blank&gt;: any coverage is permissible.</li> </ul> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required  <b>Format:</b> Char (1)  <b>Example:</b> <i>Drugcov=Y</i></p>
Pre-Birth Enrollment Allowance	birth_enr_dist_allowed	<p><b>Details:</b> This variable 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.</p> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required (default=30)  <b>Format:</b> Num (8)  <b>Example:</b> <i>birth_enr_dist_allowed=30</i></p>
Post-Birth Enrollment Allowance	enr_birth_grace_period	<p><b>Details:</b> 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.</p> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required (default=45)  <b>Format:</b> Num (8)  <b>Example:</b> <i>enr_birth_grace_period=45</i></p>
Post-Exposure Minimum Enrollment Days	post_exp_min_enr	<p><b>Details:</b> minimum number of days the patient must be enrolled after exposure</p> <p><b>Note 1:</b> To ensure only that a patient was enrolled on the exact date of exposure, enter 0. To ensure that a patient was enrolled on the day after exposure, enter 1. Therefore, to ensure a patient was enrolled at least <i>N</i> days <i>after</i> exposure, enter a value of <i>N</i>.</p> <p><b>Defined by:</b> Requester</p>

Parameter	Field Name	Description
		<p><b>Input type:</b> Required  <b>Format:</b> Num (8)  <b>Example:</b> <i>post_exp_min_enr =45</i></p>
Incidence Level for the tree	Incident_level	<p><b>Details:</b> the level of the tree that defines an incident HOI. This parameter designates that a diagnosis may be incident if there are no diagnoses that share the same node at the <i>Incident_level</i> of the tree in the HOI washout period (<i>AE_wash_up</i>).</p> <p><b>Note 1:</b> Numerically lower levels are closer to the root. Numerically higher levels are closer to the leaf.</p> <p>Allowable values for an <i>N</i>-level tree:</p> <ul style="list-style-type: none"> <li>• <b>1:</b> Tree Level 1</li> <li>• <b>2:</b> Tree Level 2</li> <li>• ...</li> <li>• <b>N:</b> Tree Level <i>N</i></li> </ul> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required  <b>Format:</b> Num (8)  <b>Example:</b> <i>Incident_level=3</i></p>
Exposure Washout Period	exp_wash_up	<p><b>Details:</b> length in days of the washout period to determine exposure incidence. Applies only to the exposure of interest that is designated as <i>primary</i>.</p> <p><b>Note 1:</b> This period is indexed to the identification of the exposure. The exposure that must be incident is given by the <i>primary</i> value in the CONCOMITANT_GROUP.sas7bdat input file for the <i>exposure_group</i> (i.e., cohort identification number of interest). The <i>primary</i> value corresponds to a <i>subgroup</i> index number. The SUBGROUP.sas7bdat file contains the code list for the exposure of interest. Incidence is assessed with respect to the <i>primary</i> exposure only, not the combination of exposures.</p> <p><b>Note2:</b> If a request wants to ensure that there are not multiple exposures in the</p>

Parameter	Field Name	Description
		<p>same observation window, it is necessary to set <math>exp\_wash\_up \geq (F\_up\_window\_end + 1)</math>.</p> <p><b>Named by:</b> Requester  <b>Input type:</b> Required  <b>Format:</b> Num (8)  <b>Example:</b> <math>exp\_wash\_up = 42</math></p>

### C. General Parameters File

The General Parameters input file (GENERAL\_PARAMETERS.sas7bdat) includes information on parameters that apply to the overall program run, regardless of cohort.

This input file should include only one record. Table 7 includes specifications for this file.

**Table 7: GENERAL\_PARAMETERS Specification**

Parameter	Field Name	Description
Start of the Age Group of Interest	Age_begin	<p><b>Details:</b> The earliest age of eligibility to be included in the program run. Since age is determined at exposure, patient age must be at least this age on the exposure date in order to be eligible for inclusion.</p> <p><b>Note 1:</b> Each run of TreeExtraction can include only one age range.</p> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required (default=0)  <b>Format:</b> Num (8)  <b>Example:</b> <math>Age\_begin=9</math></p>
End of the Age Group of Interest	Age_finish	<p><b>Details:</b> The latest age of eligibility to be included in the program run. Since age is determined at exposure, patient age must be no greater than this age on the exposure date in order to be eligible for inclusion.</p> <p><b>Note 1:</b> Each run of TreeExtraction can include only one age range.</p> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required (default=160)  <b>Format:</b> Num (8)  <b>Example:</b> <math>Age\_finish=26</math></p>
Time Increment for use with the Age Variables	Age_period	<p><b>Details:</b> This variable indicates the time increment to be associated with the <math>Age\_begin</math> and <math>Age\_finish</math>. It can be entered in any recognized standard SAS unit of date</p>

Parameter	Field Name	Description
		<p>and time (e.g., DAY, YEAR, WEEK, HOUR, etc.)</p> <p><b>Note 1:</b> time intervals are not case-sensitive.</p> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required (default=YEAR)  <b>Format:</b> Char (12)  <b>Example:</b> <i>Age_period</i>=YEAR</p>
Patient sex	sex	<p><b>Details:</b> Sex restriction for the program run. Leave blank to select all patients.</p> <p><b>Valid values:</b></p> <ul style="list-style-type: none"> <li>• M: male</li> <li>• F: female</li> <li>• U: Unknown</li> <li>• A: Ambiguous</li> </ul> <p><b>Named by:</b> Requester  <b>Input type:</b> Optional  <b>Format:</b> Char (1)  <b>Example:</b> <i>sex</i> = F</p>
ICD10 Requirement	id10_flag	<p><b>Details:</b> indicator of whether ICD-10-CM codes will be included in this request.</p> <p>Allowable values:</p> <ul style="list-style-type: none"> <li>• <b>1:</b> ICD-10-CM included</li> <li>• <b>0:</b> ICD-10-CM not included</li> </ul> <p><b>Note 1:</b> 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.</p> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required (cannot be left blank)  <b>Format:</b> Num (8)  <b>Example:</b> <i>id10_flag</i>=1</p>
Diagnosis Priority File included indicator	priority_file	<p><b>Details:</b> indicator of whether the Diagnosis Priority File has been supplied by the request programmer</p> <p>Allowable values:</p> <ul style="list-style-type: none"> <li>• <b>1:</b> Diagnosis Priority File included</li> </ul>

Parameter	Field Name	Description
		<ul style="list-style-type: none"> <li>• <b>0:</b> Diagnosis Priority File <i>not</i> included, will be generated by the program</li> </ul> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required (cannot be left blank)  <b>Format:</b> Num (8)  <b>Example:</b> <i>priority_file = 1</i></p>
Age stratifications	timestrat	<p><b>Details:</b> stratification of age at index date, i.e. exposure.</p> <p><b>Note 1:</b> Program will use units specified by <i>Age_period</i></p> <p><b>Defined by:</b> Requester  <b>Input type:</b> Optional  <b>Format:</b> Char (23)  <b>Example:</b> <i>timestrat = 18-44 45-54 55-64 65+</i></p>
Diagnosis tree codetype	tree_dx_codetype	<p><b>Details:</b> codetype of diagnosis codes in the Child-Parent Diagnosis Tree lookup file.</p> <p>Allowable values:</p> <ul style="list-style-type: none"> <li>• <b>09:</b> ICD-9</li> <li>• <b>10:</b> ICD-10</li> </ul> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required (cannot be left blank)  <b>Format:</b> Char (2)  <b>Example:</b> <i>tree_dx_codetype = 09</i></p>
Propensity Score Matching link flag	PS_flag	<p><b>Details:</b> indicator of whether the program will link to PSM results</p> <p>Allowable values:</p> <ul style="list-style-type: none"> <li>• <b>1:</b> PSM results <i>will</i> be linked</li> <li>• <b>0:</b> PSM results <i>will not</i> be linked</li> </ul> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required (cannot be left blank)  <b>Format:</b> Num (8)  <b>Example:</b> <i>PS_flag = 1</i></p>
CIDA Run Identifier	runid	<p><b>Details:</b> run identifier for the original CIDA run containing the PSM results to be linked</p> <p><b>Defined by:</b> Requester  <b>Input type:</b> Optional (required if <i>PS_flag = 1</i>)</p>

Parameter	Field Name	Description
		<b>Format:</b> Char (3) <b>Example:</b> <i>runid</i> = r01

#### D. Concomitant Exposure Information File

The optional 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 file will not be used if linking to CIDA PSM results.

This input file should include a unique row for each *exposure\_group* and *exposure\_order*. Table 8 includes specifications for this file.

**Table 8: CONCOMITANT\_GROUP Specification**

Parameter	Field Name	Description
Cohort Identification Number	<i>exposure_group</i>	<p><b>Details:</b> <i>exposure_group</i> is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the <i>exposure_group_name</i> and <i>exposure_group</i>. For example, if there are two values for <i>exposure_group_name</i>, then there will be two corresponding values for <i>exposure_group</i>.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required (cannot be left blank)  <b>Format:</b> Num (8)  <b>Example:</b> <i>exposure_group</i>=1</p>
Exposure Name	<i>group</i>	<p><b>Details:</b> name of one of the exposures of interest within a cohort as defined by <i>exposure_group</i>. For each <i>exposure_group_name</i>, there may be a grouping of same-day concomitant exposures of interest connected by underscores. The <i>group</i> should correspond to a single exposure defined using codes in the Code Information input file (SUBGROUP.sas7bdat) given by the <i>subgroup</i> value.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required (cannot be left blank)  <b>Format:</b> Char (30)  <b>Example:</b> <i>group</i> = HPV</p>

Primary Exposure of Interest Indicator	primary	<p><b>Details:</b> binary variable (1/0) indicating whether the <i>group</i> exposure is the primary exposure of interest. The exposure of interest designated as primary is the only exposure to which <i>exp_wash_up</i> parameters apply.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required (cannot be left blank)  <b>Format:</b> Num (8)  <b>Example:</b> <i>primary = 1</i></p>
Exposure Order Indicator	exposure_order	<p><b>Details:</b> ordinal variable that identifies the ordering of the exposures in a <i>exposure_group</i>. Exposures designated as <i>primary</i> will always be assigned <i>exposure_order=1</i>.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required (cannot be left blank)  <b>Format:</b> Num (8)  <b>Example:</b> <i>exposure_order = 1</i></p>
Code List Indicator	subgroup	<p><b>Details:</b> <i>subgroup</i> links to the SUBGROUP.sas7bdat input files. The <i>subgroup</i> value indicates the appropriate code list to define a particular exposure (i.e., <i>group</i>) within a cohort (i.e., <i>exposure_group</i>). For example, if <i>exposure_group=1</i>, <i>group=HPV</i> and <i>subgroup=1</i>, the code list to define HPV will be in the SUBGROUP.sas7bdat input file with <i>subgroup=1</i>.</p> <p><b>Note1:</b> The number of subgroup values will depend on the number of independent exposures of interest after all the <i>exposure_group(s)</i> are considered along with any exclusion exposures of interest.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required (cannot be left blank)  <b>Format:</b> Num (8)  <b>Example:</b> <i>subgroup=1</i></p>

### E. Exclusion Exposure Information File

The optional 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



time window 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 file will not be used if linking to CIDA PSM results.

This input file should include a unique row for each *exposure\_group* and *subgroup*. Table 9 includes specifications for this file.

**Table 9: GROUP\_EXCLUSION Specification**

Parameter	Field Name	Description
Cohort Identification Number	exposure_group	<p><b>Details:</b> <i>exposure_group</i> is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the <i>exposure_group_name</i> and <i>exposure_group</i>. For example, if there are two values for <i>exposure_group_name</i>, then there will be two corresponding values for <i>exposure_group</i>.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required (cannot be left blank)  <b>Format:</b> Num (8)  <b>Example:</b> <i>exposure_group</i>=1</p>
Code List Indicator	subgroup	<p><b>Details:</b> <i>subgroup</i> links to the SUBGROUP.sas7bdat input files. The <i>subgroup</i> value indicates the appropriate code list to define a particular exposure (i.e., <i>group</i>) within a cohort (i.e., <i>exposure_group</i>). For example, if <i>exposure_group</i>=1, <i>group</i>=HPV and <i>subgroup</i>=1, the code list to define HPV will be in the SUBGROUP.sas7bdat input file with <i>subgroup</i>=1.</p> <p><b>Note1:</b> The number of subgroup values will depend on the number of independent exposures of interest after all the <i>exposure_group</i>(s) are considered along with any exclusion exposures of interest.</p> <p><b>Note2:</b> Exclusions will be applied during the calendar period prior to the <i>primary</i> exposure of interest defined by <i>excl_start_days</i> and <i>excl_end_days</i>.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required (cannot be left blank)</p>

		<b>Format:</b> Num (8) <b>Example:</b> <i>subgroup=1</i>
Start of Exclusion Period Prior to Index Date	excl_dist_start	<b>Details:</b> number of days prior to the primary exposure of interest that defines the <i>lower bound</i> of the exclusion period  <b>Note 1:</b> To start the exclusion period 7 days prior to exposure, enter -7.  <b>Note 2:</b> To allow an exclusion period of 7 days, not counting the day of exposure, enter -8 (and enter -1 for <i>excl_dist_end</i> ).  <b>Defined by:</b> SOC request programmer <b>Input type:</b> Required (cannot be left blank) <b>Format:</b> Num (8) <b>Example:</b> <i>excl_dist_start = -183</i>
End of Exclusion Period Prior to Index Date	excl_dist_end	<b>Details:</b> number of days prior to the primary exposure of interest that defines the <i>upper bound</i> of the exclusion period  <b>Note 1:</b> To end the exclusion period on the primary exposure date, enter 0.  <b>Note 2:</b> To end the exclusion period the day prior to exposure, enter -1.  <b>Defined by:</b> SOC request programmer <b>Input type:</b> Required (cannot be left blank) <b>Format:</b> Num (8) <b>Example:</b> <i>excl_dist_end = -1</i>

#### F. Code Information File

The optional 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 10 includes specifications for this file. This file is required unless linking to CIDA PSM results.

**Table 10: SUBGROUP Specification**

Parameter	Field Name	Description
Code List Indicator	subgroup	<p><b>Details:</b> The <i>subgroup</i> value indicates the appropriate code list to define a particular exposure (i.e., <i>group</i>) within a cohort (i.e., <i>exposure_group</i>). Other files are linked by this value to obtain the list of exposure codes.</p> <p><b>Note1:</b> The number of subgroup values will depend on the number of independent exposures of interest after all the <i>exposure_group(s)</i> are considered along with any exclusion exposures of interest.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required (cannot be left blank)  <b>Format:</b> Num (8)  <b>Example:</b> <i>subgroup=1</i></p>
Code	code	<p><b>Details:</b> National Drug Codes (NDCs), procedure and/or diagnosis codes of interest.</p> <p><b>Note 1:</b> There is no ability to use a wildcard in the program. If an end-user submits a request with a wildcard, then an SOC request programmer must translate that request into codes that can be read by the program.</p> <p><b>Note 2:</b> 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.</p> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required  <b>Format:</b> Char (11)  <b>Example:</b> (<i>code_category=RX; code_type=11</i>), <i>code=12345678911</i></p>
Code Category	code_category	<p><b>Details:</b> type of each code category value included in the <i>code_type</i> field (below) of this file.</p> <p>Valid values include:</p> <ul style="list-style-type: none"> <li>• <b>RX:</b> NDC</li> <li>• <b>DX:</b> Diagnosis code</li> </ul>

Parameter	Field Name	Description
		<ul style="list-style-type: none"> <li>• <b>PX</b>: Procedure code</li> </ul> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required  <b>Format:</b> Char (2)  <b>Example:</b> <code>code_category=PX</code></p>
Code Type	code_type	<p><b>Details:</b> type of each code value included in the <code>code_category</code> field (above) of this file.</p> <p><b>Note1:</b> This variable is optional if all codes in the table have <code>code_category = RX</code>.</p> <p>Valid values include:</p> <p><u>If <code>code_category= RX</code>:</u></p> <ul style="list-style-type: none"> <li>• <b>11</b>: 11-digit NDC</li> </ul> <p><u>If <code>code_category = DX</code>:</u></p> <ul style="list-style-type: none"> <li>• <b>09</b>: ICD-9-CM</li> <li>• <b>10</b>: ICD-10-CM</li> <li>• <b>11</b>: ICD-11-CM</li> <li>• <b>OT</b>: Other</li> </ul> <p><u>If <code>code_category = PX</code>:</u></p> <ul style="list-style-type: none"> <li>• <b>09</b>: ICD-9-CM</li> <li>• <b>10</b>: ICD-10-CM</li> <li>• <b>11</b>: ICD-11-CM</li> <li>• <b>ND</b>: 11-digit NDC Code</li> <li>• <b>C4</b>: CPT-4 (<i>i.e.</i>, HCPCS Level I)</li> <li>• <b>HC</b>: HCPCS (<i>i.e.</i>, HCPCS Level II)</li> <li>• <b>H3</b>: HCPCS Level III</li> <li>• <b>C2</b>: CPT Category II</li> <li>• <b>C3</b>: CPT Category III</li> <li>• <b>RE</b>: Revenue</li> <li>• <b>LO</b>: Local homegrown</li> <li>• <b>OT</b>: Other</li> </ul> <p><b>Defined by:</b> Requester  <b>Input type:</b> Optional  <b>Format:</b> Char (2)  <b>Example:</b> (<code>code_category=DX</code>), <code>code_type=09</code></p>

### G. 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 11 includes specifications for this file.

**Table 11: AE\_ENC\_SETTING Specification**

Parameter	Field Name	Description
Cohort Identification Number	exposure_group	<p><b>Details:</b> <i>exposure_group</i> is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the <i>exposure_group_name</i> and <i>exposure_group</i>. For example, if there are two values for <i>exposure_group_name</i>, then there will be two corresponding values for <i>exposure_group</i>.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required (cannot be left blank)  <b>Format:</b> Num (8)  <b>Example:</b> <i>exposure_group</i>=1</p>
Setting to Define Incidence	incident_enc_setting	<p><b>Details:</b> eligible settings in which an incident health outcome of interest can occur.</p> <p>Allowable values:</p> <ul style="list-style-type: none"> <li>• <b>AV:</b> outpatient</li> <li>• <b>ED:</b> emergency department</li> <li>• <b>IP:</b> inpatient</li> <li>• <b>IS:</b> institutional stay</li> <li>• <b>OA:</b> other ambulatory encounter</li> </ul> <p><b>Note 1:</b> if multiple care settings must be specified, one row per setting per <i>exposure_group</i> must be listed.</p> <p><b>Named by:</b> Requester  <b>Input type:</b> Required  <b>Format:</b> Char (2)  <b>Example:</b> <i>Incident_enc_setting</i>=IP</p>
Primary Diagnosis to Define Incidence	PDX	<p><b>Details:</b> primary diagnosis values which determine whether a health outcome of interest can be counted as incident.</p>

Parameter	Field Name	Description
		<p>Allowable values:</p> <ul style="list-style-type: none"> <li>• <b>P:</b> primary diagnosis required</li> <li>• <b>blank:</b> any value is allowed</li> </ul> <p><b>Note 1:</b> if multiple care settings must be specified, one row per setting per <i>exposure_group</i> must be listed, and this value must be repeated for all applicable care settings.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Optional  <b>Format:</b> Char (1)  <b>Example:</b> <i>PDX</i> = P</p>

#### H. Adverse Event Wash-Up Encounter Setting Information File

The Adverse Event Wash-Up Encounter Setting Information input file (WASH\_UP\_ENC\_SETTING.sas7bdat) includes information on the encounter settings that a HOI must occur in to be counted in the wash-up period. The Adverse Event Wash-Up 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 *wash\_up\_enc\_setting*. Table 12 includes specifications for this file.

**Table 12: WASH\_UP\_ENC\_SETTING Specification**

Parameter	Field Name	Description
Cohort Identification Number	<i>exposure_group</i>	<p><b>Details:</b> <i>exposure_group</i> is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the <i>exposure_group_name</i> and <i>exposure_group</i>. For example, if there are two values for <i>exposure_group_name</i>, then there will be two corresponding values for <i>exposure_group</i>.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required (cannot be left blank)  <b>Format:</b> Num (8)  <b>Example:</b> <i>exposure_group</i>=1</p>
Setting to Define Wash-up Period	<i>wash_up_enc_setting</i>	<p><b>Details:</b> eligible settings in which a health outcome of interest can be counted in the wash-up period.</p> <p>Allowable values:</p> <ul style="list-style-type: none"> <li>• <b>AV:</b> outpatient</li> </ul>

Parameter	Field Name	Description
		<ul style="list-style-type: none"> <li>• <b>ED:</b> emergency department</li> <li>• <b>IP:</b> inpatient</li> <li>• <b>IS:</b> institutional stay</li> <li>• <b>OA:</b> other ambulatory encounter</li> </ul> <p><b>Note 1:</b> if multiple care settings must be specified, one row per setting per <i>exposure_group</i> must be listed.</p> <p><b>Named by:</b> Requester  <b>Input type:</b> Required  <b>Format:</b> Char (2)  <b>Example:</b> <i>wash_up_enc_setting</i> = IP</p>
Primary Diagnosis to Define Wash-up Period	PDX	<p><b>Details:</b> primary diagnosis values which determine whether a health outcome of interest can be counted in the wash-up period.</p> <p>Allowable values:</p> <ul style="list-style-type: none"> <li>• <b>P:</b> primary diagnosis required</li> </ul> <p><b>Note 1:</b> if multiple care settings must be specified, one row per setting per <i>exposure_group</i> must be listed, and this value must be repeated for all applicable care settings.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Optional  <b>Format:</b> Char (1)  <b>Example:</b> <i>PDX</i> = P</p>

### I. 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 CHILD\_PARENT.sas7bdat files; however, these fields are repeated for each *exposure\_group* in the DXTREE\_PRIORITY.sas7bdat file.

The Diagnosis Priority file may be 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. This file may also be created during TreeExtraction program execution on the actual data being queried, but only if the programmer assigns *priority\_file* = 0 in the

General Parameters file. If the program creates the file, a permanent copy will be saved to the DPLocal library. Table 13 includes specifications for this file.

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 CHILD\_PARENT.sas7bdat and the conversion will be done with the ID10\_ID9\_MAP.sas7bdat file.

**Table 13: DXTREE\_PRIORITY Specifications**

Parameter	Field Name	Description
Cohort Identification Number	exposure_group	<p><b>Details:</b> <i>exposure_group</i> is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the <i>exposure_group_name</i> and <i>exposure_group</i>. For example, if there are two values for <i>exposure_group_name</i>, then there will be two corresponding values for <i>exposure_group</i>.</p> <p><b>Defined by:</b> SOC request programmer  <b>Input type:</b> Required (cannot be left blank)  <b>Format:</b> Num (8)  <b>Example:</b> <i>exposure_group</i>=1</p>
Diagnosis Code	dx	<p><b>Details:</b> Diagnosis codes of interest that are input as strings.</p> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required  <b>Format:</b> Char (11)  <b>Example:</b> <i>dx</i>=242.01</p>
Diagnosis Code Type	dx_codetype	<p><b>Details:</b> The code type that is used to populate the Diagnosis Tree.</p> <p>Valid values include:</p> <ul style="list-style-type: none"> <li>• <b>09:</b> ICD-9-CM</li> </ul> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required  <b>Format:</b> Char (2)  <b>Example:</b> <i>dx_codetype</i>=09</p>
Incidence Level for the tree	Incident_level	<p><b>Details:</b> the level of the tree that defines an incident HOI. This parameter designates that a diagnosis may be incident if there are no diagnoses that share the same node at the <i>Incident_level</i> of the tree in the HOI washout period (<i>AE_wash_up</i>).</p> <p>Allowable values for an <i>N</i>-level tree:</p>



Parameter	Field Name	Description
		<ul style="list-style-type: none"> <li>• <b>1</b>: Tree Level 1</li> <li>• <b>2</b>: Tree Level 2</li> <li>• ...</li> <li>• <b>N</b>: Tree Level <i>N</i></li> </ul> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required  <b>Format:</b> Num (8)  <b>Example:</b> <i>Incident_level=3</i></p>
Priority Level of Diagnosis	priority	<p><b>Details:</b> 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 <i>incident_level</i> of the diagnosis tree.</p> <p><b>Note1:</b> 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.</p> <p><b>Defined by:</b> Requester  <b>Input type:</b> Required  <b>Format:</b> Num (8)  <b>Example:</b> <i>priority=33</i></p>

## VI. OUTPUT FILES

There are eight 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.

If CIDA PSM results are linked, the output files may have different structure and naming conventions. See the description of each output file below for details.

### A. Output to the Sentinel Operations Center (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 14: SELF\_CONTROL\_ORIG\_DX Specifications**

Parameter	Field Name	Description
Data Partner ID	DPID	<p><b>Details:</b> standard Sentinel Data Partner identifier.</p> <p><b>Format:</b> Char (2)</p> <p><b>Example:</b> <i>DPID</i> = MS</p>
Site ID	SITEID	<p><b>Details:</b> standard Sentinel site identifier.</p> <p><b>Format:</b> Char (4)</p> <p><b>Example:</b> <i>SITEID</i> = OC</p>
Cohort Identification Number	exposure_group	<p><b>Details:</b> <i>exposure_group</i> is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the <i>exposure_group_name</i> and <i>exposure_group</i>. For example, if there are two values for <i>exposure_group_name</i>, then there will be two corresponding values for <i>exposure_group</i>.</p> <p><b>Format:</b> Num (8)</p> <p><b>Example:</b> <i>exposure_group</i>=1</p>
Cohort Name	exposure_group_name	<p><b>Details:</b> name of the cohort of interest defined as either a singular exposure of interest or a grouping of same-day exposures. Defining unique <i>exposure_group_name</i> values allows for the specification of multiple cohorts in a single execution of the program package.</p> <p><b>Note 1:</b> The ordering of exposures in <i>exposure_group_name</i> 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 <i>primary</i> exposure of interest, which is explained below. For example, if the <i>exposure_group_name</i> = TDaP_HP4, then the primary exposure of interest is TDaP, but there is also an interest in collecting data on same-day exposures to HPV4.</p> <p><b>Format:</b> Char (30)</p> <p><b>Example:</b> <i>exposure_group_name</i> = TDaP_HP4</p>
Incidence Level for the tree	incident_level	<p><b>Details:</b> the level of the tree that defines an incident HOI. This variable designates that a diagnosis may be incident if there are no diagnoses that share the same</p>

Parameter	Field Name	Description
		<p>node at the <i>Incident_level</i> of the tree in the HOI washout period (<i>AE_wash_up</i>).</p> <p>Allowable values for an <i>N</i>-level tree:</p> <ul style="list-style-type: none"> <li>• <b>1</b>: Tree Level 1</li> <li>• <b>2</b>: Tree Level 2</li> <li>• ...</li> <li>• <b>N</b>: Tree Level <i>N</i></li> </ul> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>Incident_level=3</i></p>
Time to censor	time_to_censor	<p><b>Details:</b> number of days after exposure that the patient becomes ineligible due to break in enrollment, patient death, or reaching the end of the study period.</p> <p><b>Note1:</b> Default value is 9999 if the patient is not censored within the minimum required follow-up window.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>time_to_censor=44</i></p>
Exposure Combination Identification Number	exp_comb	<p><b>Details:</b> concatenation of binary (0/1) indicators for presence or absence of exposure of interest that corresponds to the <i>exposure_group_name</i>. For example, if the <i>exposure_group_name</i> is TDaP_HP4 and the result pertains to the ascertainment of a TDaP vaccination without a same-day concomitant HPV4 vaccination, then the <i>exp_comb</i>=10. The number of characters is dictated by the <i>exposure_group_name</i> that has the most same-day exposure groupings included. For example, if there is an <i>exposure_group_name</i> with 3 exposures, then the <i>exp_comb</i> file will have 3 characters.</p> <p><b>Note 1:</b> This variable will not be included when linking with CIDA PSM results.</p> <p><b>Note 2:</b> '.' characters indicate that the program is not recording any information on other exposures. For example, if the <i>exposure_group_name</i> is TDaP_HP4 but another <i>exposure_group_name</i> in the execution of the program has 3 exposures, then the potential <i>exp_comb</i> values for TDaP_HP4 will be "10." or "11."</p>

Parameter	Field Name	Description
		<p><b>Format:</b> Char (30)</p> <p><b>Example:</b> <i>exp_comb</i> = 101</p>
Age Grouping at Index Date	age_group	<p><b>Details:</b> stratification of age at index date, i.e. exposure. Stratifications are defined by <i>timestrat</i> in the General Parameters input file.</p> <p><b>Format:</b> Char (5)</p> <p><b>Example:</b> <i>age_group</i> = 20-24</p>
Original Incident Health Outcome of Interest	orig_dx	<p><b>Details:</b> 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.</p> <p><b>Note1:</b> If the <i>orig_dx</i> happens to be an ICD-9-CM code, then no conversion is necessary, and ICD-9-CM code is used.</p> <p><b>Format:</b> Char (18)</p> <p><b>Example:</b> <i>orig_dx</i> = 780.2 (if ICD-9-CM) <i>orig_dx</i>=R56.0 (if ICD-10-CM)</p>
Original Diagnosis Code Type	orig_dx_codetype	<p><b>Details:</b> Code type required by lookup file.</p> <p>Valid values include:</p> <ul style="list-style-type: none"> <li>• <b>09:</b> ICD-9-CM</li> <li>• <b>10:</b> ICD-10-CM</li> </ul> <p><b>Format:</b> Char (2)</p> <p><b>Example:</b> <i>orig_dx_codetype</i>=09</p>
Ascertained Incident Health Outcome of Interest	dx	<p><b>Details:</b> incident health outcome of interest that is included in the analytic dataset. It will be a string.</p> <p><b>Note1:</b> If the <i>orig_dx</i> 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).</p> <p><b>Format:</b> Char (18)</p> <p><b>Example:</b> <i>dx</i> = 780.2</p>
Ascertained Diagnosis Code Type	dx_codetype	<p><b>Details:</b> code type that is used to populate the Diagnosis Tree (CHILD_PARENT.sas7bdat).</p> <p>Valid values include:</p> <ul style="list-style-type: none"> <li>• <b>09:</b> ICD-9-CM</li> </ul>

Parameter	Field Name	Description
		<b>Format:</b> Char (2) <b>Example:</b> <i>dx_codetype=09</i>
Time to Event for the Ascertained Health Outcome of Interest	days_from_exp_to_outcome	<b>Details:</b> time to event of the incident health outcome of interest that is indexed on the day of exposure where day of exposure=0. <i>days_from_exp_to_outcome</i> must fall within the allowable follow-up window and is given in days. For example, it is 12 if the diagnosis occurred 12 days after vaccination.  <b>Format:</b> Num (8) <b>Example:</b> <i>days_from_exp_to_outcome = 23</i>
Eligible Outcome Counts	exp_dx_dist_cnt	<b>Details:</b> 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.  <b>Format:</b> Num (8) <b>Example:</b> <i>exp_dx_dist_cnt = 2</i>

## 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*).

When linking with CIDA PSM results, the file is named by adding the *MATCHID* name to the end of the filename, i.e. ELIG\_EXP\_DOSES\_[*MATCHID*].sas7bdat, where name depends on the PSM analysis used. Valid values for TreeScan analysis include:

- **PREDEFINEDPSMATCHID1\_1:** predefined covariate estimation model; 1:1 matching strategy
- **HDPSPREDEFINEDMATCHID1\_1:** predefined covariate/ empirically identified covariate hybrid estimation model; 1:1 matching strategy
- **HDPSONLYMATCHID1\_1:** empirically identified covariate estimation model; 1:1 matching strategy

The file structure is identical, except the *exp\_comb* variable does not appear in the CIDA PSM version.

**Table 15: ELIG\_EXP\_DOSES Specifications**

Parameter	Field Name	Description
Cohort Identification Number	exposure_group	<p><b>Details:</b> <i>exposure_group</i> is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the <i>exposure_group_name</i> and <i>exposure_group</i>. For example, if there are two values for <i>exposure_group_name</i>, then there will be two corresponding values for <i>exposure_group</i>.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>exposure_group</i>=1</p>
Cohort Name	exposure_group_name	<p><b>Details:</b> name of the cohort of interest defined as either a singular exposure of interest or a grouping of same-day exposures. Defining unique <i>exposure_group_name</i> values allows for the specification of multiple cohorts in a single execution of the program package.</p> <p><b>Note 1:</b> The ordering of exposures in <i>exposure_group_name</i> 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 <i>primary</i> exposure of interest, which is explained below. For example, if the <i>exposure_group_name</i> = TDaP_HP4, then the primary exposure of interest is TDaP, but there is also an interest in collecting data on same-day exposures to HPV4.</p> <p><b>Format:</b> Char (30)  <b>Example:</b> <i>exposure_group_name</i> = TDaP_HP4</p>
Exposure Combination	exp_comb	<p><b>Details:</b> concatenation of binary (0/1) indicators for presence or absence of exposure of interest that corresponds to</p>

Parameter	Field Name	Description
Identification Number		<p>the <i>exposure_group_name</i>. For example, if the <i>exposure_group_name</i> is TDaP_HP4 and the result pertains to the ascertainment of a TDaP vaccination without a same-day concomitant HPV4 vaccination, then the <i>exp_comb</i>=10. The number of characters is dictated by the <i>exposure_group_name</i> that has the most same-day exposure groupings included. For example, if there is a <i>exposure_group_name</i> with 3 exposures, then the <i>exp_comb</i> file will have 3 characters.</p> <p><b>Note1:</b> ‘.’ characters indicate that the program is not recording any information on other exposures. For example, if the <i>exposure_group_name</i> is TDaP_HP4 but another <i>exposure_group_name</i> in the execution of the program has 3 exposures, then the potential <i>exp_comb values</i> for TDaP_HP4 will be “10.” or “11.”</p> <p><b>Format:</b> Char (30)  <b>Example:</b> <i>exp_comb</i> = 101</p>
Age Grouping at Index Date	age_group	<p><b>Details:</b> stratification of age at index date, i.e. exposure. Stratifications are defined by <i>timestrat</i> in the General Parameters input file.</p> <p><b>Format:</b> Char (5)  <b>Example:</b> <i>age_group</i> = 20-24</p>
Eligible Exposure Dose Count	elig_exp_dose_cnt	<p><b>Details:</b> number of unique episodes of exposure that have met all enrollment and inclusion/exclusion criteria.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>elig_exp_dose_cnt</i> = 45000</p>

### 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*.

**Table 16: EXP\_AGE Specification**

Parameter	Field Name	Description
Cohort Identification Number	exposure_group	<p><b>Details:</b> <i>exposure_group</i> is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the <i>exposure_group_name</i> and <i>exposure_group</i>. For example, if there are two values for <i>exposure_group_name</i>, then there will be two corresponding values for <i>exposure_group</i>.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>exposure_group</i>=1</p>
Exposure Combination Identification Number	exp_comb	<p><b>Details:</b> concatenation of binary (0/1) indicators for presence or absence of exposure of interest that corresponds to the <i>exposure_group_name</i>. For example, if the <i>exposure_group_name</i> is TDaP_HP4 and the result pertains to the ascertainment of a TDaP vaccination without a same-day concomitant HPV4 vaccination, then the <i>exp_comb</i>=10. The number of characters is dictated by the <i>exposure_group_name</i> that has the most same-day exposure groupings included. For example, if there is a <i>exposure_group_name</i> with 3 exposures, then the <i>exp_comb</i> file will have 3 characters.</p> <p><b>Note 1:</b> This variable will not be included when linking with CIDA PSM results.</p> <p><b>Note 2:</b> '.' characters indicate that the program is not recording any information on other exposures. For example, if the <i>exposure_group_name</i> is TDaP_HP4 but another <i>exposure_group_name</i> in the execution of the program has 3 exposures, then the potential <i>exp_comb</i> values for TDaP_HP4 will be "10." or "11."</p> <p><b>Format:</b> Char (30)  <b>Example:</b> <i>exp_comb</i> = 101</p>



Parameter	Field Name	Description
Age at Date of Exposure	age_at_exp_dt	<p><b>Details:</b> This variable indicates age at exposure in the units specified by <i>Age_period</i>.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>age_at_exp_dt</i> =330</p>
Eligible Exposures Count	COUNT	<p><b>Details:</b> Counts of eligible exposures that occur at each <i>age_at_exp_dt</i>.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>COUNT</i> =8</p>
Eligible Exposures Percentage	PERCENT	<p><b>Details:</b> Percentage of eligible exposures that occur at each <i>age_at_exp_dt</i>.</p> <p><b>Note1:</b> Percentages are calculated within <i>exposure_group</i> such that all percentages associated with a particular <i>exposure_group</i> add up to 100.</p> <p><b>Note2:</b> Valid values will be between 0 and 100.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>PERCENT</i> =2.5</p>

#### 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 17: EXPOSURE\_GROUP\_LEVEL\_OUTC Specification**

Parameter	Field Name	Description
Cohort Name	exposure_group_name	<p><b>Details:</b> name of the cohort of interest defined as either a singular exposure of interest or a grouping of same-day exposures. Defining unique <i>exposure_group_name</i> values allows for the specification of multiple cohorts in a single execution of the program package.</p> <p><b>Note 1:</b> The ordering of exposures in <i>exposure_group_name</i> 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</p>

		<p>exposures are separated by underscores but concatenated in a single name. The first listed exposure of interest is the <i>primary</i> exposure of interest, which is explained below. For example, if the <i>exposure_group_name</i> = TDaP_HP4, then the primary exposure of interest is TDaP, but there is also an interest in collecting data on same-day exposures to HPV4.</p> <p><b>Format:</b> Char (30)  <b>Example:</b> <i>exposure_group_name</i> = TDaP_HP4</p>
Cohort Identification Number	<i>exposure_group</i>	<p><b>Details:</b> <i>exposure_group</i> is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the <i>exposure_group_name</i> and <i>exposure_group</i>. For example, if there are two values for <i>exposure_group_name</i>, then there will be two corresponding values for <i>exposure_group</i>.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>exposure_group</i>=1</p>
Incidence Level for the tree	<i>incident_level</i>	<p><b>Details:</b> the level of the tree that defines an incident HOI. This variable designates that a diagnosis may be incident if there are no diagnoses that share the same node at the <i>Incident_level</i> of the tree in the HOI washout period (<i>AE_wash_up</i>).</p> <p>Allowable values for an <i>N</i>-level tree:</p> <ul style="list-style-type: none"> <li>• <b>1:</b> Tree Level 1</li> <li>• <b>2:</b> Tree Level 2</li> <li>• ...</li> <li>• <b>N:</b> Tree Level <i>N</i></li> </ul> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>Incident_level</i>=3</p>
Patients with Exposed Adverse Events Count	<i>exp_diag_pnt_cnt</i>	<p><b>Details:</b> Counts of unique patients having an incident adverse event in the observation window following exposure.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>exp_diag_pnt_cnt</i> =8</p>

Exposed Adverse Events Count	exp_diag_event_cnt	<p><b>Details:</b> Counts of unique incident adverse events in the observation window following exposure.</p> <p><b>Format:</b> Num (8)</p> <p><b>Example:</b> <i>exp_diag_event_cnt</i> =2.5</p>
------------------------------	--------------------	---

### 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 18: ORIG\_DX\_WKDAY Specification**

Parameter	Field Name	Description
Cohort Identification Number	exposure_group	<p><b>Details:</b> <i>exposure_group</i> is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the <i>exposure_group_name</i> and <i>exposure_group</i>. For example, if there are two values for <i>exposure_group_name</i>, then there will be two corresponding values for <i>exposure_group</i>.</p> <p><b>Format:</b> Num (8)</p> <p><b>Example:</b> <i>exposure_group</i>=1</p>
Incidence Level for the tree	incident_level	<p><b>Details:</b> the level of the tree that defines an incident HOI. This variable designates that a diagnosis may be incident if there are no diagnoses that share the same node at the <i>Incident_level</i> of the tree in the HOI washout period (<i>AE_wash_up</i>).</p> <p>Allowable values for an <i>N</i>-level tree:</p> <ul style="list-style-type: none"> <li>• <b>1:</b> Tree Level 1</li> <li>• <b>2:</b> Tree Level 2</li> <li>• ...</li> <li>• <b>N:</b> Tree Level <i>N</i></li> </ul> <p><b>Format:</b> Num (8)</p> <p><b>Example:</b> <i>Incident_level</i>=3</p>
Exposure Combination	exp_comb	<p><b>Details:</b> concatenation of binary (0/1) indicators for presence or absence of exposure of interest that corresponds to the</p>

Parameter	Field Name	Description
Identification Number		<p><i>exposure_group_name</i>. For example, if the <i>exposure_group_name</i> is TDaP_HP4 and the result pertains to the ascertainment of a TDaP vaccination without a same-day concomitant HPV4 vaccination, then the <i>exp_comb</i>=10. The number of characters is dictated by the <i>exposure_group_name</i> that has the most same-day exposure groupings included. For example, if there is a <i>exposure_group_name</i> with 3 exposures, then the <i>exp_comb</i> file will have 3 characters.</p> <p><b>Note 1:</b> This variable will not be included when linking with CIDA PSM results.</p> <p><b>Note 2:</b> ‘.’ characters indicate that the program is not recording any information on other exposures. For example, if the <i>exposure_group_name</i> is TDaP_HP4 but another <i>exposure_group_name</i> in the execution of the program has 3 exposures, then the potential <i>exp_comb</i> values for TDaP_HP4 will be “10.” or “11.”</p> <p><b>Format:</b> Char (30)  <b>Example:</b> <i>exp_comb</i> = 101</p>
Original Incident Health Outcome of Interest	orig_dx	<p><b>Details:</b> original health outcome of interest. It will be a string.</p> <p><b>Format:</b> Char (18)  <b>Example:</b> <i>orig_dx</i> = 780.2 (if ICD-9-CM)  <i>orig_dx</i>=R56.0 (if ICD-10-CM)</p>
Day of the Week of the Original Health Outcome of Interest	dx_wkday	<p><b>Details:</b> stratification of all eligible health outcomes of interest by the day of the week of their occurrence.</p> <p><b>Note1:</b> 1=Sunday</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>dx_wkday</i> =3</p>
Eligible Outcomes Count	COUNT	<p><b>Details:</b> Counts of original health outcomes of interest that fall on a particular weekday.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>COUNT</i> =8</p>

Parameter	Field Name	Description
Eligible Outcomes Percentage	PERCENT	<p><b>Details:</b> Percentage of original health outcomes of interest that fall on a particular weekday.</p> <p><b>Note1:</b> Percentages are calculated within <i>exposure_group</i> such that all percentages associated with a particular <i>exposure_group</i> add up to 100.</p> <p><b>Note2:</b> Valid values will be between 0 and 100.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> PERCENT =2.5</p>

#### 6. Ascertained Health Outcome of Interest Day of the Week 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 19: DX\_WKDAY Specification**

Parameter	Field Name	Description
Cohort Identification Number	exposure_group	<p><b>Details:</b> <i>exposure_group</i> is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the <i>exposure_group_name</i> and <i>exposure_group</i>. For example, if there are two values for <i>exposure_group_name</i>, then there will be two corresponding values for <i>exposure_group</i>.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>exposure_group</i>=1</p>
Incidence Level for the tree	incident_level	<p><b>Details:</b> the level of the tree that defines an incident HOI. This variable designates that a diagnosis may be incident if there are no diagnoses that share the same node at the <i>Incident_level</i> of the tree in the HOI washout period (<i>AE_wash_up</i>).</p> <p>Allowable values for an N-level tree:</p> <ul style="list-style-type: none"> <li>• <b>1:</b> Tree Level 1</li> <li>• <b>2:</b> Tree Level 2</li> </ul>

Parameter	Field Name	Description
		<ul style="list-style-type: none"> <li>• ...</li> <li>• <i>N</i>: Tree Level <i>N</i></li> </ul> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>Incident_level=3</i></p>
Exposure Combination Identification Number	exp_comb	<p><b>Details:</b> concatenation of binary (0/1) indicators for presence or absence of exposure of interest that corresponds to the <i>exposure_group_name</i>. For example, if the <i>exposure_group_name</i> is TDaP_HP4 and the result pertains to the ascertainment of a TDaP vaccination without a same-day concomitant HPV4 vaccination, then the <i>exp_comb</i>=10. The number of characters is dictated by the <i>exposure_group_name</i> that has the most same-day exposure groupings included. For example, if there is a <i>exposure_group_name</i> with 3 exposures, then the <i>exp_comb</i> file will have 3 characters.</p> <p><b>Note 1:</b> This variable will not be included when linking with CIDA PSM results.</p> <p><b>Note 2:</b> '.' characters indicate that the program is not recording any information on other exposures. For example, if the <i>exposure_group_name</i> is TDaP_HP4 but another <i>exposure_group_name</i> in the execution of the program has 3 exposures, then the potential <i>exp_comb</i> values for TDaP_HP4 will be "10." or "11."</p> <p><b>Format:</b> Char (30)  <b>Example:</b> <i>exp_comb = 101</i></p>
Ascertained Incident Health Outcome of Interest	dx	<p><b>Details:</b> ascertained incident health outcome of interest that is included in the analytic dataset. It will be a string, and decimals will not be compressed.</p> <p><b>Format:</b> Char (18)  <b>Example:</b> <i>dx = 780.2</i></p>
Day of the Week of Ascertained Health Outcome of Interest	dx_wkday	<p><b>Details:</b> stratification of all ascertained health outcomes of interest by the day of the week of their occurrence.</p>

Parameter	Field Name	Description
		<b>Note1:</b> 1=Sunday <b>Format:</b> Num (8) <b>Example:</b> <i>dx_wkday</i> =3
Eligible Outcomes Count	COUNT	<b>Details:</b> Counts of ascertained health outcomes of interest that fall on a particular weekday. <b>Format:</b> Num (8) <b>Example:</b> <i>COUNT</i> =8
Eligible Outcomes Percentage	PERCENT	<b>Details:</b> Percentage of ascertained health outcomes of interest that fall on a particular weekday. <b>Note1:</b> Percentages are calculated within <i>exposure_group</i> such that all percentages associated with a particular <i>exposure_group</i> add up to 100. <b>Note2:</b> Valid values will be between 0 and 100. <b>Format:</b> Num (8) <b>Example:</b> <i>PERCENT</i> =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 20: EXP\_WKDAY Specification**

Parameter	Field Name	Description
Cohort Identification Number	<i>exposure_group</i>	<b>Details:</b> <i>exposure_group</i> is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the <i>exposure_group_name</i> and <i>exposure_group</i> . For example, if there are two values for <i>exposure_group_name</i> , then there will be two corresponding values for <i>exposure_group</i> . <b>Format:</b> Num (8) <b>Example:</b> <i>exposure_group</i> =1

<p>Exposure Combination Identification Number</p>	<p>exp_comb</p>	<p><b>Details:</b> concatenation of binary (0/1) indicators for presence or absence of exposure of interest that corresponds to the <i>exposure_group_name</i>. For example, if the <i>exposure_group_name</i> is TDaP_HP4 and the result pertains to the ascertainment of a TDaP vaccination without a same-day concomitant HPV4 vaccination, then the <i>exp_comb</i>=10. The number of characters is dictated by the <i>exposure_group_name</i> that has the most same-day exposure groupings included. For example, if there is a <i>exposure_group_name</i> with 3 exposures, then the <i>exp_comb</i> file will have 3 characters.</p> <p><b>Note 1:</b> This variable will not be included when linking with CIDA PSM results.</p> <p><b>Note 2:</b> '.' characters indicate that the program is not recording any information on other exposures. For example, if the <i>exposure_group_name</i> is TDaP_HP4 but another <i>exposure_group_name</i> in the execution of the program has 3 exposures, then the potential <i>exp_comb values</i> for TDaP_HP4 will be "10." or "11."</p> <p><b>Format:</b> Char (30)  <b>Example:</b> <i>exp_comb</i> = 101</p>
<p>Day of the Week of Eligible Exposure</p>	<p>exp_wkday</p>	<p><b>Details:</b> stratification of all eligible exposures by the day of the week of their administration.</p> <p><b>Note1:</b> 1=Sunday</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>exp_wkday</i> =3</p>
<p>Eligible Exposures Count</p>	<p>COUNT</p>	<p><b>Details:</b> Counts of eligible exposures that fall on a particular day of the week.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>COUNT</i> =8</p>
<p>Eligible Exposures Percentage</p>	<p>PERCENT</p>	<p><b>Details:</b> Percentage of eligible exposures that fall on a particular day of the week</p>



		<p><b>Note1:</b> Percentages are calculated within <i>exposure_group</i> such that all percentages associated with a particular <i>exposure_group</i> add up to 100.</p> <p><b>Note2:</b> Valid values will be between 0 and 100.</p> <p><b>Format:</b> Num (8)</p> <p><b>Example:</b> <i>PERCENT</i> =2.5</p>
--	--	--

### 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 file will not be included in the output when linking with CIDA PSM results.

This output file should include a row for each unique cohort (*exposure\_group*).

**Table 21: EXPOSURE\_GROUP\_PROCESS\_FLOW Specification**

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

Parameter	Field Name	Description
		<p><b>Note 1:</b> The ordering of exposures in <i>exposure_group_name</i> 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 <i>primary</i> exposure of interest, which is explained below. For example, if the <i>exposure_group_name</i> = TDaP_HP4, then the primary exposure of interest is TDaP, but there is also an interest in collecting data on same-day exposures to HPV4.</p> <p><b>Format:</b> Char (30)  <b>Example:</b> <i>exposure_group_name</i> = TDaP_HP4</p>
Count of Exposed Patients	exp_pnt_cnt	<p><b>Details:</b> Counts of unique patients having at least one exposure.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>exp_pnt_cnt</i> =8</p>
Count of Exposures	exp_event_cnt	<p><b>Details:</b> Counts of unique exposures.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>exp_event_cnt</i> =8</p>
Count of Patients with Primary Exposure	prim_exp_pnt_cnt	<p><b>Details:</b> Counts of unique patients having at least one <i>primary</i> exposure.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>prim_exp_pnt_cnt</i> =8</p>
Count of Primary Exposures	prim_exp_event_cnt	<p><b>Details:</b> Counts of unique <i>primary</i> exposures.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>prim_exp_event_cnt</i> =8</p>
Count of Patients with Incident Exposures	inc_exp_pnt_cnt	<p><b>Details:</b> Counts of unique patients having at least one <i>incident</i> primary exposure.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>COUNT</i> =8</p>
Count of Incident Exposures	inc_exp_event_cnt	<p><b>Details:</b> Counts of unique <i>incident</i> primary exposures.</p> <p><b>Format:</b> Num (8)</p>

Parameter	Field Name	Description
		<b>Example:</b> <code>inc_exp_event_cnt =8</code>
Count of Patients having Exposures with Eligible Drug Enrollment and Demographics	<code>drug_enr_demo_elig_pnt_cnt</code>	<p><b>Details:</b> Counts of unique patients having at least one incident primary exposure that meets <i>drug enrollment and demographic eligibility</i> criteria.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <code>drug_enr_demo_elig_pnt_cnt =8</code></p>
Count of Exposures with Eligible Drug Enrollment and Demographics	<code>drug_enr_demo_exp_event_elig</code>	<p><b>Details:</b> Counts of unique incident primary exposures that meet <i>drug enrollment and demographic eligibility</i> criteria.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <code>drug_enr_demo_exp_event_elig =8</code></p>
Count of Patients having Exposures with Eligible Medical Enrollment	<code>med_enr_demo_elig_pnt_cnt</code>	<p><b>Details:</b> Counts of unique patients having at least one eligible exposure that also meets <i>medical enrollment eligibility</i> and minimum post-exposure enrollment criteria.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <code>med_enr_demo_elig_pnt_cnt =8</code></p>
Count of Exposures with Eligible Medical Enrollment	<code>med_enr_demo_exp_event_elig</code>	<p><b>Details:</b> Counts of unique eligible exposures that also meet <i>medical enrollment eligibility</i> and minimum post-exposure enrollment criteria.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <code>med_enr_demo_exp_event_elig =8</code></p>
Count of Patients having Eligible Exposures after Exclusions	<code>elig_pnt_cnt_after_excl</code>	<p><b>Details:</b> Counts of unique patients having at least one fully-eligible exposure <i>after applying exclusions</i>.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <code>elig_pnt_cnt_after_excl =8</code></p>
Count of Eligible Exposures after Exclusions	<code>exp_event_elig_after_excl</code>	<p><b>Details:</b> Counts of unique fully-eligible exposures <i>after applying exclusions</i>.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <code>exp_event_elig_after_excl =8</code></p>

## 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 22: CRSWLK\_FOR\_FREEZE Specification**

Parameter	Field Name	Description
Cohort Name	exposure_group	<p><b>Details:</b> <i>exposure_group</i> is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the <i>exposure_group_name</i> and <i>exposure_group</i>. For example, if there are two values for <i>exposure_group_name</i>, then there will be two corresponding values for <i>exposure_group</i>.</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>exposure_group</i>=1</p>
Incidence Level for the tree	incident_level	<p><b>Details:</b> the level of the tree that defines an incident HOI. This variable designates that a diagnosis may be incident if there are no diagnoses that share the same node at the <i>Incident_level</i> of the tree in the HOI washout period (<i>AE_wash_up</i>).</p> <p>Allowable values for an <i>N</i>-level tree:</p> <ul style="list-style-type: none"> <li>• <b>1:</b> Tree Level 1</li> <li>• <b>2:</b> Tree Level 2</li> <li>• ...</li> <li>• <b>N:</b> Tree Level <i>N</i></li> </ul> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>Incident_level</i>=3</p>
Exposure Combination Identification Number	exp_comb	<p><b>Details:</b> concatenation of binary (0/1) indicators for presence or absence of exposure of interest that corresponds to the <i>exposure_group_name</i>. For example, if the <i>exposure_group_name</i> is TDaP_HP4 and the result pertains to the ascertainment of a TDaP vaccination without a same-day concomitant</p>

Parameter	Field Name	Description
		<p>HPV4 vaccination, then the <i>exp_comb</i>=10. The number of characters is dictated by the <i>exposure_group_name</i> that has the most same-day exposure groupings included. For example, if there is a <i>exposure_group_name</i> with 3 exposures, then the <i>exp_comb</i> file will have 3 characters.</p> <p><b>Note1:</b> ‘.’ characters indicate that the program is not recording any information on other exposures. For example, if the <i>exposure_group_name</i> is Tdap_HP4 but another <i>exposure_group_name</i> in the execution of the program has 3 exposures, then the potential <i>exp_comb values</i> for Tdap_HP4 will be “10.” or “11.”</p> <p><b>Format:</b> Char (30)  <b>Example:</b> <i>exp_comb</i> = 101</p>
Unique Patient Identifier	PATID	<p><b>Details:</b> unique patient identifier.</p> <p><b>Format:</b> Char (30)  <b>Example:</b> <i>PATID</i> = APEIufyq39845</p>
Patient age at exposure date	age_at_exp_dt	<p><b>Details:</b> patient age at exposure date</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>age_at_exp_dt</i> = 30</p>
Incident node	node	<p><b>Details:</b> the incident node that contains the ascertained health outcome of interest</p> <p><b>Format:</b> Char (30)  <b>Example:</b> <i>node</i> = 10.01.08</p>
Original Incident Health Outcome of Interest	orig_dx	<p><b>Details:</b> 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.</p> <p><b>Note1:</b> If the <i>orig_dx</i> happens to be an ICD-9-CM code, then no conversion is necessary.</p> <p><b>Format:</b> Char (18)  <b>Example:</b> <i>orig_dx</i> = 780.2 (if ICD-9-CM)  <i>orig_dx</i>=R56.0 (if ICD-10-CM)</p>

Parameter	Field Name	Description
Original Diagnosis Code Type	orig_dx_codetype	<p><b>Details:</b> Code type required by lookup file.</p> <p>Valid values include:</p> <ul style="list-style-type: none"> <li>• <b>09:</b> ICD-9-CM</li> <li>• <b>10:</b> ICD-10-CM</li> </ul> <p><b>Format:</b> Char (2)  <b>Example:</b> <i>orig_dx_codetype=09</i></p>
Ascertained Incident Health Outcome of Interest	dx	<p><b>Details:</b> incident health outcome of interest that is included in the analytic dataset. It will be a string.</p> <p><b>Note1:</b> If the <i>orig_dx</i> 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).</p> <p><b>Format:</b> Char (18)  <b>Example:</b> <i>dx = 780.2</i></p>
Ascertained Diagnosis Code Type	dx_codetype	<p><b>Details:</b> code type that is used to populate the Diagnosis Tree (CHILD_PARENT.sas7bdat).</p> <p>Valid values include:</p> <ul style="list-style-type: none"> <li>• <b>09:</b> ICD-9-CM</li> </ul> <p><b>Format:</b> Char (2)  <b>Example:</b> <i>dx_codetype=09</i></p>
Date for Incident Health Outcome of Interest	dx_adata	<p><b>Details:</b> date of the qualifying incident health outcome of interest.</p> <p><b>Format:</b> SAS date (date9.)  <b>Example:</b> <i>dx_adata=08/15/2004</i></p>
Date for Exposure of Interest	exp_adata	<p><b>Details:</b> date for the exposure of interest.</p> <p><b>Format:</b> SAS date (date9.)  <b>Example:</b> <i>exp_adata=08/15/2004</i></p>
Time to censor	time_to_censor	<p><b>Details:</b> number of days after exposure that the patient becomes ineligible due to break in enrollment, patient death, or reaching the end of the study period.</p> <p><b>Note 1:</b> Default value is 9999 if the patient is not censored within the minimum required follow-up window.</p>

Parameter	Field Name	Description
		<b>Format:</b> Num (8) <b>Example:</b> <i>time_to_censor=44</i>

## 2. PSM Analytic Dataset Crosswalk

The PSM Analytic Dataset Crosswalk file (CRSWLK\_FOR\_FREEZE\_[MATCHID].sas7bdat), used in place of the Analytic Dataset Crosswalk file when linking with CIDA PSM results, 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.

The *MATCHID* value in the dataset name depends on the CIDA PSM analysis used. Valid values for TreeScan analysis include:

- **PREDEFINEDPSMATCHID1\_1:** predefined covariate estimation model; 1:1 matching strategy
- **HDPSPREDEFINEDMATCHID1\_1:** predefined covariate/ empirically identified covariate hybrid estimation model; 1:1 matching strategy
- **HDPONLYMATCHID1\_1:** empirically identified covariate estimation model; 1:1 matching strategy

**Table 23: CRSWLK\_FOR\_FREEZE [MATCHID] Specification**

Parameter	Field Name	Description
Cohort Name	exposure_group	<b>Details:</b> <i>exposure_group</i> is a numeric identification number used by all the input files as a key. There is a 1:1 correspondence between the <i>exposure_group_name</i> and <i>exposure_group</i> . For example, if there are two values for <i>exposure_group_name</i> , then there will be two corresponding values for <i>exposure_group</i> .  <b>Format:</b> Num (8) <b>Example:</b> <i>exposure_group=1</i>
Incidence Level for the tree	incident_level	<b>Details:</b> the level of the tree that defines an incident HOI. This variable designates that a diagnosis may be incident if there are no diagnoses that share the same node at the <i>Incident_level</i> of the tree in the HOI washout period ( <i>AE_wash_up</i> ).  Allowable values for an <i>N</i> -level tree: <ul style="list-style-type: none"> <li>• <b>1:</b> Tree Level 1</li> <li>• <b>2:</b> Tree Level 2</li> <li>• ...</li> <li>• <b>N:</b> Tree Level <i>N</i></li> </ul>

		<p><b>Format:</b> Num (8)  <b>Example:</b> <i>Incident_level=3</i></p>
Unique Patient Identifier	PATID	<p><b>Details:</b> unique patient identifier.</p> <p><b>Format:</b> Char (30)  <b>Example:</b> <i>PATID = APElufyq39845</i></p>
Match ID value	[MATCHID]	<p><b>Details:</b> MatchID value for the analysis requested.</p> <p><b>Note1:</b> Name of the variable is based on the CIDA PSM analysis requested. Allowable values include:</p> <ul style="list-style-type: none"> <li>• <b>PREDEFINEDPSMATCHID1_1:</b> predefined covariate estimation model; 1:1 matching strategy</li> <li>• <b>HDPSPREDEFINEDMATCHID1_1:</b> predefined covariate/ empirically identified covariate hybrid estimation model; 1:1 matching strategy</li> <li>• <b>HDPSONLYMATCHID1_1:</b> empirically identified covariate estimation model; 1:1 matching strategy</li> </ul> <p><b>Format:</b> Char (50)  <b>Example:</b> <i>PREDEFINEDPSMATCHID1_1=1467</i></p>
Patient age at exposure date	age_at_exp_dt	<p><b>Details:</b> patient age at exposure date</p> <p><b>Format:</b> Num (8)  <b>Example:</b> <i>age_at_exp_dt = 30</i></p>
Patient age grouping	age_group	<p><b>Details:</b> patient age grouping</p> <p><b>Format:</b> Char (30)  <b>Example:</b> <i>age_group = 21-44</i></p>
Incident node	node	<p><b>Details:</b> the incident node that contains the ascertained health outcome of interest</p> <p><b>Format:</b> Char (30)  <b>Example:</b> <i>node = 10.01.08</i></p>
Original Incident Health Outcome of Interest	orig_dx	<p><b>Details:</b> 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</p>



		<p>included in the analytic dataset. It will be a string.</p> <p><b>Note1:</b> If the <i>orig_dx</i> happens to be an ICD-9-CM code, then no conversion is necessary.</p> <p><b>Format:</b> Char (18)  <b>Example:</b> <i>orig_dx</i> = 780.2 (if ICD-9-CM)  <i>orig_dx</i>=R56.0 (if ICD-10-CM)</p>
Original Diagnosis Code Type	orig_dx_codetype	<p><b>Details:</b> Code type required by lookup file.</p> <p>Valid values include:</p> <ul style="list-style-type: none"> <li>• <b>09:</b> ICD-9-CM</li> <li>• <b>10:</b> ICD-10-CM</li> </ul> <p><b>Format:</b> Char (2)  <b>Example:</b> <i>orig_dx_codetype</i>=09</p>
Ascertained Incident Health Outcome of Interest	dx	<p><b>Details:</b> incident health outcome of interest that is included in the analytic dataset. It will be a string.</p> <p><b>Note1:</b> If the <i>orig_dx</i> 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).</p> <p><b>Format:</b> Char (18)  <b>Example:</b> <i>dx</i> = 780.2</p>
Ascertained Diagnosis Code Type	dx_codetype	<p><b>Details:</b> code type that is used to populate the Diagnosis Tree (CHILD_PARENT.sas7bdat).</p> <p>Valid values include:</p> <ul style="list-style-type: none"> <li>• <b>09:</b> ICD-9-CM</li> </ul> <p><b>Format:</b> Char (2)  <b>Example:</b> <i>dx_codetype</i>=09</p>
Date for Incident Health Outcome of Interest	dx_adata	<p><b>Details:</b> date of the qualifying incident health outcome of interest.</p> <p><b>Format:</b> SAS date (date9.)  <b>Example:</b> <i>dx_adata</i>=08/15/2004</p>
Date for Exposure of Interest	exp_adata	<p><b>Details:</b> date for the exposure of interest.</p> <p><b>Format:</b> SAS date (date9.)  <b>Example:</b> <i>exp_adata</i>=08/15/2004</p>

Time to censor	time_to_censor	<p><b>Details:</b> number of days after exposure that the patient becomes ineligible due to break in enrollment, patient death, or reaching the end of the study period.</p> <p><b>Note1:</b> Default value is 9999 if the patient is not censored within the minimum required follow-up window.</p> <p><b>Format:</b> Num (8) <b>Example:</b> <i>time_to_censor=44</i></p>
Enrollment end date	enr_end_dt	<p><b>Details:</b> end date of enrollment span where index exposure was identified.</p> <p><b>Format:</b> SAS date (date9.) <b>Example:</b> <i>enr_end_dt =44</i></p>

### 3. 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. If *priority\_file=0* in the General Parameters file, then first create the DXTREE\_PRIORITY file from the available data.
5. Combine ICD-9-CM diagnosis tree codes with ICD-10-CM diagnosis tree codes according to ID10-ID9\_MAPPING.sas7bdat file.
6. Convert vertical Child-Parent Diagnosis Tree file into horizontal structure.
7. Gather relevant data from SCDM tables.
  - a. 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.
  - b. From these inputs, the TreeExtraction program subsets the SCDM utilization tables to the data that are necessary for the entire subsequent TreeExtraction execution.
8. Gather all records with relevant exposure codes as determined by prior step.
9. Delete patients according to the ASO\_EXCL\_FLAG input file if *ASO\_excl\_flag=1*.
10. Create table based on exposure records linked to *exposure\_group* and exclusion *subgroup* using EXCLUSION\_GROUP.sas7bdat parameters.
11. Link to Sentinel Common Data Model Death table, and exclude exposures having death date < exposure date.

12. Subset pool of patients to those that have *primary* exposure event per *exposure\_group* and attach *exposure\_group* parameters to each record.
  - a. Given that this is a SCRI design, only members with the exposure of interest will be included in the study population.
  - b. The exposure of interest will be identified using a combination of diagnosis, procedure, and dispensing codes.
  - c. This program can only identify an exposure using “OR” Boolean logic. It does not have the capability to identify exposures with “AND” logic.
13. Limit *primary* exposure of interest to only incident exposures as specified by *exp\_wash\_up*.
14. Add enrollment and demographics information for the patients previously identified.
15. Create enrollment episodes for each potential member that could be included in the cohort. Bridge gaps less than or equal to *enr\_gap* days.
16. Make birth adjustments to enrollment criteria.
17. Determine that member meets enrollment criteria for study population.
  - a. For member with enrollment within the study period dates, include only exposures having post-exposure enrollment greater than or equal to the value of the *post\_exp\_min\_enr* parameter.
  - b. For members with enrollment within the study period dates, members will have a required pre-exposure enrollment period.
    - i. 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*).
18. Determine eligible follow-up time for the patient, measured as the minimum of:
  - i. the number of days between exposure and patient death;
  - ii. the number of days between exposure and end of required continuous enrollment type;
  - iii. the number of days between exposure and end of study period.
  - b. Assign this value to a variable called *time\_to\_censor* for inclusion in appropriate output datasets.
19. Determine that patient 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.
20. Exclude records with prior exposures based on exclusion criteria. Add non-primary exposures to the selected *primary* exposures for each *exposure\_group*.
21. If *PS\_flag*=1 in the General Parameters file, then link to the Propensity Score Master and Match files held in the CIDA library.
  - a. These results will be used in place of the cohort and exposure information described above.
22. Calculate descriptive statistics for each *exposure\_group*.
  - i. 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.
23. Find eligible incident health outcomes of interest.
- a. 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.
  - b. 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.
  - c. 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 not have occurred in the *AE\_wash\_up* days at the *Incident\_level* from the CHILD\_PARENT.sas7bdat file with the specified *EncType* and *PDX* values using the WASH\_UP\_ENC\_SETTINGS.sas7bdat file.
  - d. Break any ties for same day incident diagnoses that occur using the *priority* parameter.
24. Create EXPOSURE\_GROUP\_LEVEL\_OUTC.sas7bdat file with patient and event counts for exposure/AE pairs by *exposure\_group* and specified *incident\_level*.
25. Calculate 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.
26. 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
27. 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.