

# MINI-SENTINEL PROSPECTIVE ROUTINE OBSERVATIONAL MONITORING PROGRAM TOOL: CONTINUOUS AND GROUP SEQUENTIAL ANALYSIS WITH SELF-CONTROL DESIGN

**Technical Users' Guide version: 1.0**

**Prepared by:** Claudia Coronel-Moreno, MPH,<sup>1</sup> Yury Vilks, PhD,<sup>1</sup> Ivair Silva, PhD,<sup>1</sup> Katherine Yih, PhD, MPH,<sup>1</sup> Martin Kulldorff, PhD<sup>1</sup>

**Author Affiliations:** 1. Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA

**May 19, 2014**

Mini-Sentinel is a pilot project sponsored by the [U.S. Food and Drug Administration \(FDA\)](#) to inform and facilitate development of a fully operational active surveillance system, the Sentinel System, for monitoring the safety of FDA-regulated medical products. Mini-Sentinel is one piece of the [Sentinel Initiative](#), a multi-faceted effort by the FDA to develop a national electronic system that will complement existing methods of safety surveillance. Mini-Sentinel Collaborators include Data and Academic Partners that provide access to health care data and ongoing scientific, technical, methodological, and organizational expertise. The Mini-Sentinel Coordinating Center is funded by the FDA through the Department of Health and Human Services (HHS) Contract number HHSF223200910006I.

# Mini-Sentinel Prospective Routine Observational Monitoring Program Tool: Continuous and Group Sequential Analysis with Self-Control Design

## Technical Users' Guide version: 1.0

### Table of Contents

<b>I. OVERVIEW .....</b>	<b>1</b>
<b>II. DESIGN PARAMETER SETTINGS .....</b>	<b>2</b>
A. DATA EXTRACTION.....	2
B. DATA ANALYSIS.....	5
C. STATISTICAL POWER AND EXPECTED TIME TO SIGNAL .....	5
<b>III. DATA EXTRACTION (SAS CODE) .....</b>	<b>5</b>
A. FOR INVESTIGATORS.....	5
B. FOR MINI-SENTINEL OPERATIONS CENTER.....	12
1. <i>Documentation</i> .....	12
2. <i>Program Steps</i> .....	21
C. FOR DATA PARTNERS .....	22
<b>IV. SEQUENTIAL ANALYSIS (R FUNCTIONS).....</b>	<b>23</b>
A. DOWNLOADING R AND LOADING FUNCTIONS.....	24
B. RUNNING FUNCTIONS .....	29
1. <i>Function "AnalyzeSetUp.Binomial"</i> .....	30
2. <i>Function "Analyze.Binomial"</i> .....	31
C. R OUTPUT.....	33
D. EXAMPLE TESTS.....	35
1. <i>Analyze.Binomial(name="TestFile",test=1,cases&lt;- 5,controls&lt;- 1)</i> .....	35
2. <i>Analyze.Binomial(name="TestFile",test=2,cases&lt;- 3,controls&lt;- 0)</i> .....	37
3. <i>Analyze.Binomial(name="TestFile",test=3,cases&lt;- 7,controls&lt;- 1)</i> .....	39
E. EXTENDED SURVEILLANCE .....	40
F. REPORT FORMAT.....	41
<b>V. APPENDIX 1.....</b>	<b>45</b>
A. MINI-SENTINEL WORKPLAN.....	45
1. <i>Project Name</i> .....	45
2. <i>Overall Project Objective</i> .....	45
3. <i>MSOC Contact Information</i> .....	45
4. <i>Budget Item</i> .....	45
5. <i>Instructions</i> .....	45
6. <i>For KP Mini-Sentinel Programmers Only</i> .....	46
7. <i>Workplan Timeline</i> .....	46
8. <i>Package Documents</i> .....	46

## I. OVERVIEW

When a new vaccine or pharmaceutical drug enters the market, there are always questions about its safety profile. Post-marketing drug and vaccine safety surveillance is important in order to detect serious adverse events that are too rare to be detected during Phase III clinical trials. Such surveillance has traditionally been based on spontaneous adverse event reporting systems, but electronic health records from health insurance plans are increasingly being used instead.

If there is a major safety problem, prompt detection is critical. Electronic health data can now be accessed on a quarterly, monthly, or even weekly basis to prospectively monitor the safety of a newly approved drug or vaccine, using gradually accumulating data. This repeated monitoring requires use of sequential statistical analysis, which adjusts for the multiple testing inherent in the many looks at the data.

Self-control methods are a powerful and often used design for drug and vaccine safety studies when looking for acute adverse events that occur soon after initial exposure. Through the self-control design, all non-time varying confounders are automatically adjusted for, including gender, insurance coverage, education, income, geography, etc. The method uses a pre-defined risk window, such as 1-21 days after initial exposure, and a pre-defined control window of the same or different length. Depending on the drug/vaccine-event pair under study, the control window may be either before exposure or after the risk window. Technically, it is also possible to have a control window immediately after the initial exposure if an increased risk is implausible during that period, and a risk window after the control window, but this would not be a common study design. In the current version, the ratio of the risk to control window must be the same for all patients. Future versions of the system will allow for different ratios for different patients.

The self-control method does not adjust for time-varying confounders unless explicitly modified to do so. Hence, it should be used with caution for very young children, whose baseline risk of certain health outcomes can change quickly within small increments of age, such as at 3 versus 5 weeks of age. It is then important to adjust for age using an offset term. The self-control design should also be used with caution if there is seasonality in both drug/vaccine initiation and the outcome under study, in which case seasonality must be adjusted for. If season influences only one of these two, i.e. either the drug/vaccine exposure or the outcome, the method will still be unbiased, and no adjustment is needed. Another potential source of bias is if there is something that triggers both the exposure and the outcome. For example, patients are often given the Pneumococcal Polysaccharide Vaccine (PPSV) vaccine just before an organ transplant, and organ transplants may cause adverse events. Hence, in a self-control analysis, there may be more adverse events just after PPSV vaccination even though the vaccine is just an innocent by-stander. If occurrence of the outcome influences or could influence whether the drug/vaccine is given to the patient, it is important to use a control window after the risk window rather than a pre-exposure control window in order to avoid bias by indication or contra-indication.

This manual describes how self-control sequential analysis can be conducted within the Mini-Sentinel PROMPT project, utilizing SAS code and R functions that we have developed. The SAS code is written exclusively for the sequential self-control design while the R functions are written both for the self-control design as well as for the propensity score matched concurrent controls.

This is not a guide or tutorial on the use of sequential analysis, nor does it provide details about specific sequential statistical analysis or self-control designs, for which we refer the reader to the papers listed in the reference section. After this introduction, the manual is divided into four parts: (i) study design parameter definitions, (ii) SAS code for data extraction, (iii) R functions for sequential analysis, and (iv) the template for reporting the sequential analysis results. Detailed step-by-step guidance and multiple sample screen shots are provided.

## II. DESIGN PARAMETER SETTINGS

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

### A. DATA EXTRACTION

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

**Table 1. Specified Features for Self-Controlled Design**

	LISINOPRIL	CLINDAMYCIN	MMRV-1	MMRV-2	NOTES
Minimum period of continuous enrollment with pharmacy and medical benefits required prior to dispensing date	183 days	183 days	183 days	All fields same as for MMRV-1 except for control window	An apparent gap in enrollment can be specified such that the gap is bridged (ignored).
Age range at exposure	≥18 yrs.	≥45 yrs.	12-23 mo.		

	LISINOPRIL	CLINDAMYCIN	MMRV-1	MMRV-2	NOTES
Exposure exclusions	ACEi's, ARBs, direct renin inhibitors	Antibiotics			Specify if patients with certain exposures on the same day as or in the x days prior to the exposure of interest are to be excluded.
Exposure look-back period from dispensing date*	0-183 days	0-183 days	0-183 days		
Health outcome (HOI)	Angioedema	AMI	Seizures		
HOI settings	All	Inpatient	Inpatient, ED (inpatient has higher priority within a day)	Inpatient, ED (inpatient has higher priority within a day)	Possible settings in which to look for potential cases include inpatient, ED, and clinic (outpatient).
HOI exclusions	Prior angioedema	Prior AMI	Prior seizures		
HOI exclusion settings	All	IP	All		For MMRV, cases of interest are those in inpatient and ED settings but only if there were no cases in the prior 183 days in <i>any</i> setting, including clinic.

	LISINOPRIL	CLINDAMYCIN	MMRV-1	MMRV-2	NOTES
HOI look-back period from diagnosis date**	1-183 days	1-28 days	1-183 days		The case must be the patient's first one to occur in this period of time in order to count.
Risk window	1-28 days	1-14 days	7-10 days		This is the post-exposure (Day 0) period during which an increased risk of the HOI is considered possible and will be examined.
Comparison window	-35 to -8	15 to 28	1 to 6 days	15 to 28	This is the period, either pre- or post-exposure (Day 0), during which the risk is considered to be at baseline.
Maximum period to collect to include both windows	-35 to +28 days	1 to 28 days	1 to 28 days	1 to 28 days	

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

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

## B. DATA ANALYSIS

Before analyzing the data, the user must specify several parameters:

**Table 2. Several Parameters to Consider Before Data Analysis**

PARAMETER	EXPLANATION	CONSIDERATION
alpha	Desired alpha level overall	0.05 is the default, if no other value is chosen
M	Minimum number of cases in risk + control intervals to signal	1 is the default, which means that the system will signal as soon as there are enough events in the risk interval to reject the null hypothesis. (It does not mean that it will signal after one event.)
N	Cumulative number of cases in risk + control intervals at which surveillance is to stop even if $H_0$ has not been rejected (sometimes called the “upper limit on the length of surveillance”)	No default; should be selected to ensure that there is enough statistical power to detect the lowest risk of concern for public health

## C. STATISTICAL POWER AND EXPECTED TIME TO SIGNAL

Outside of this project, R code has been developed to estimate: a) the statistical power to detect a range of relative risks; b) for different maximum lengths of surveillance; as well as c) the expected time to signal when the null hypothesis is rejected. The statistical power can also be looked up in reference tables available in the manuscripts by Kulldorff et al<sup>12</sup>

## III. DATA EXTRACTION (SAS CODE)

### A. FOR INVESTIGATORS

To implement the criteria and features of data to be extracted that were presented in Section B.1., one must first assign and define some codes, as in the following look-up table, LKP\_FOR\_ALL\_CODES. **Note about terminology:** In the look-up and mapping tables, the word “query” is related to the exposure of interest, and the word “event” is related to the health outcome of interest. Also, please note that “exposure” and “dispensing” are used somewhat interchangeably, and “health outcome,” “outcome,” and “HOI” are also used somewhat interchangeably, in this Section C.1.

<sup>1</sup> Kulldorff, M., Silva, I.R. (2012). Continuous Sequential Analysis with a Delayed Start. Manuscript submitted for publication.

<sup>2</sup> Silva, I.R., Kulldorff M. (2012). Continuous versus Group Sequential Analysis for Vaccine and Drug Safety Surveillance. Unpublished manuscript.

**Table 3. LKP\_FOR\_ALL\_CODES**

COLUMN_NAME	CODE_VALUE	CODE_VALUE_C	CODE_DESC
EVENT_GRP	1		LISINOPRIL
EVENT_GRP	2		CLINDAMYCIN
EVENT_GRP	3		MMRV-1
EVENT_GRP	4		MMRV-2
GROUP	1		LISINOPRIL
GROUP	2		CLINDAMYCIN
GROUP	3		MMRV-1
GROUP	4		MMRV-2
QUERY_GRP	1		LISINOPRIL
QUERY_GRP	2		CLINDAMYCIN
QUERY_GRP	3		MMRV
SUBGROUP	1		LISINOPRIL
SUBGROUP	2		CLINDAMYCIN
SUBGROUP	3		MMRV
SUBGROUP	11		ACEI'S, ARBS, DIRECT RENIN INHIBITORS
SUBGROUP	12		ANTIBIOTICS
SUBGROUP	101		ANGIOEDEMA
SUBGROUP	102		AMI
SUBGROUP	103		SEIZURES
ENCTYPE		IP	INPATIENT HOSPITAL STAYS
ENCTYPE		IS	NON-ACUTE INSTITUTIONAL STAYS
ENCTYPE		ED	EMERGENCY DEPARTMENT
ENCTYPE		AV	AMBULATORY VISIT
ENCTYPE		QA	OTHER AMBULATORY VISIT
PDX		P	PRINCIPAL DIAGNOSIS
PDX			NON-PRINCIPAL DIAGNOSIS

In the LKP\_FOR\_ALL\_CODES table above, there are three QUERY\_GRPs, corresponding to lisinopril, clindamycin, and MMRV exposures and the conditions that define those, and four EVENT\_GRPs, representing the health outcomes and the conditions that define those; there are four instead of three EVENT\_GRPs because the MMRV exposure is being studied using two different control intervals. These groups are linked to each other in the GROUPS\_MAPPING table:

**Table 4. GROUPS\_MAPPING**

GROUP	QUERY_GRP	EVENT_GRP
1	1	1
2	2	2
3	3	3
4	3	4

The minimum enrollment criteria are specified (with other criteria, to be discussed later) in the QUERY table. In these studies, 183 days of continuous enrollment prior to the exposure (“WASHPER”) are required for inclusion. If apparent gaps in enrollment are to be bridged (ignored), the macro parameter ENROLGAP should be filled out with the maximum gap to be bridged in terms of days, for example, “14.” The default value of ENROLGAP is 45.

**Table 5. QUERY**

QUERY_GRP	SUBGROUP	WASHTYP	WASHPER	EXCL_COMBO_MEDS_ON_SAME_DAY
1	1	MULT	183	1
2	2	MULT	183	1
3	3	MULT	183	1

The age range is specified in the AGE GROUPS table below, which also specifies and assigns codes to the age groups within the range. The age groups within the ranges are not needed for the self-controlled analysis but are used to characterize cases in the event that a signal appears that warrants further investigation.

**Table 6. AGE\_GROUPS**

QUERY_GRP	AGE_UNIT	AGE_MIN	AGE_MAX	AGE_GROUP
1	Y	18	44	4
1	Y	45	54	5
1	Y	55	64	6
1	Y	65	200	7
2	Y	45	54	5
2	Y	55	64	6
2	Y	65	74	7
2	Y	75	84	8
2	Y	85	200	9
3	M	12	15	1
3	M	16	23	2

The NDC codes identifying the exposures of interest (lisinopril, clindamycin, and MMRV) and the kinds of prior exposures for which exposures of interest are to be excluded (referred to as “exposure exclusions” in Section B.1.) are in the QUERY\_CODES table. This table is very long and is shown only schematically below, with just one row per SUBGROUP. The CODEs associated with SUBGROUPs 1, 2, and 3 are the

11-digit NDC codes for lisinopril, clindamycin, and MMRV products, respectively. SUBGROUPS 11 and 12 are discussed below the table in the context of other exposures for which an exposure of interest should be excluded.

**Table 7. QUERY\_CODES**

SUBGROUP	CODETYPE	CODE	QUERY_TBL	INCQUERY_1
1	Rx11	[multiple 11-digit numeric NDC codes]	1	0
2	Rx11	[ditto]	1	0
3	Rx11	[ditto]	1	0
11	Rx11	[ditto]	0	1
12	Rx11	[ditto]	0	1

The specifications for exposure exclusion criteria appear in multiple tables: the QUERY and QUERY\_CODES tables discussed above and the INCQUERY table. In the example studies, the minimum enrollment period is the same as the look-back period to check for prior exposures, thus, the QUERY table serves both purposes (if investigators wish to specify distinct minimum enrollment and look-back periods, an additional parameter will be needed for minimum enrollment). The QUERY\_CODES table contains not only the codes for the exposures (QUERY\_GRPs) being studied but also codes for types of exposures meant to lead to the exclusion of an exposure of interest. These two categories of product are distinguished by the QUERY\_TBL and INCQUERY\_1 flags, the first of which is set to 1 for the exposures being studied and the second of which is set to 1 for the exposures leading to exclusion.

There are two kinds of exposure exclusions: one is if the exposure of interest is not incident, the other is if the exposure of interest was preceded by exposure to some other product within a certain period of time. In our example studies, we defined incident exposure as the first dispensing in a 183-day period—this is specified by setting WASHPER = 183 in the QUERY table, which leads to exclusion of a dispensing occurring < 183 days after a previous one (for the same SUBGROUP). If exposure to some *other* product within a certain period of time is to lead to exclusion, one first creates a separate SUBGROUP of the product codes in the QUERY\_CODES table. For example, SUBGROUP 11 consists of ACEi's, ARBS, and direct renin inhibitors, all of which are to lead to exclusion from Study (GROUP) 1 if a dispensing of any of them appeared in the 183 days prior to the lisinopril exposure being considered; and SUBGROUP 12 consists of antibiotics, which are to lead to exclusion from Study (GROUP) 2 if a dispensing of any of them appeared in the 183 days prior to the clindamycin exposure being considered. The INCQUERY table links the exposure exclusions (SUBGROUPs) to their respective exposures of interest (QUERY\_GRPs):

**Table 8. INC\_QUERY**

QUERY_GRP	SUBGROUP
1	11
2	12

Finally, in the QUERY table, which we present again below, the field “EXCL\_MEDS\_ON\_SAME\_DAY” is for the purpose of indicating whether Day 0 (the day of the exposure of interest) is to be included in the look-back period for exclusion. The value should be set to 1 if receipt of any product in the exclusion list on the same day as the exposure of interest (Day 0) (within the respective QUERY\_GRP) disqualifies a dispensing from inclusion. Setting the value to 1 means that no product in the exposure exclusion list and no other formulation of the drug/vaccine of interest can be given on the same day as the drug/vaccine of interest if that latter exposure is to be included. A limitation of the current data extraction program is that the look-back period to determine whether an exposure of interest is incident cannot differ from the look-back period for the other exposure exclusions.

**Table 9. QUERY**

QUERY_GRP	SUBGROUP	WASHTYP	WASHPER	EXCL_MEDS_ON_SAME_DAY
1	1	MULT	183	1
2	2	MULT	183	1
3	3	MULT	183	1

The health outcomes of interest are defined by diagnosis or procedure codes, medical settings, and whether a code or codes were the principal diagnosis. In our examples, all the outcomes of interest are defined by ICD9 diagnosis codes rather than CPT4 procedure codes. The EVENT\_CODES table is populated with those codes for each of the outcomes, namely, angioedema (SUBGROUP 101), AMI (SUBGROUP 102), and seizures (SUBGROUP 103). The EVENT\_TBL\_FLAG is set to 1 for codes corresponding to the health outcome of interest. Settings (AV, ED, IP) can be used with diagnosis or procedure codes to define health outcomes of interest, as explained below the EVENT\_CODES table. However, algorithms of greater complexity (e.g. Dx Code A OR B OR C, AND fever of X degrees; OR Dx Code D AND E) cannot be accommodated by the current program.

**Table 10. EVENT\_CODES**

SUBGROUP	CODETYPE	CODE	EVENT_TBL_FLAG	INCEVENT_TBL_FLAG
101	DX09	995.1	1	1
102	DX09	410.00	1	1
102	DX09	410.01	1	1
102	DX09	410.10	1	1
102	DX09	410.11	1	1
102	DX09	410.20	1	1
102	DX09	410.21	1	1
102	DX09	410.30	1	1
102	DX09	410.31	1	1

SUBGROUP	CODETYPE	CODE	EVENT_TBL_FLAG	INCEVENT_TBL_FLAG
102	DX09	410.40	1	1
102	DX09	410.41	1	1
102	DX09	410.50	1	1
102	DX09	410.51	1	1
102	DX09	410.60	1	1
102	DX09	410.61	1	1
102	DX09	410.70	1	1
102	DX09	410.71	1	1
102	DX09	410.80	1	1
102	DX09	410.81	1	1
102	DX09	410.90	1	1
102	DX09	410.91	1	1
103	DX09	345.0%	1	1
103	DX09	345.1%	1	1
103	DX09	345.2	1	1
103	DX09	345.3	1	1
103	DX09	345.4%	1	1
103	DX09	345.5%	1	1
103	DX09	345.6%	1	1
103	DX09	345.7%	1	1
103	DX09	345.8%	1	1
103	DX09	345.9%	1	1
103	DX09	780.31	1	1
103	DX09	780.32	1	1
103	DX09	780.39	1	1

The EVENT table specifies the medical setting(s) in which to identify the health outcomes of interest (i.e. the setting(s) included in the algorithm), whether there is a requirement that a DX or PX code be a principal diagnosis, and the risk and control intervals relative to the exposure (Day 0).

**Table 11. EVENT**

EVENT_GRP	SUBGROUP	ENCTYPE	PDX	WASHTYP	WASHPER	MIN_DAYS_FR_INDEX_FOR_COMP_WND	MAX_DAYS_FR_INDEX_FOR_COMP_WND	MIN_DAYS_FR_INDEX_FOR_RISK_WND	MAX_DAYS_FR_INDEX_FOR_RISK_WND
1	101			MULT	183	-35	-8	1	28
2	102	IP		MULT	28	15	28	1	14
3	103	IP		MULT	183	1	6	7	10
3	103	ED		MULT	183	1	6	7	10
4	103	IP		MULT	183	15	28	7	10
4	103	ED		MULT	183	15	28	7	10

The health outcome exclusion criteria are specified in the EVENT\_CODES, EVENT, and INCEVENT tables. As is the case for the exposure of interest, there are two kinds of HOI exclusions: one is if the HOI is not incident, the other is if the HOI was preceded by a certain other kind of outcome within a specified period of time. For excluding non-incident cases, we first set the look-back period. In our example studies, we defined incident HOI as the first of its kind to occur in a 183-day period—this is specified by setting WASHPER = 183 in the EVENT table, which leads to exclusion of an HOI occurring < 183 days after a previous one (for the same SUBGROUP). Where the control interval (also called the comparison window, or COMP\_WND) is prior to exposure, the look-back will go back farther than the minimum enrollment specified in the QUERY table. We also need to specify the setting and principal diagnosis criteria for prior cases. This is done using the ENCTYPE and PDX fields in the INCEVENT table. For example, in the clindamycin-AMI study (EVENT\_GRP 2), we want to look for prior AMI in the inpatient (IP) setting only (and if any such case is found in the 28 days prior to a case being considered, the more recent case is excluded). ENCTYPE is left blank if all three settings are to be used. For example, in the MMRV-seizure studies (EVENT\_GRPs 3 and 4), the seizure outcomes to be used must be in either the ED or inpatient setting, but must not follow, within 183 days, a seizure in *any* setting.

**Table 12. INCEVENT**

EVENT_GRP	SUBGROUP	ENCTYPE	PDX
1	101		
2	102	IP	
3	103		
4	103		

The other kind of HOI exclusion that can be specified is if some *other* health outcome occurred within a certain period of time prior to the one of interest. To specify this, one goes to the EVENT\_CODES table and sets EVENT\_TBL-FLG to 0 and INCEVENT\_TBL\_FLAG to 1. In our example studies, we did not have

HOI exclusions based on other health outcomes. If we had, we would have added them as SUBGROUPs to the EVENT\_CODES and INCEVENT tables.

Finally, the exposure period for which data are to be extracted from the M-S distributed dataset must be specified for each data partner in the EXPOSURE\_PERIODS table. Specifying exposure date range does not restrict the follow-up period, only the dispensing, or exposure, period. Because of differences among data partners with respect to date ranges, frequency of data updates, and completeness of data in each update, etc., it will likely be appropriate to select different exposure date ranges for different Data Partners. Also, it might not be possible to pre-specify all date ranges within each Data Partner into the future. For each new data update from a specific Data Partner, a date range immediately following the date range used the last time that Partner’s data were extracted and analyzed should be chosen. It is important that there be no overlap in the sequence of exposure periods specified for each data partner, because the analysis method does not allow for refreshment, correction, and re-use of previously analyzed data. In the examples below, HM-HPHC data were obtained monthly, with one additional consecutive month’s worth of dispensings collected each time, while chunks of dispensings for AE-OS and KP-NC were 3 months long.

**Table 13. EXPOSURE\_PERIODS**

DPID	SITEID	EXPOSURE_PERIOD_CD	EXPOS_PERIOD_STRT_DT	EXPOS_PERIOD_END_DT
HM	HPHC	1	7/1/2012	7/31/2012
HM	HPHC	2	8/1/2012	8/31/2012
HM	HPHC	3	9/1/2012	9/30/2012
HM	HPHC	4	10/1/2012	10/31/2012
HM	HPHC	5	11/1/2012	11/30/2012
HM	HPHC	6	12/1/2012	12/31/2012
AE	OS	1	1/1/2012	3/31/2012
AE	OS	2	4/1/2012	6/30/2012
KP	NC	1	1/1/2012	3/31/2012
KP	NC	2	4/1/2012	6/30/2012

If these tables are filled out in Access, they must be imported into SAS.

## **B. FOR MINI-SENTINEL OPERATIONS CENTER**

### **1. Documentation**

This documentation exists as a stand-alone document called “Mod\_prg\_410\_Activity\_Documentation.docx.”

This document has documentation for two programs in the package for Mini-Sentinel activity 4.10.

1. **Program run\_mp410\_distrib\_prg.sas. This is a wrapper for running modular program mp410.**  
**In the first section of the program one needs to update parameters. The rest of the program has “include” statements for standard MSCDM macros, the main macro mp410.sas and finally the macro call %mp410.**
2. **Macro mp410.sas. This is the main macro which performs the actual data manipulation and calculation for Mini-Sentinel activity 4.10**

```

/*****
*****

```

```

*Program:      run_mp410_distrib_prg.sas
*Project:      Mini-Sentinel Activity 4.10
*Purpose:      Wrapper to run all necessary programs and macros;
-----

```

\* Input parameters:

Standard MSCDM input parameters including parameters for libraries, directories, site id, etc.

**EXPOSURE\_PERIOD\_CD:** Macro parameter. Should be used in conjunction with the dataset **EXPOSURE\_EPISODES\_FILE**. See below for details. Numeric (1, 2, 3,...). Required.

The dataset **EXPOSURE\_EPISODES\_FILE** is distributed as a part of the package. In addition, to the column EXPOSURE\_PERIOD\_CD the dataset has two columns: EXPOS\_PERIOD\_STRT\_DT and EXPOS\_PERIOD\_END\_DT. These dates can be different for different data partner sites depending on sites data availability and refresh rate. For example, the exposure period can be a month length for one site and a quarter for another. This is possible because the study compares two statistics calculated at same data partner site: the number of adverse outcomes in risk time window with the number of adverse outcomes in comparison time window. The time periods appropriate to the site should be supplied by MINI-SENTINEL OPERATIONS CENTER or updated by local site per leading site instructions.

The allowed values are either a single number or a list of numbers separated by spaces. The latter is mostly used on historical data to determine background rates.

Examples:

```

%LET exposure_period_cd=1;
%LET exposure_period_cd=1 2 3;
*-----;
*-----;
*-----;

```

```

/* HEADER -----
|

```

```

| PROGRAM: mp410
|
| PURPOSE:   Create summary tables for drug of interests and adverse events.
|
| INPUT:     Patients level tables: Tables in MSCDM format

```

Parameters, Lookup and Mapping tables used by the program.

The Lookup and Mapping tables are distributed as a part of a package for modular program.

\*-----;

**EXPOSURE\_PERIOD\_CD:** Macro parameter.  
Should be used in conjunction with the dataset EXPOSURE\_EPISODES.  
See below for details.  
Numeric (1, 2, 3,...). Required.

\*-----;

**ENROLGAP:** Macro parameter.  
Grace period used to bridge intervals of enrollment.  
Numeric. Default 45 days.

\*-----;

The dates QUERYFROM and QUERYTO below should be set very wide.  
The actual restrictions by dates are set by the mapping tables:  
EXPOSURE\_PERIODS\_FILE, QUERYFILE, EVENTFILE (WASHPER).  
The only case when the above dates will be used is when WASTYP="MIN" which rarely necessary.

\*-----;

**TMP:** Macro Parameter for temporary library. Default is WORK.  
It can be changed to the permanent library for troubleshooting.

\*-----;

**TMP2:** Macro Parameter for temporary library. Default is WORK.  
This parameter should be changed only if your MSCDM tables are in the relational database and you have permission to create temporary tables in the same database. If this is the case, set **TMP2** to the library associated with a schema in the database in which you can create tables. This can significantly improve performance.

\*-----;

**LKP\_FOR\_ALL\_CODES\_FILE:** Macro parameter.  
Lookup table that has complete set of code for columns used by various mapping tables below.

COLUMNS:

**COLUMN\_NAME:** Name of the code column used in mapping tables defined below.  
For example, GROUP, QUERY\_GRP, etc.

**CODE\_VALUE:** Value for numeric code.

**CODE\_VALUE\_C:** Value for character code.

**CODE\_DESC:** Code description

\*-----;

**GROUPS\_MAPPING:** Macro parameter.  
Mapping table between study group code (GROUP) and QUERY\_GRP  
and EVENT\_GRP.

COLUMNS:

**GROUP:** Code for a study group.  
A code which identifies both population selection (QUERY) and adverse  
effects (EVENT) set of conditions.  
Numeric (1, 2, 3,...). Required.

**QUERY\_GRP:** A code which identifies set of conditions for population selection.  
Usually, defines population who had exposures to a set of drugs.  
Numeric (1, 2, 3,...). Required.

**EVENT\_GRP:** A code which identifies set of conditions for adverse events of interest.  
Numeric (1, 2, 3,...). Required.

\*-----;

**EXPOSURE\_PERIODS\_FILE:** Macro parameter. Mapping table for exposure periods.

The dataset is distributed as a part of the package.

In addition, to the column EXPOSURE\_PERIOD\_CD the dataset has two columns:

EXPOS\_PERIOD\_STRT\_DT and EXPOS\_PERIOD\_END\_DT. These dates can be  
different for different data partner sites depending on sites data availability  
and refresh rate. For example, the exposure period can be a month length  
for one site and a quarter for another. This is possible because the study compares  
two statistics calculated at same data partner site: the number of adverse outcomes in  
risk time window with the number of adverse outcomes in comparison time window.

The time periods appropriate to the site should be supplied by  
MINI-SENTINEL OPERATIONS CENTER or updated by local site per leading site  
instructions.

COLUMNS:

**EXPOSURE\_PERIOD\_CD:** Numeric (1, 2, 3,...)

**EXPOS\_PERIOD\_END\_DT:** SAS date.

**EXPOS\_PERIOD\_STRT\_DT:** SAS date.

\*-----;

**QUERYFILE:** Parameter.  
Mapping table for population selection variables (QUERY).

COLUMNS:

**QUERY\_GRP:** A code which identifies set of conditions for population selection. Usually, defines population who had exposures to a set of drugs. Numeric (1, 2, 3,...). Required.

**SUBGROUP:** Identifies group of codes, typically, a set of drugs. Numeric (1, 2, 3,...). Required.

**WASHTYP:** Selects how incidence events will be defined. Character \$ 3. Possible values the same as in MP3: "MIN", "SING", "MULT".

**WASHPER:** Length of event washout period in days. Possible values the same as in MP3

**EXCL\_MEDS\_ON\_SAME\_DAY:** When this flag is set to 1 then cases with more than one NDC from the same QUERY\_GRP and on the same day are excluded. In other words, the look back period starts from 0 days i.e. look back period [0, WASHPER]. When this flag is set to 0 then the look back period starts with previous day i.e. look back period is [1,WASHPER]. Possible values are 0 and 1.

\*-----;

**INCQUERYFILE:** Macro parameter.  
Mapping table with additional parameters for population selection. Identifies conditions for previous incidence of exposure which are used to exclude exposure encounters (define by QUERYFILE). Used in conjunction with WASHTYP and WASHPER from QUERYFILE. Optional.

COLUMNS:

**QUERY\_GRP:** A code which identifies set of conditions for population selection. Usually, defines population who had exposures to a set of drugs. Numeric (1, 2, 3,...). Required.

**SUBGROUP:** Identifies group of codes, typically, a set of drugs. Numeric (11, 12, 13,...). Required.

\*-----;

**QUERY\_CODES\_FILE:** Macro parameter.

Mapping table between group of codes (SUBGROUP) and actual codes used for population selection (QUERY/INCQUERY).

**COLUMNS:**

- SUBGROUP:** Identifies group of codes, typically, a set of drugs. Numeric (11, 12, 13,...). Required.
- CODE:** NDC, diagnosis or procedure code mapped to a subgroup above. The wild cards in code values are allowed. One can use either UNIX style wild cards(?,\*) or SQL style wild cards (\_,%).
- The presence or absence of a period in PX or DX code does not affect the results.
- CODETYPE:** Details: type of each code value included in the CODE field above. Character \$ 4. Required

Valid values are:

- RX11: 11-digit NDC
- RX09: 9-digit NDC
- PX09: ICD-9-CM procedure
- PX10: ICD-10-CM procedure
- PX11: ICD-11-CM procedure
- PXC4: CPT-4 procedure (i.e., HCPCS Level I)
- PXHC: HCPCS procedure (i.e., HCPCS Level II)
- PXH3: HCPCS Level III
- PXC2: CPT Category II
- PXC3: CPT Category III

QUERY\_TBL\_FLAG/INCQUERY\_TBL\_FLAG:Optional fields that identify if SUBGROUP is QUERY and/or INCQUERY table.

They have no affect on the program but can be useful for validation.

\*-----;

**EVENTFILE:** Macro parameter.  
 ~~~~~ Mapping table for adverse events table by EVENT\_GRP.

The tables **EVENTFILE** and **INCEVENTFILE** allow NULL values that must be interpreted as any allowed values. For example, ENCTYPEIS NULL in one of these tables means that all five possible values of ENCTYPE are allowed. The JOIN to the table LKP\_FOR\_ALL\_CODES create a table which has explicit set of all possible codes;

**COLUMNS:**

- EVENT\_GRP:** A code which identifies set of conditions for adverse events of interest. Numeric (1, 2, 3,...). Required.
- SUBGROUP:** Identifies group of codes, typically, a set of diagnosis. Numeric (101, 102,...) Required.
- WASHTYP:** Selects how incidence events will be defined. Character \$ 3. Possible values the same as in MP3: "MIN", "MULT".
- WASHPER:** Length of event washout period in days. Possible values the same as in MP3  
 MP3  
 Numeric. Required (0 must be entered if no WASHPER is required).
- ENCTYPE:** Encounter type. The same as in MSCDM data model.  
**PDX:** Principal diagnosis. As in MSCDM in data model it is defined only for inpatient stays.

**Columns for comparison and risk windows:**

- MAX\_DAYS\_FR\_INDEX\_FOR\_COMP\_WND**  
**MAX\_DAYS\_FR\_INDEX\_FOR\_RISK\_WND**  
**MIN\_DAYS\_FR\_INDEX\_FOR\_COMP\_WND**  
**MIN\_DAYS\_FR\_INDEX\_FOR\_RISK\_WND**

\*-----;

**INCEVENTFILE:** Macro parameter.

Mapping table with additional parameters for adverse outcomes selection. Identifies conditions for previous outcomes which are used to exclude adverse outcomes (defined by EVENTFILE). Used in conjunction with WASHTYP and WASHPER from EVENTFILE. Optional.

**COLUMNS:** (Same meaning and data types as in **EVENTFILE**)

- EVENT\_GRP**  
**SUBGROUP**  
**ENCTYPE**  
**PDX**

**EVENT\_CODES\_FILE:** Macro parameter.

Mapping table between group of codes (SUBGROUP) and actual codes used for adverse events (EVENTS/INCEVENTS). Has similar structure as QUERY\_CODES\_FILE. The two tables are separated for convenience.

The wild cards in code values are allowed. One can use either UNIX style wild cards(?,\*) or SQL style wild cards (\_,%).

The presence or absence of a period in PX or DX code does not affect the results.

\*-----;

**AGE\_GROUPS\_FILE:** Macro parameter.  
Mapping table for age groups.

**COLUMNS:**

**QUERY\_GRP:** See QUERYFILE above  
**AGE\_GROUP:** Code for age group.Numeric (1, 2, 3,...). Required.  
**AGE\_UNIT:** Unit for measuring age. Character \$ 1.

Valid values:

- D: days
- W: weeks
- Q: quarters
- M: months
- Y: years (default value)

**AGE\_MAX** Minimum age at exposure (index) date. Numeric. Default 0.  
**AGE\_MIN** Maximum age at exposure (index) date. Numeric. Default 200.

\*-----;

**OUTPUT:**

All output tables, reports and logs use output prefix below

out\_tbl\_prefix: **&DPID.&REQUESTID;**  
 out\_rpt\_prefix: **&out\_tbl\_prefix.&DateStamp.&TimeStamp;**  
 \*-----;

Summary datasets on event\_date (relative to exposure\_dt) level:

**SUM\_EVNT:**

Summarization level:

**GROUP EXPOSURE\_PERIOD\_CD EXPOS\_PERIOD\_STRT\_DT AGE\_GROUP SEX**  
**DAYS\_FROM\_EXPOS\_TO\_EVENT DAYS\_ENR\_AFTER\_EXPOS**

Statistics:

CNT\_EXP\_EPISODES  
CNT\_IN\_COMP\_WND  
CNT\_IN\_RISK\_WND  
CNT\_CENSORED\_EVENTS

\*-----;

**SUM\_EXPS\_PERIOD:** Summary datasets for adverse events on exposure period summarization level stratified by gender and age\_group.

Summarization level:

**GROUP EXPOSURE\_PERIOD\_CD EXPOS\_PERIOD\_STRT\_DT AGE\_GROUP SEX**

Statistics:

CNT\_EXP\_EPISODES  
CNT\_IN\_COMP\_WND  
CNT\_IN\_RISK\_WND  
CNT\_CENSORED\_EVENTS

MIN\_DAYS\_FROM\_EXPOS\_TO\_EVENT  
MAX\_DAYS\_FROM\_EXPOS\_TO\_EVENT  
MIN\_DAYS\_ENR\_AFTER\_EXPOS  
MAX\_DAYS\_ENR\_AFTER\_EXPOS

\*-----;

**SUM\_EXPS\_PERIOD\_FIN:** Summary datasets for adverse events on exposure period summarization level without gender and age stratification. This is the dataset that will be used as input by R program.

Summarization level:

**DPID\_SITEID GROUP EXPOSURE\_PERIOD\_CD EXPOS\_PERIOD\_STRT\_DT**

Statistics:

CNT\_EXP\_EPISODES  
CNT\_IN\_COMP\_WND  
CNT\_IN\_RISK\_WND

## 2. Program Steps

- Import mapping and lookup tables supplied with the package.
- Do some initial processing of mapping and lookup tables.
- These include setting some defaults, joining some of the tables together and creating.
- macro variables for min and max dates which will be used latter in the program.
- Create tables with complete set of all NDC, PX, and DX codes ever used by the MSCDM data warehouse at a given site.
- This will be used latter to create an explicit list of codes of interest without wild cards and
- in the format used by a site (for example, with or without a period in DX or PX codes).
- Extract medical claims from the diagnosis and procedure files.
- Extract drug claims from the outpatient pharmacy file.
- Combine NDC, DX and PX data extracts.
- Summarize RX data for dispenses on the same date.
- Keep one record per combination of MSCDM variables that suppose to be unique per MSCDM data model.
- The last two steps are added in case a site does not completely comply with MSCDM data model.
- Separate extracted encounters into ones that will be used for population selection (query conditions) and the ones that will be used for adverse outcome (events conditions).
- Reset PDX to missing when it is not principal diagnosis because we need to treat missing and secondary diagnosis in the same way.
- Add population selection parameters from the mapping table to the "QUERY/ INQUERY" data extract.
  
- Apply population selection criteria to find patients with exposure to the drugs
- or other encounters of interest.
- Extract enrollment and demographics information for the population of interest.
- Create enrollment spans which bridge enrollment intervals separated by allowed grace period;
- Add enrollment and demographics data to the table with patients who had exposures;
- Calculate age in units (years, months, etc) defined by parameter AGE\_INTERVAL\_TYPE;
- Exclude patients that do not have age in the required age intervals;
- Find patients that can have outcomes of interest and add parameters for adverse event conditions;
- Find patients who satisfy WASHPER condition for adverse events.
- Combine exposure information and adverse outcome information into a single dataset;
- Match query conditions (QUERY\_GRP) with the corresponding event conditions (EVEN\_GRP).
- Add study group code;
- Create flags for events in comparison and risk windows.
- Create variable for number days enrolled after exposure.
- Set CENSORED\_FLAG=1 when enrollment ends within a comparison or risk window.
- The variable can be used to censor events.
- Create Final datasets.
- Create patient exposure episode level dataset in **DPLocal** library.
- Keep at a datapartner site; **&out\_tbl\_prefix.\_PAT\_EVENT**
- Create summary exposure-event level dataset in **DPLocal** library

- Keep at datapartner site; **&out\_tbl\_prefix.\_SUM\_EXPS\_EVNT\_LEV**
- Create summary dataset on event level in **MSOC** library;
- To send to MSOC; **MSOC.&out\_tbl\_prefix.\_SUM\_EVNT\_LEV**
  
- Create summary dataset on exposure period level stratified by gender and age\_group in **MSOC** library
- To send to MSOC; **MSOC.&out\_tbl\_prefix.\_SUM\_EXPS\_PERIOD\_LEV**
  
- Create final summary dataset on exposure period level in **MSOC** library (no gender and age stratification).
- To send to MSOC; **MSOC.&out\_tbl\_prefix.\_SUM\_EXPS\_PERIOD\_LEV\_FIN**

### **C. FOR DATA PARTNERS**

For the extraction of data, the Data Partners will each receive a zip file which contains the SAS distributed programs and instructions for using them. There are two programs for the extraction of data and the creation of output tables and a log file. The values from the output tables will be used in the R-code analysis. Below are general instructions for using the program:

Download zip file containing SAS program, and mapping and lookup tables. The first program is called **run\_mp410\_distrib\_prg.sas** and will extract data based on the parameters chosen. It contains a macro, **mp4\_10.sas**, that will perform calculations from extracted data to create variables for SAS output tables.

The file QUERY\_CODES\_FILE is a macro parameter that will organize the codes from lkp\_for\_all\_codes.sas7bdat into groups for selection of the cohort with the HOI and the exposure.

Join the mapping tables EVENTFILE (HOI, defined) and QUERYFILE (exposure group, defined) as directed by instructions in zip file. A link will be established between HOI and exposure criteria.

Extract medical claims from MSCDM diagnosis and procedure files.

Extract drug claims from MSCDM outpatient and pharmacy files.

Combine NDC, DX and PX data extracts. Check the data for missing fields, and set PDX as “missing” when it appears as a secondary (or any non-principal) diagnosis.

Using the QUERYFILE and EXPOSURE\_PERIODS\_FILE mapping tables, apply population selection criteria to locate all patients with exposure.

Utilize the EVENTFILE and INCEVENTFILE mapping tables to define the population of patients exposed, those exposed patients who also were diagnosed with the HOI within the risk or control windows of exposure. The WASHPER condition (length of event washout period in days) from the EVENTFILE mapping table will specify additional conditional parameters for the identified HOI-exposure pairs that fall into the risk window. Utilize AGEGROUPS\_FILE to separate out age groups.

Utilize ENROLGAP macro parameter to bridge enrollment intervals separated by the grace period. Use this information to further specify the cohort created in STEP 8.

Create final datasets: 1 local (remains at Data Partner site), 3 MSCDM:

Local: Patient exposure level, of every HOI in the surveillance period

MSCDM: Summary of exposure-period level

MSCDM: Summary of event level

MSCDM: Summary of combined exposure-event level

Return Log to MSCDM

See Appendix 1 for example of instructions sent to Data Partners.

#### IV. SEQUENTIAL ANALYSIS (R FUNCTIONS)

Data analysis will use the free and open-source R statistical program. The R code is written to be used both for PROMT module 1 using a self-control sequential design and for PROMPT module 2 using a concurrent propensity score matched design. Based on the binomial maxSPRT, both use exact sequential analysis with a binomial probability model. If each data feed include at most one case or control, the

functions perform continuous sequential analysis. If data feeds include more than one case or control, the function performs a group sequential analysis with variable group sizes. If some data feeds have only one case or control, while others have more than one, the functions perform a hybrid continuous-group sequential analysis.

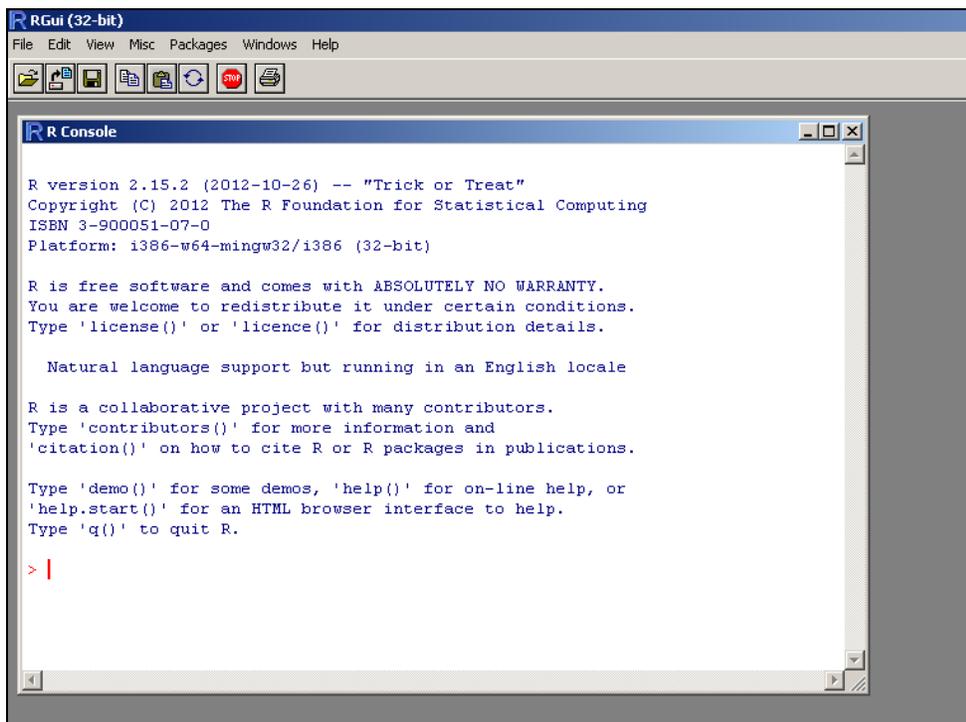
The R code script named Combined\_Analyze.Binomial\_v1.0 was created for self-controlled risk interval continuous and group sequential analysis. No more than the pre-defined parcel of Type I error ( $\alpha$ ) is spent at each sequential test (see the descriptive readme.txt file included with the Combined\_Analyze.Binomial\_v1.0 script). The data set to be used will be from the data extraction described previously in this manual and will consist simply of counts of the health outcome of interest in a risk interval after the drug or vaccine exposure and in a control interval, referred to as “cases” and “controls,” respectively, in the R code output. To utilize the R-program for sequential analysis, you will open the R program, open the R-script Combined\_Analyze.Binomial\_v1.0, run the R-script Combined\_Analyze.Binomial\_v1.0, enter the parameters of up to two separate functions, and subsequently run the function to analyze your dataset. What follows are step by step instructions to load and run the R code functions for a dataset analysis, as well as explanations of the R code output.

## A. DOWNLOADING R AND LOADING FUNCTIONS

The R-program is available for download from the Comprehensive R Archive Network, “CRAN” (<http://cran.r-project.org>). Detailed instructions for various operating systems are listed on the CRAN website.

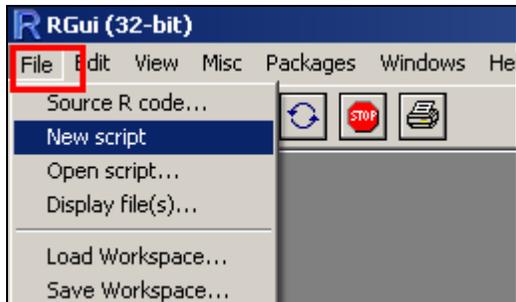
Once you have downloaded the R program, open it, and click on the R Console window.

**Figure 1.** R Console Window



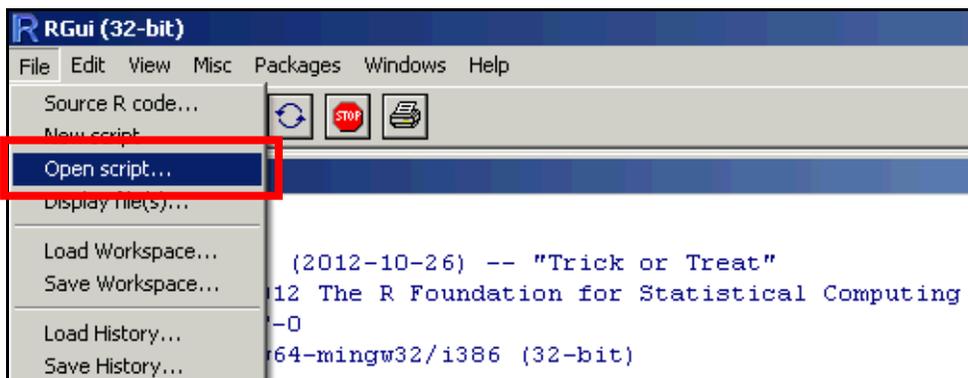
From the Menu Bar, select “File”

**Figure 2.**



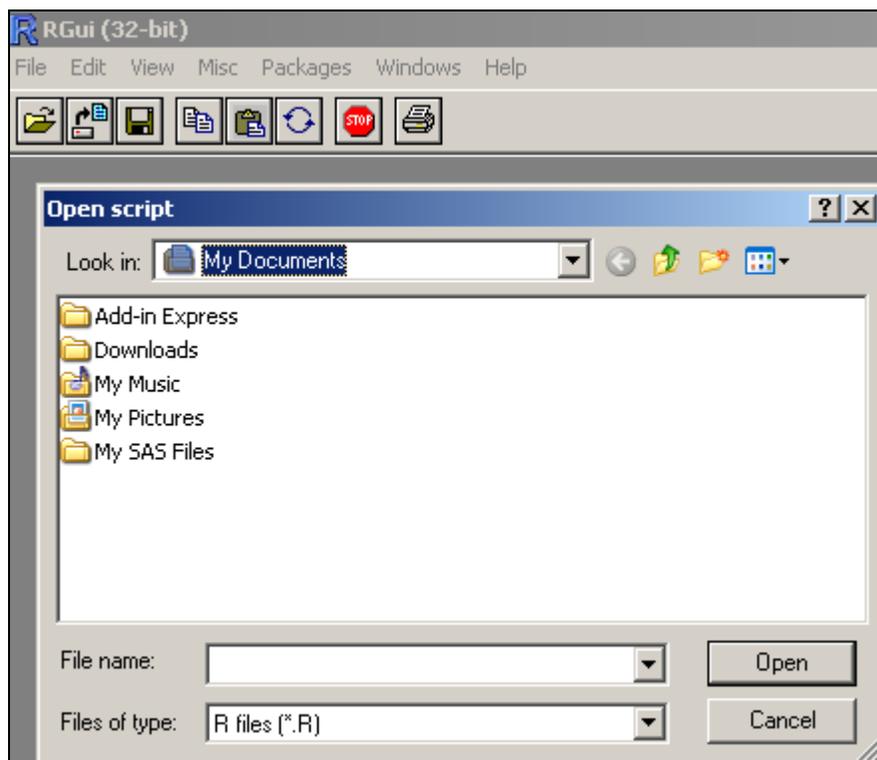
From the drop down of the File Menu, select “Open Script...”

**Figure 3.**



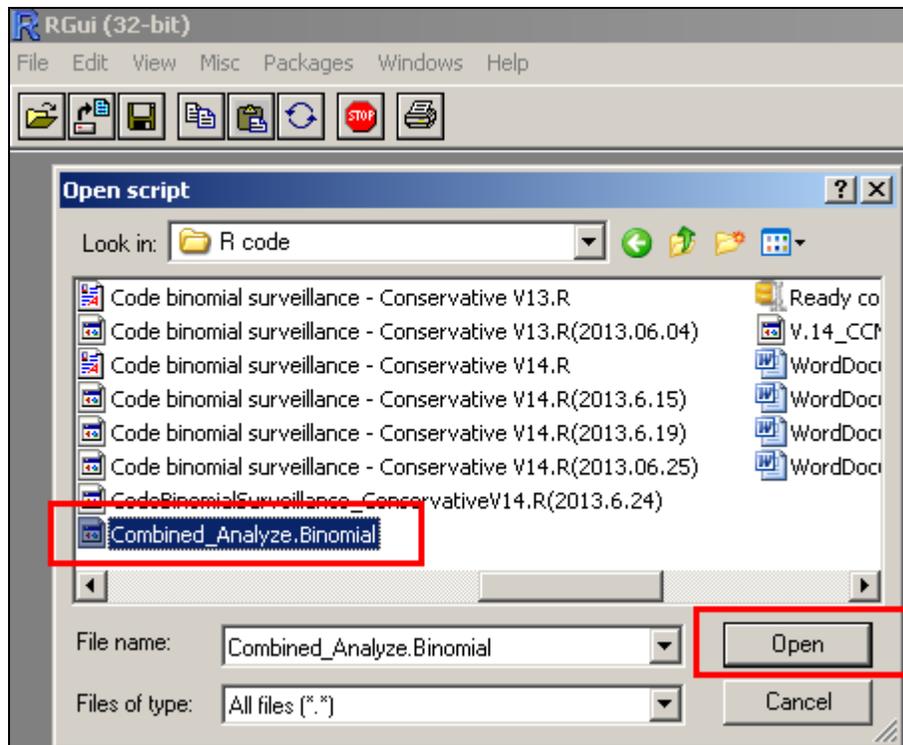
A pop-up window will ask the user to locate the R-script. The R-script is distributed separately from the User Manual and is named Combined\_Analyze.Binomial\_v1.0. The user is responsible for deciding where this script is saved.

**Figure 4.**



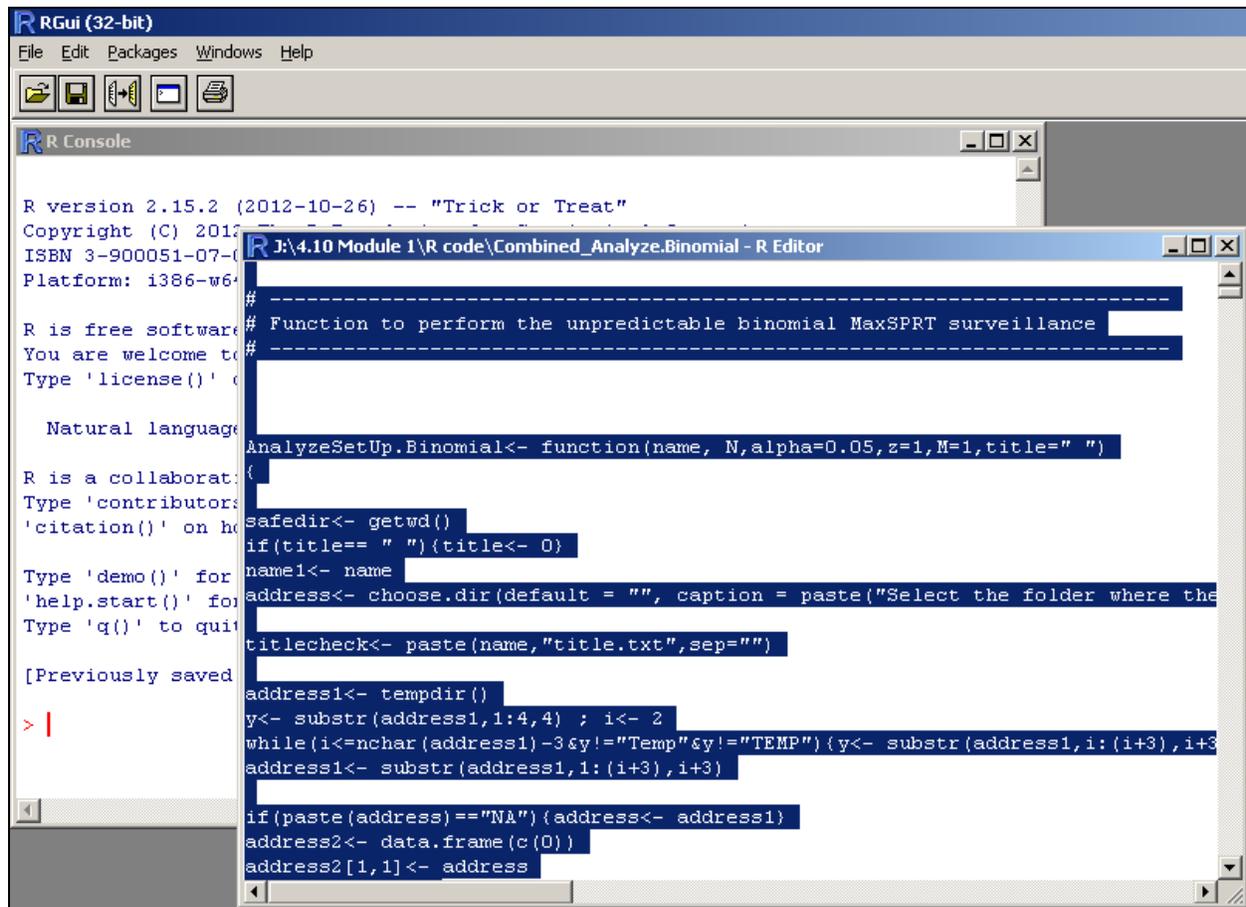
Select “Combined\_Analyze.Binomial\_v1.0” and open this file.

**Figure 5.**



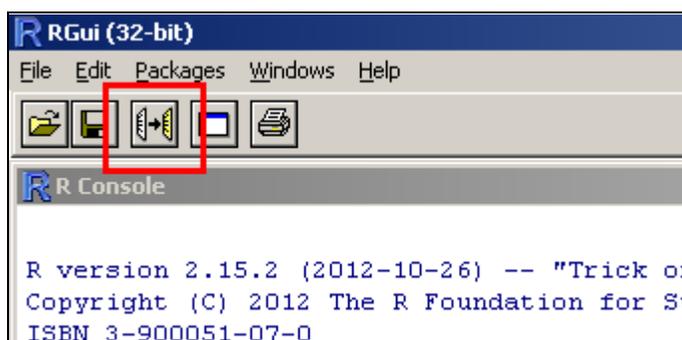
Click on the (newly opened) window of the Combined\_Analyze.Binomial\_v1.0 script, highlight all of the script.

Figure 6.



Once the script has been highlighted, click the “run “  button that appears in the Menu Bar

Figure 7.





The second function, “Analyze.Binomial,” is where the user will call a filename, and values from the SAS extraction table (from Data Partners) will be used as cases and controls for the sequential analysis.

### 1. Function “AnalyzeSetUp.Binomial”

Prior to any analysis, the user will specify a location for and name for each cumulative data file and unique parameters for *all* Drug:HOI pairs. The cumulative files will be created with

**AnalyzeSetUp.Binomial** function in the **R Console** window with a (case sensitive) template as follows:

**AnalyzeSetUp.Binomial(name= “...”,N=#, alpha=#,z=#, M=#,title= “ ... ”)**, where  
**name** = name of the file that will store cumulative analyses of one Drug : HOI pair.

The other items to be entered into the template include the N, alpha, and M discussed in Section B.2., namely:

**N** = number of cases in risk and control windows combined, at which to stop surveillance without rejecting the null hypothesis

**alpha** = alpha level for the surveillance applied to all tests (optional, default = 0.05)

**M** = minimum number of cases to signal (optional)

**z** = ratio of length of the control window to length of the risk window

Lastly, a title may be chosen:

**title** = title for display across the tables and graphical output produced (optional)

Once the parameters of **AnalyzeSetUp.Binomial** function have been filled and a filename has been chosen, for example:

**name** = “TestFile”

**N** = 50

**alpha** = 0.05

**z** = 1

**M** = 3

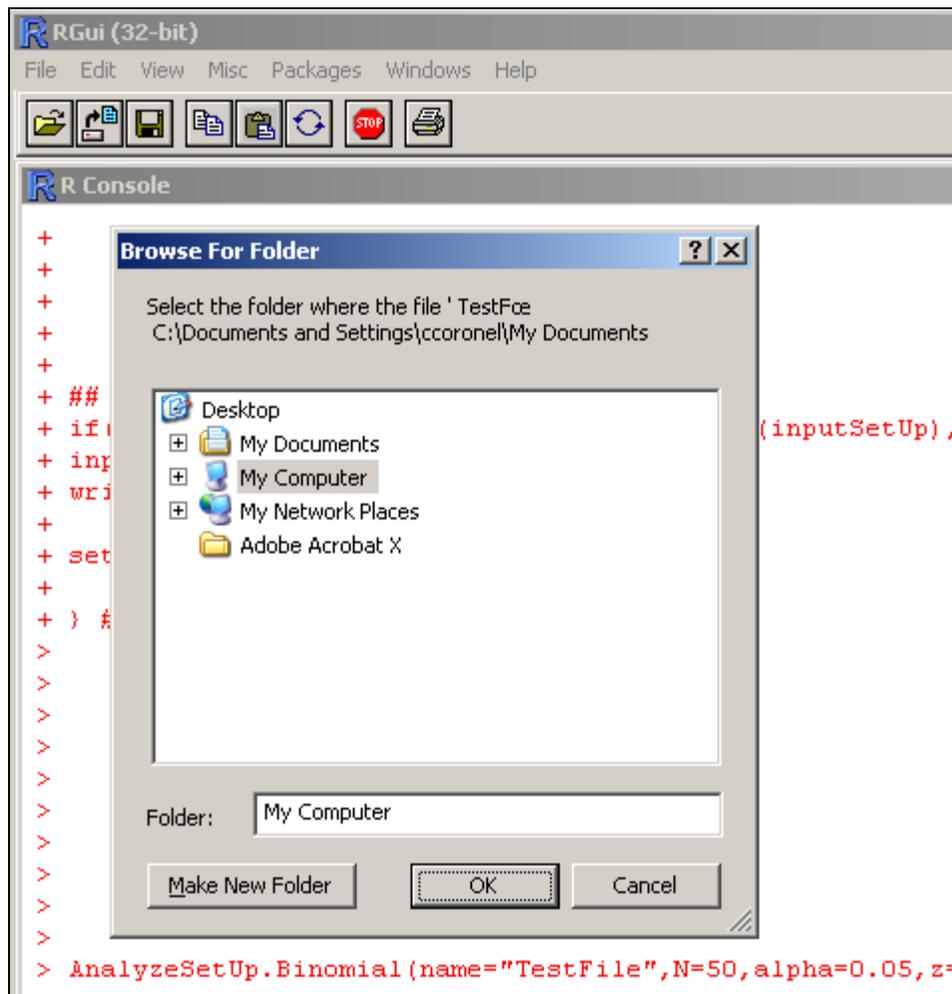
**title** = “ExampleRun”

```
> AnalyzeSetUp.Binomial(name="TestFile",N=50,alpha=0.05,z=1,M=3,title="ExampleRun")
```

the user will run the function by pressing “Enter.”

Immediately following this, the user will be asked to specify the location of the file in a pop-up

Figure 9.



## 2. Function “Analyze.Binomial”

The user will next be executing the function **Analyze.Binomial** to analyze the counts of HOI events returned by the Data Partners in the extraction table. Events that occur in the risk window will be called “cases,” and events that occur in the control window will be called “controls”. An example of output returned from a Data Partner is below.

Figure 10.



| Table      | View   | DP          | Exposure      | Start Date   | End Date     | Total Events    | Events in Control Window "Controls" | Events in Risk Window "Cases" |
|------------|--------|-------------|---------------|--------------|--------------|-----------------|-------------------------------------|-------------------------------|
| DPID_SITID | GROUP  | STUDY_GROUP | EXPOSURE_PERI | EXPOS_PERIOD | EXPOS_PERIOD | CNT_EXP_EPISODE | CNT_IN_COMP_W                       | CNT_IN_RISK_W                 |
| 1          | hmhphc | 1           | LISINOPRIL    | 01SEP2012    | 30SEP2012    | 216             | 1                                   | 5                             |
| 2          | hmhphc | 2           | CLINDAMYCIN   | 01SEP2012    | 30SEP2012    | 21              | 2                                   | 1                             |
| 3          | hmhphc | 3           | MMRV-1        | 01SEP2012    | 30SEP2012    | 1               | 0                                   | 0                             |
| 4          | hmhphc | 4           | MMRV-2        | 01SEP2012    | 30SEP2012    | 1               | 0                                   | 0                             |

The analysis will be executed by entering values into the function **Analyze.Binomial** in the **R Console** window with a (case sensitive) template as follows:

**Analyze.Binomial**(name= "...",test=#,cases<- #,controls<- #), where  
**name** = name of the text file which will store all associated cumulative tests associated with the unique parameters for each Drug:HOI pair  
**test** = number of the test, whether it is the 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>, etc  
**cases <-** = numeric value from extraction table of the HOI events that occur in the risk window (when entering the number, leave one space after the left-pointing arrow)  
**controls <-** = numeric value from extraction table of the HOI events that occur in the control window (again, when entering the number, leave one space after the left-pointing arrow)

As an example, we have chosen the following values to run our analysis:

```
name = "TestFile"
test = 1
cases<- = 5
controls<- = 1
```

and the **Analyze.Binomial** input for the R Console is as follows:

```
> Analyze.Binomial(name="TestFile",test=1,cases<- 5,controls<- 1)
```

It is important to note that for **Analyze.Binomial**, as time goes on and new data are received for analysis, all values for "test" must be entered in chronological order, in the order the corresponding data were received. If any incorrect values for a test are entered, then it will be necessary to delete the corresponding text file established previously via **AnalysisSetUp.Binomial** and run all tests again. The reason for deleting the file is that the values from all previous tests are stored and used, and the file and any stored analyses and descriptive statistics will no longer be correct. It is strongly suggested that the user keep a separate, detailed log of all data extractions from contributing Data Partners.

After entering the values into the R Console, press the Enter key to run the function **Analyze.Binomial**. The **Analyze.Binomial** function will run sequential analysis and produce an output of a table and four graphs. This may take several minutes, depending on the number of cases and controls.

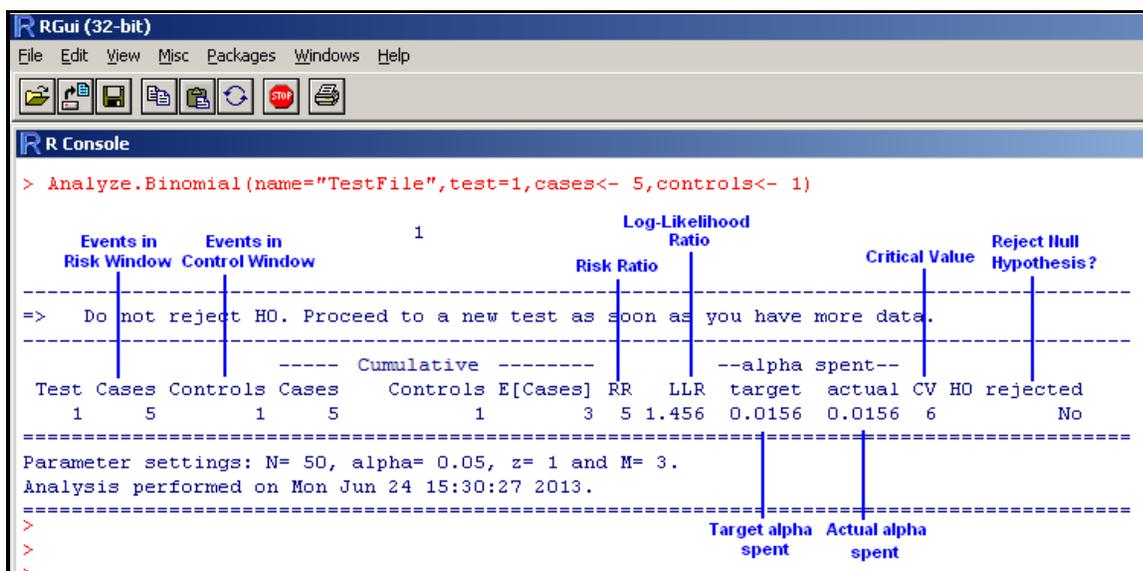
In order to run additional sequential tests using the same analysis parameters for an exposure:HOI pair, the user will repeat the steps in Section D.1 (Downloading R and Loading Functions) to define the function **Analyze.Binomial** and then follow the steps in D.2.b. (**Analyze.Binomial**), using the

Analyze.Binomial function and referencing the already existing file containing the stored values from previous analyses.

### C. R OUTPUT

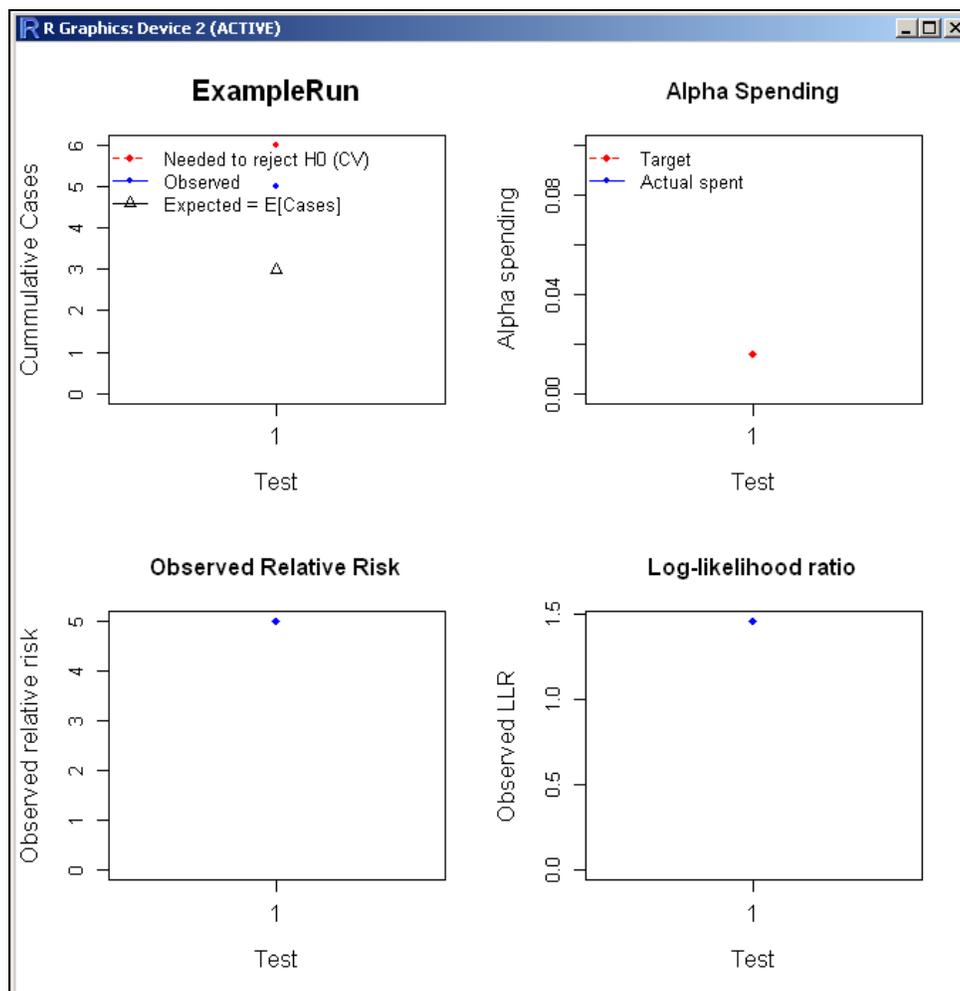
The output table in the R Console will include for each test: the counts of cases and controls (per test and cumulatively), the expected number of cases under the null hypothesis (“E[Cases]”), the risk ratio (“RR”), log likelihood ratio test statistic (“LLR”), target alpha spent (“target”), actual alpha spent (“actual”), cumulative number of cases needed in the risk window in order to reject the null hypothesis (critical value, “CV”), and whether the null hypothesis was rejected (“HO rejected”).

Figure 11.



Graphic output from R showing the numbers of cases observed, expected, and needed to reject the null hypothesis; alpha spending; observed relative risk; and log-likelihood ratio **for the last test run** will appear in a separate window within the R program:

Figure 12.

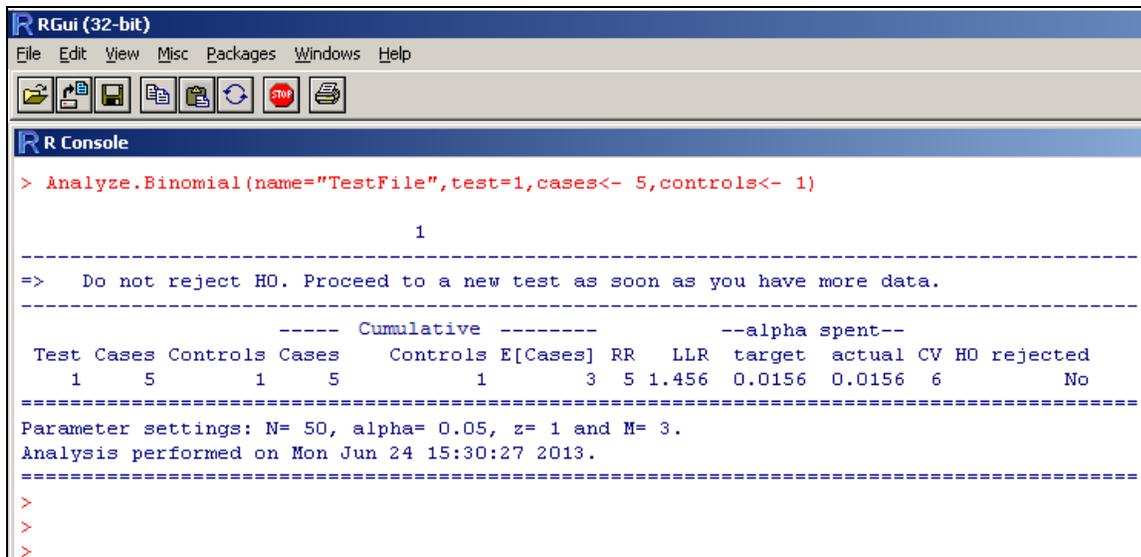


## D. EXAMPLE TESTS

### 1. Analyze.Binomial(name="TestFile",test=1,cases<- 5,controls<- 1)

Test 1 is explained in figures 13 and 14.

Figure 13.



```

RGui (32-bit)
File Edit View Misc Packages Windows Help
[Icons]

R Console
> Analyze.Binomial(name="TestFile",test=1,cases<- 5,controls<- 1)

1

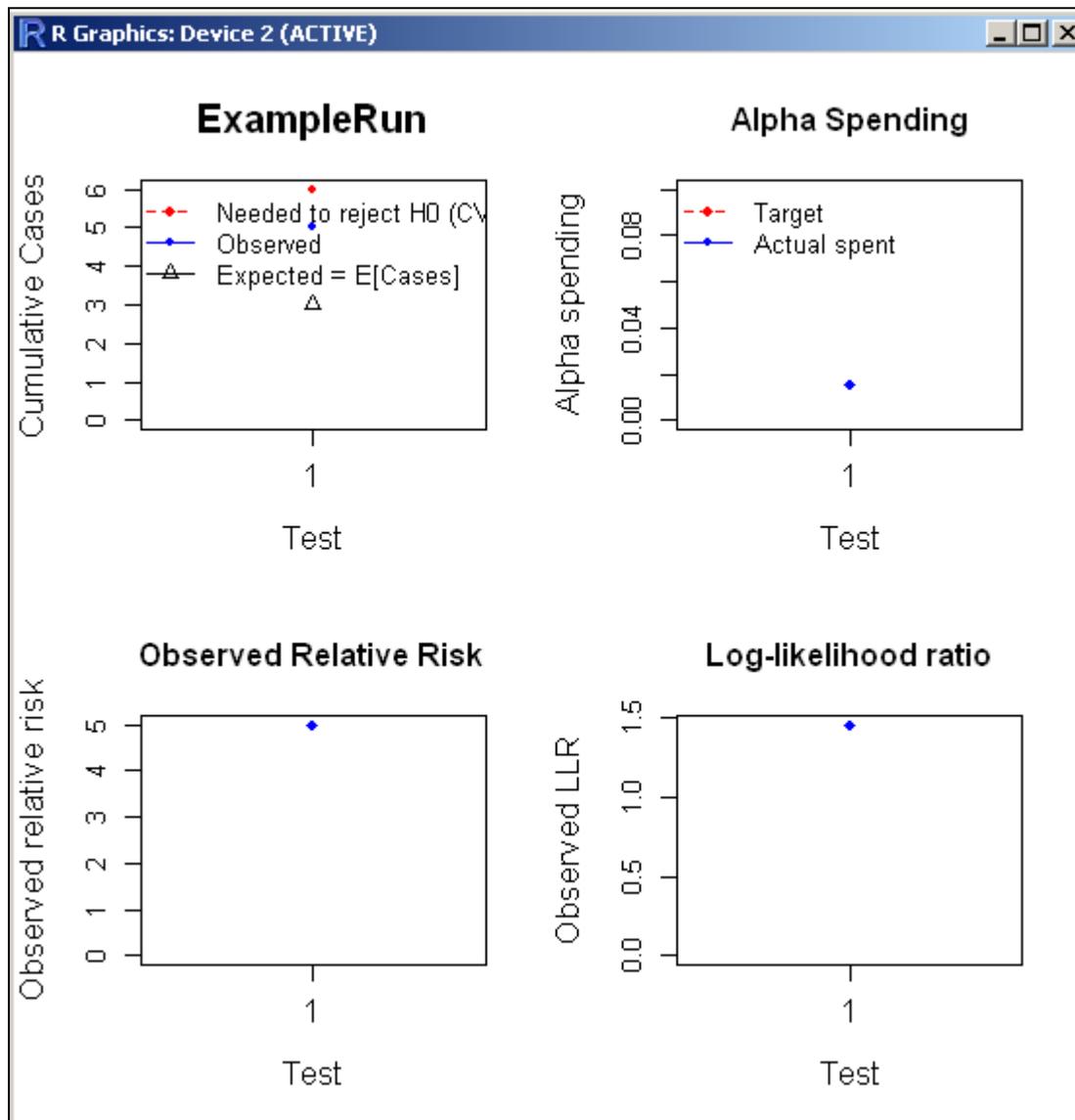
-----
=> Do not reject H0. Proceed to a new test as soon as you have more data.
-----

----- Cumulative ----- --alpha spent--
Test Cases Controls Cases Controls E[Cases] RR LLR target actual CV HO rejected
1 5 1 5 1 3 5 1.456 0.0156 0.0156 6 No
-----

Parameter settings: N= 50, alpha= 0.05, z= 1 and M= 3.
Analysis performed on Mon Jun 24 15:30:27 2013.
-----
>
>
>

```

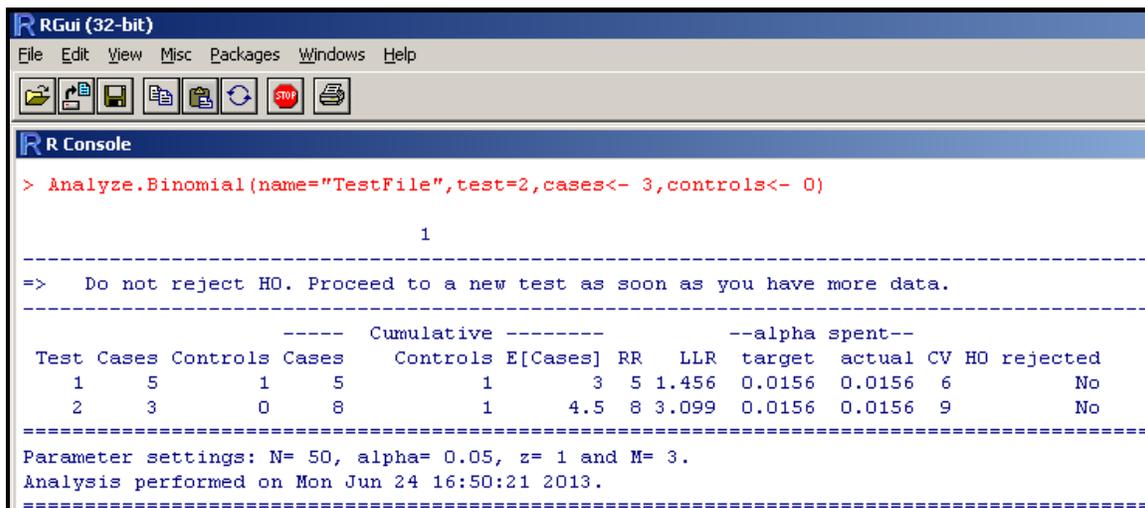
Figure 14.



## 2. Analyze.Binomial(name="TestFile",test=2,cases<- 3,controls<- 0)

Test 2 is explained in figures 15 and 16.

Figure 15.



```

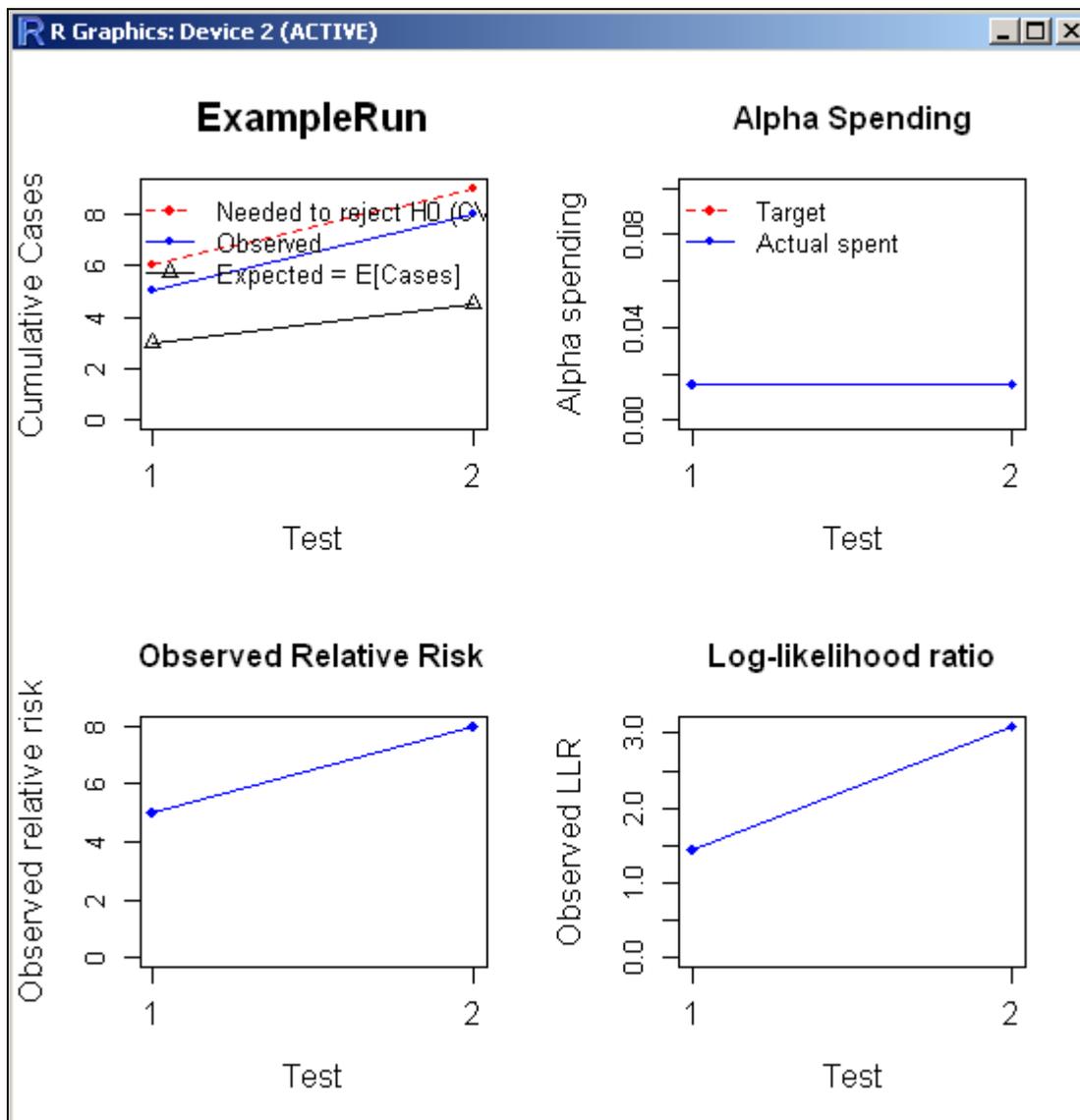
RGui (32-bit)
File Edit View Misc Packages Windows Help
[Icons]

R Console
> Analyze.Binomial(name="TestFile",test=2,cases<- 3,controls<- 0)

1
-----
=> Do not reject H0. Proceed to a new test as soon as you have more data.
-----
Test Cases Controls Cases Cumulative Controls E[Cases] RR LLR target actual CV HO rejected
1 5 1 5 1 3 5 1.456 0.0156 0.0156 6 No
2 3 0 8 1 4.5 8 3.099 0.0156 0.0156 9 No
-----
Parameter settings: N= 50, alpha= 0.05, z= 1 and M= 3.
Analysis performed on Mon Jun 24 16:50:21 2013.
-----

```

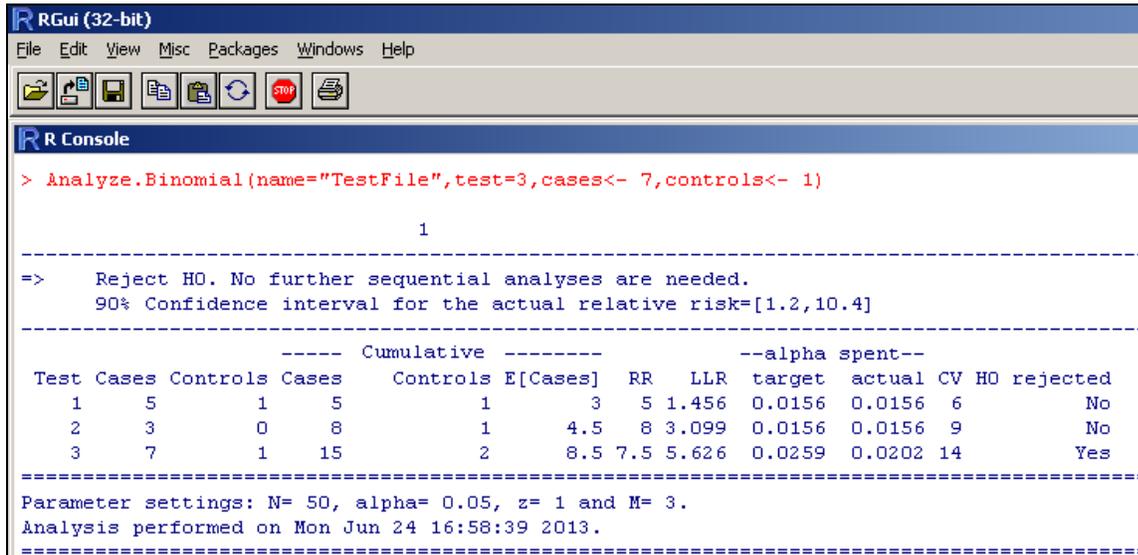
Figure 16.



### 3. Analyze.Binomial(name="TestFile",test=3,cases<- 7,controls<- 1)

Test 3 is explained in figures 17 and 18.

Figure 17.



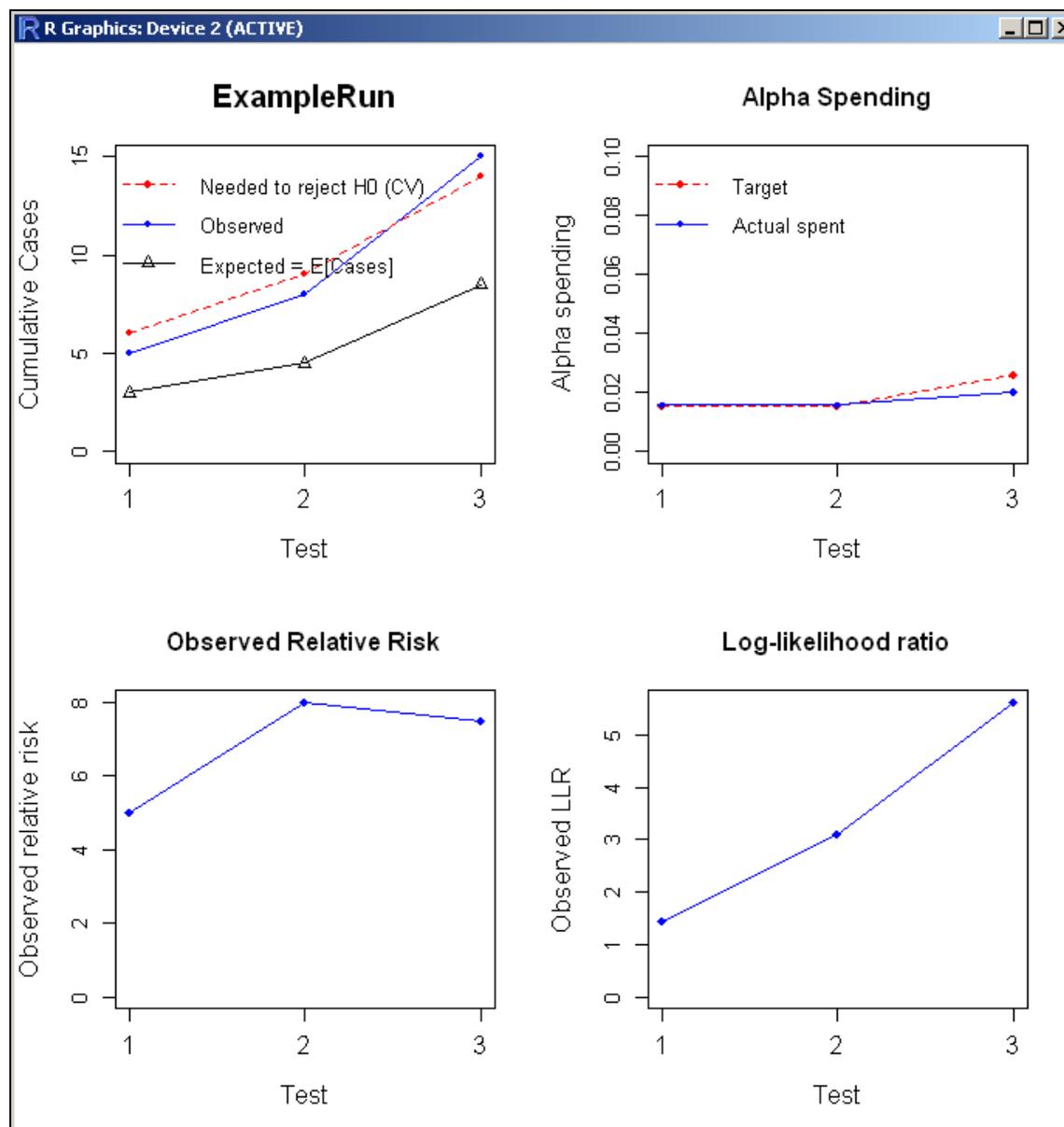
```

RGui (32-bit)
File Edit View Misc Packages Windows Help
[Icons]

R Console
> Analyze.Binomial(name="TestFile",test=3,cases<- 7,controls<- 1)

1
-----
=> Reject H0. No further sequential analyses are needed.
    90% Confidence interval for the actual relative risk=[1.2,10.4]
-----
      Test Cases Controls Cases      Cumulative      E[Cases] RR  LLR  --alpha spent--
      1     5     1     5      Controls      [Cases]  LLR  target  actual  CV  HO rejected
-----
      1     5     1     5         1         3  5  1.456  0.0156  0.0156  6      No
      2     3     0     8         1         4.5  8  3.099  0.0156  0.0156  9      No
      3     7     1    15         2         8.5  7.5  5.626  0.0259  0.0202  14     Yes
-----
Parameter settings: N= 50, alpha= 0.05, z= 1 and M= 3.
Analysis performed on Mon Jun 24 16:58:39 2013.
  
```

Figure 18.



## E. EXTENDED SURVEILLANCE

It is possible to continue analyzing the data even after the null hypothesis has been rejected or after the upper limit for surveillance (N) has been reached, even though the sequential analysis has formally ended. One might want to continue after rejection of the null in order to get a more stable estimate of the relative risk, for example. And, if the null hypothesis is not rejected by the time the upper limit (N) is reached, one might want to continue beyond the upper limit in order to gain additional assurance that the product is safe. However, it must be noted that no more alpha is spent, and the decision of whether to reject the null hypothesis or not will not be changed.

## **F. REPORT FORMAT**

A template for reporting the results of the sequential analyses starts on the next page. Currently, its production is manual, drawing from the outputs of the SAS data extraction programs and the R programs.

**Lisinopril-angioedema**

Self-controlled group sequential analysis

Risk interval: Days 1-28; control interval: Days -35 through -8 (where Day 0 is the first day of exposure)

Test no. and report date:

Last exposure date:

Cumulative exposed:

**Sequential analysis history and results**

| Tracking |                                | New events                                   |                                                    | Cumulative              |                                          |                                                |                       | Risk estimates |                 | Hypothesis-testing statistics and results |                       |                    |                                       |               |
|----------|--------------------------------|----------------------------------------------|----------------------------------------------------|-------------------------|------------------------------------------|------------------------------------------------|-----------------------|----------------|-----------------|-------------------------------------------|-----------------------|--------------------|---------------------------------------|---------------|
| Test #   | Most recent batch(es) included | No. of new events in risk interval ("cases") | No. of new events in control interval ("controls") | No. of exposed patients | No. of events in risk interval ("cases") | No. of events in control interval ("controls") | Expected no. of cases | Relative risk  | Risk difference | Log likelihood ratio test statistic       | Target alpha to spend | Actual alpha spent | No. of cases needed to reject H0 (CV) | H0 rejected ? |
| 1        |                                |                                              |                                                    |                         |                                          |                                                |                       |                |                 |                                           |                       |                    |                                       |               |
| 2        |                                |                                              |                                                    |                         |                                          |                                                |                       |                |                 |                                           |                       |                    |                                       |               |
| 3        |                                |                                              |                                                    |                         |                                          |                                                |                       |                |                 |                                           |                       |                    |                                       |               |
| 4        |                                |                                              |                                                    |                         |                                          |                                                |                       |                |                 |                                           |                       |                    |                                       |               |
| 5        |                                |                                              |                                                    |                         |                                          |                                                |                       |                |                 |                                           |                       |                    |                                       |               |
|          |                                |                                              |                                                    |                         |                                          |                                                |                       |                |                 |                                           |                       |                    |                                       |               |

Date of last analysis: 5/21/2013

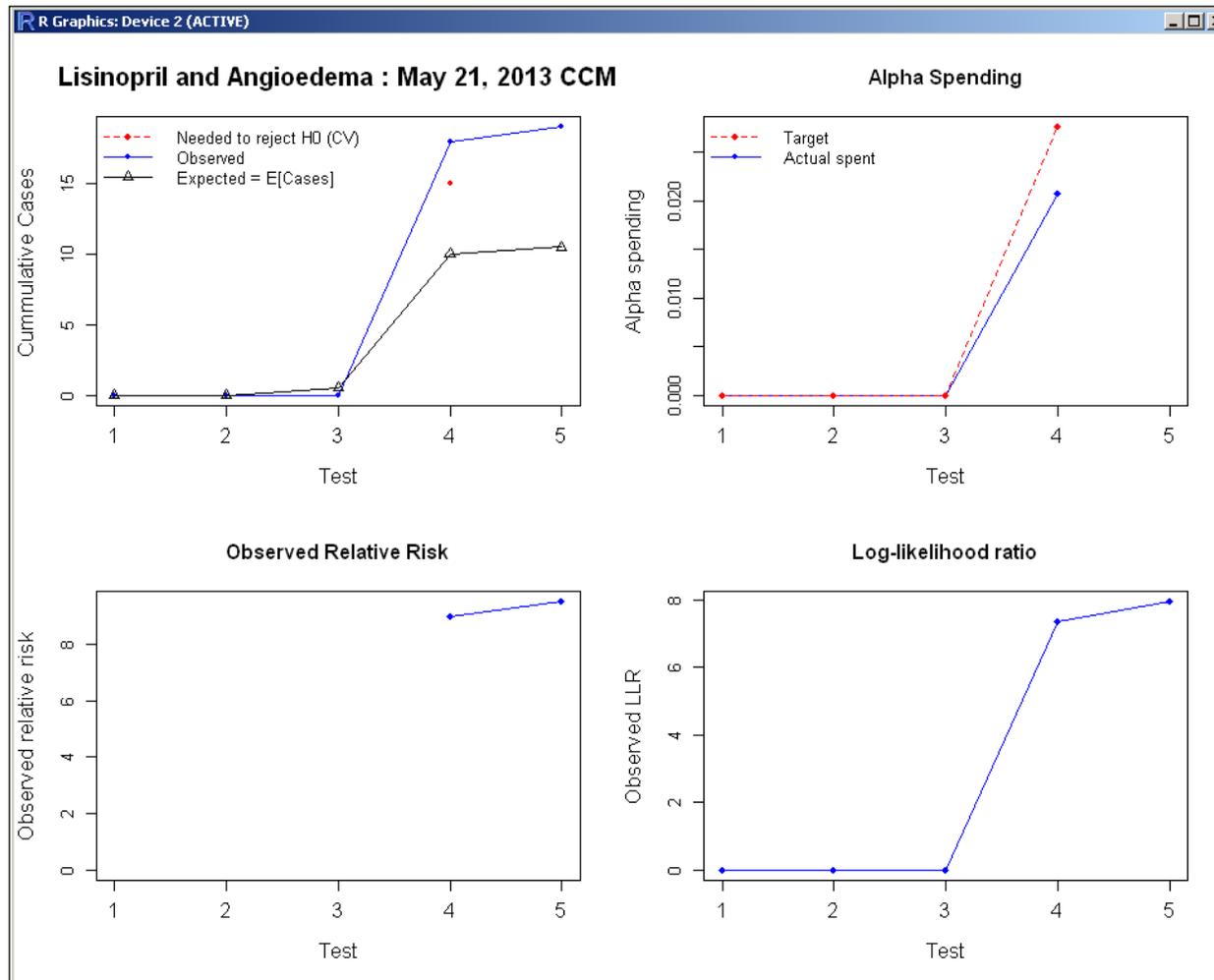
**Sequential analysis parameter settings**

|                                                               |      |
|---------------------------------------------------------------|------|
| Maximum total no. of events in risk and control intervals (N) | 30   |
| Alpha level                                                   | 0.05 |
| Ratio of risk interval to control interval (z=control:risk)   | 1:1  |
| Minimum total no. of events to signal (M)                     | 3    |

**Performance metrics**

|                                                     |     |     |     |     |     |    |
|-----------------------------------------------------|-----|-----|-----|-----|-----|----|
| Relative risk:                                      | 1.0 | 1.5 | 2.0 | 3.0 | 5.0 | 10 |
| Estimated statistical power                         |     |     |     |     |     |    |
| Expected total no. of events when H0 rejected       |     |     |     |     |     |    |
| Expected total no. of events at end of surveillance |     |     |     |     |     |    |

Graphical representation of subset of results and statistics



**Data extraction history**

| Batch # | Data partner | No. of exposed patients | No. of events in risk interval ("cases") | No. of events in control interval ("controls") | Earliest possible dispensing date in dataset | Latest possible dispensing date in dataset | Date dataset created |
|---------|--------------|-------------------------|------------------------------------------|------------------------------------------------|----------------------------------------------|--------------------------------------------|----------------------|
| 1       |              |                         |                                          |                                                |                                              |                                            |                      |
| 2       |              |                         |                                          |                                                |                                              |                                            |                      |
| 3       |              |                         |                                          |                                                |                                              |                                            |                      |
| 4       |              |                         |                                          |                                                |                                              |                                            |                      |
| 5       |              |                         |                                          |                                                |                                              |                                            |                      |
|         |              |                         |                                          |                                                |                                              |                                            |                      |

## V. APPENDIX 1

### A. MINI-SENTINEL WORKPLAN

Date submitted: June 18, 2013

#### 1. Project Name

Task Order 8: Foundational Elements – Statistical Methods Development

**Mini-Sentinel Request ID:** mini\_sentinel\_to08\_scd\_wp1\_b2 (Task Order 8 – Self Control Design Sequential Methods Workplan 1 – Beta 2)

#### 2. Overall Project Objective

The purpose of Mini-Sentinel Activity 4.10 is to build the capability for active surveillance for medical product safety for FDA through approaches that are flexible enough to be easily adapted for the study of a variety of products and outcomes. This data request creates analytic datasets that will be used to conduct sequential analysis using the R statistical package.

#### 3. MSOC Contact Information

Name: Jen Popovic

Institution: Mini-Sentinel Operations Center, Harvard Pilgrim Health Care Institute

Phone #: 617-509-9811

Email: jennifer\_popovic@harvardpilgrim.org

#### 4. Budget Item

Task Order 8 (Year 3 Foundational Elements)

#### 5. Instructions

Please follow the instructions in the header section of the main SAS program (mini\_sentinel\_to08\_scd\_wp1\_b2.sas). The sections in the header that will need user-input are:

Edit DPID and Site ID according to the table below,

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

Edit this section to reflect locations for the Libraries and Folders for Mini-Sentinel.

**\*\*\*For this run, please do NOT alter the value of “%LET exposure\_period\_cd=2;” statement at the very beginning of the header.\*\*\***

When done please:

1. zip/compress the MSOC output folder (file name should be “mini\_sentinel\_to08\_scd\_wp1\_b2.zip”)
2. upload it to the Mini-Sentinel secure portal (<https://portal.mini-sentinel.org>) in your site’s Private section (Private/\*\*DPID+SiteID\*\*/)
3. notify the MSOC

## 6. For KP Mini-Sentinel Programmers Only

Instead of uploading the results to the Mini-Sentinel secure portal, please place the SAS datasets and log files in a zip file and post them on to the Mini-Sentinel Content Area on the KP Secure File Transfer Web Site (<https://www.kpchr.org/securefiletransfer/apps/default.aspx>).

## 7. Workplan Timeline

Please complete by June 25, 2013.

## 8. Package Documents

1 main SAS Program (sasprograms subdirectory): mini\_sentinel\_to08\_scd\_wp1\_b2.sas

2 supporting SAS Programs files (inputfiles subdirectory):

mp4\_10.sas

ms\_episoderec.sas

10 lookup files (inputfiles subdirectory):

exposure\_periods.sas7bdat

event.sas7bdat

incevent.sas7bdat

incquery.sas7bdat

lkp\_for\_all\_codes.sas7bdat

age\_groups.sas7bdat

groups\_mapping.sas7bdat

event\_codes.sas7bdat

query.sas7bdat

query\_codes.sas7bdat

This workplan document and technical specification file:

mini\_sentinel\_to08\_scd\_wp1\_b2\_workplan.pdf

Timeframe for Data to be Included

April 1, 2013 – June 30, 2013

MSCDM Files Accessed

Demographic

Diagnosis

Dispensing

## Procedure

### EnrollmentOutput Files

This program generates: 1) 3 SAS datasets, 2) 1 log file, and 3) 1 lst file, to the msoc subdirectory. Please return all 3 files to MSOC.

3 SAS output datasets returned to MSOC:

```
&dpid.&siteid.&out_tbl_prefix._sum_evnt.sas7bdat  
&dpid.&siteid.&out_tbl_prefix._sum_exps_period.sas7bdat  
&dpid.&siteid.&out_tbl_prefix._signature.sas7bdat  
&dpid.&siteid.&out_tbl_prefix.sum_exps_period_fin
```

1 LOG file returned to MSOC:

```
&dpid.&siteid.&out_rpt_prefix..log
```

1 LST file returned to MSOC:

```
&dpid.&siteid.&out_rpt_prefix._valid_rpt.lst
```

This program also creates 4 SAS datasets that are saved into the dplocal subdirectory. These files are to remain with the Data Partner and retained for troubleshooting.

4 SAS output dataset to remain at the Data Partner in the dplocal subdirectory:

```
&dpid.&siteid.&out_tbl_prefix._pat_evnt  
&dpid.&siteid.&out_tbl_prefix._sum_exps_evnt  
&dpid.&siteid.&out_tbl_prefix._pat_exps_excl  
&dpid.&siteid.&out_tbl_prefix._pat_evnt_excl
```