

A Macro for Generating the Adverse Events Summary for ClinicalTrials.gov

Andrew Moseby and Maya Barton, Rho, Inc.®

ABSTRACT

In the clinical trials industry, the website ClinicalTrials.gov serves as a publicly accessible outlet of information on trial outcomes. Once a federally sponsored trial concludes, the study team is responsible for publishing adverse events data to ClinicalTrials.gov. These data must be delivered in a very particular structure, uniform across all studies. This uniformity allows for generalization via a SAS® macro. The purpose of this paper is to introduce such a macro. Macro variables offer flexibility to account for study-specific characteristics such as number of treatment groups and variable naming conventions. Input datasets may be modified in the work directory in cases where the macro variables do not accommodate a particularly non-standard input dataset.

INTRODUCTION

The ClinicalTrials.gov deliverable is due within 12 months of last patient last visit (LPLV). In addition to adverse events (AEs), there are several aspects of a clinical trial that must be reported on, including participant flow, baseline characteristics, and outcome measures. This paper specifically addresses the AE summary upload format and explains a macro we have developed and successfully used to generate the AE summary part of the deliverable. The macro allows for variation in study protocol, such as number of treatment groups, as well as different file paths, variable names, and conventions. This macro generates a SAS dataset ready to be converted to XML for upload.

CLINICALTRIALS.GOV ADVERSE EVENTS SUMMARY FORMAT

The dataset output by the macro separates adverse events by serious adverse event (SAE; true or false), and then sorts by system organ class, preferred term, and treatment group. Each type of AE according to this sorting has 3 observations associated with it, each containing a statistic required by ClinicalTrials.gov. The statistic is indicated by the variable CAT_VAR. A value of "numEvents" corresponds to the number of AEs experienced. "numSubjectsAffected" gives the number of distinct subjects who experienced those events, and "numSubjects" gives the total number of subjects in the treatment group. There are also 2 extra rows per treatment group for both SAEs and non-SAEs, denoted by ORGANSYSTEMNAME = "Total AE", which give totals of AEs and subjects experiencing AEs across all system organ classes and preferred terms.

	SAE	Assessment Type	SourceVocabulary	OrganSystemName	Term	ReportingGroupID	cat_var	Counts
217	False	Systematic Assessment	MedDRA 11.1	Skin and subcutaneous tis...	Urticaria	ReportedEvents-InterventionGroup.1	numEvents	5
218	False	Systematic Assessment	MedDRA 11.1	Skin and subcutaneous tis...	Urticaria	ReportedEvents-InterventionGroup.1	numSubjectsAffected	5
219	False	Systematic Assessment	MedDRA 11.1	Skin and subcutaneous tis...	Urticaria	ReportedEvents-InterventionGroup.1	numSubjects	235
220	False	Systematic Assessment	MedDRA 11.1	Skin and subcutaneous tis...	Urticaria	ReportedEvents-InterventionGroup.2	numEvents	15
221	False	Systematic Assessment	MedDRA 11.1	Skin and subcutaneous tis...	Urticaria	ReportedEvents-InterventionGroup.2	numSubjectsAffected	13
222	False	Systematic Assessment	MedDRA 11.1	Skin and subcutaneous tis...	Urticaria	ReportedEvents-InterventionGroup.2	numSubjects	228
223	False	Systematic Assessment	MedDRA 11.1	Skin and subcutaneous tis...	Urticaria	ReportedEvents-InterventionGroup.3	numEvents	5
224	False	Systematic Assessment	MedDRA 11.1	Skin and subcutaneous tis...	Urticaria	ReportedEvents-InterventionGroup.3	numSubjectsAffected	5
225	False	Systematic Assessment	MedDRA 11.1	Skin and subcutaneous tis...	Urticaria	ReportedEvents-InterventionGroup.3	numSubjects	47
226	False	Systematic Assessment	MedDRA 11.1	Skin and subcutaneous tis...	Urticaria	ReportedEvents-InterventionGroup.4	numEvents	3
227	False	Systematic Assessment	MedDRA 11.1	Skin and subcutaneous tis...	Urticaria	ReportedEvents-InterventionGroup.4	numSubjectsAffected	2
228	False	Systematic Assessment	MedDRA 11.1	Skin and subcutaneous tis...	Urticaria	ReportedEvents-InterventionGroup.4	numSubjects	46
229	False	Systematic Assessment	MedDRA 11.1	Total AE	Total AE	ReportedEvents-InterventionGroup.1	numSubjectsEvents	206
230	False	Systematic Assessment	MedDRA 11.1	Total AE	Total AE	ReportedEvents-InterventionGroup.1	partAtRiskEvents	235
231	False	Systematic Assessment	MedDRA 11.1	Total AE	Total AE	ReportedEvents-InterventionGroup.2	numSubjectsEvents	182
232	False	Systematic Assessment	MedDRA 11.1	Total AE	Total AE	ReportedEvents-InterventionGroup.2	partAtRiskEvents	228
233	False	Systematic Assessment	MedDRA 11.1	Total AE	Total AE	ReportedEvents-InterventionGroup.3	numSubjectsEvents	45
234	False	Systematic Assessment	MedDRA 11.1	Total AE	Total AE	ReportedEvents-InterventionGroup.3	partAtRiskEvents	47
235	False	Systematic Assessment	MedDRA 11.1	Total AE	Total AE	ReportedEvents-InterventionGroup.4	numSubjectsEvents	38
236	False	Systematic Assessment	MedDRA 11.1	Total AE	Total AE	ReportedEvents-InterventionGroup.4	partAtRiskEvents	46
237	True	Systematic Assessment	MedDRA 11.1	Immune system disorders	Anaphylactic reacti...	ReportedEvents-InterventionGroup.1	numEvents	3
238	True	Systematic Assessment	MedDRA 11.1	Immune system disorders	Anaphylactic reacti...	ReportedEvents-InterventionGroup.1	numSubjectsAffected	2
239	True	Systematic Assessment	MedDRA 11.1	Immune system disorders	Anaphylactic reacti...	ReportedEvents-InterventionGroup.1	numSubjects	235
240	True	Systematic Assessment	MedDRA 11.1	Immune system disorders	Anaphylactic reacti...	ReportedEvents-InterventionGroup.2	numEvents	1
241	True	Systematic Assessment	MedDRA 11.1	Immune system disorders	Anaphylactic reacti...	ReportedEvents-InterventionGroup.2	numSubjectsAffected	1

Figure 1. Output generated by the macro on data from the Persistence of Oral Tolerance to Peanut (LEAP-On) study. The ReportingGroupID variable corresponds to treatment group, and the cat_var variable describes what the Counts variable means.

VARIANCE FROM STUDY TO STUDY

The macro takes into account several aspects of studies which may differ. It can handle any number of treatment groups, the presence or absence of a randomization dataset, and different variable naming conventions. For example, the ID variable in the AE analysis dataset may be named USUBJID, and the ID variable in the subject-level analysis dataset may be named SUBJID. The macro can take the 2-3 datasets (AE, subject-level, and optionally randomization) from any location on the network. If these considerations still do not accommodate a particularly non-standard dataset, then modifications can be made in the work directory in the program containing the macro call.

TYPICAL AE CODED DATASET

Figure 2 shows an example of a standard AE coded dataset. A clinical dataset is converted into a coded dataset by using a predefined coding dictionary, such as MedDRA, to standardize “verbatim terms” typically entered in a text field for adverse events to a discrete set of “preferred terms” and “System Organ Class”. This is the AE dataset that is to be used as input data for the macro.

	USUBJID	STRATUM	TRT	AESER	AESOC	AEPT
1	1	1	2	No	Infections and infestations	Rhinitis
2	2	1	1	No	Infections and infestations	Impetigo
3	3	1	2	No	Gastrointestinal disorders	Vomiting
4	4	1	2	No	General disorders and administration site c...	Pyrexia
5	5	1	1	No	Infections and infestations	Respiratory tract infection
6	6	1	1	No	Gastrointestinal disorders	Diarrhoea
7	7	1	1	No	Infections and infestations	Respiratory tract infection
8	8	1	1	No	Ear and labyrinth disorders	Motion sickness
9	9	1	1	No	Eye disorders	Conjunctivitis
10	10	1	1	No	Gastrointestinal disorders	Abdominal pain

Figure 2. AE coded dataset from LEAP-On (modified). This study used 2 strata and 2 treatment groups, the 4 combinations of which constituting the reporting groups. The USUBJID variable values have been changed from the data used in the study.

This AE dataset is one record per AE. Coded AE datasets typically contain many more variables, but the ones shown in figure 2 are the ones that are relevant for ClinicalTrials.gov upload. The ID variable USUBJID attributes the AE to a particular subject. The LEAP-On study used two variables, STRATUM and TRT, to specify groups for analysis. AESER, AESOC, and AEPT give the SAE status, system organ class, and preferred term of the AE respectively.

CALLING THE MACRO

The macro, named %CT_Upload, can be called with code similar to the following:

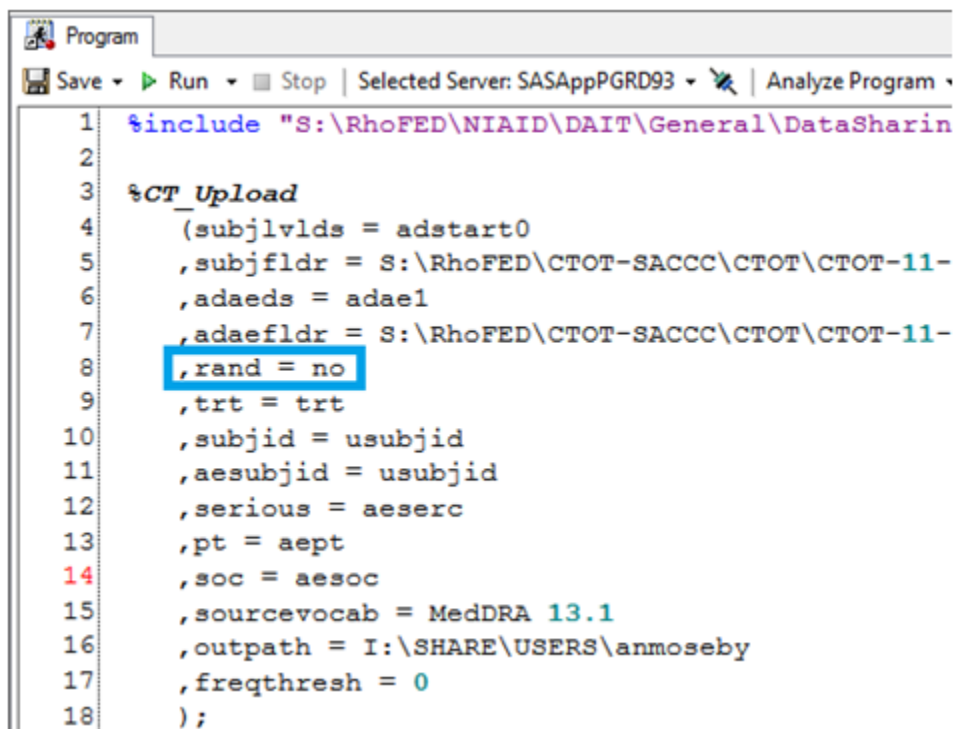
```
%include "&CTpath\CT_Upload.sas";

%CT_Upload
(
  subjlvlds = adstart0
  , subjfldr = S:\RhoFED\CTOT-SACCC\CTOT\CTOT-11-Sayegh\Stats\Data\Derive
  , adaeds = adael
  , adaefldr = S:\RhoFED\CTOT-SACCC\CTOT\CTOT-11-Sayegh\Stats\Data\Derive
  , randds = patient_data
  , randfldr = S:\RhoFED\CTOT-SACCC\CTOT\CTOT-11-Sayegh\Stats\Data\RhoRAND
  , rand = rand
  , trt = trt
  , subjid = usubjid
  , aesubjid = usubjid
  , randsubjid = id
  , serious = aeser
  , pt = aept
  , soc = aesoc
  , freqthresh = 0.05
  , sourcevocab = MedDRA 11.1
  , outpath = I:\SHARE\USERS\anmoseby
);
```

- SUBJLVLD: the name of the subject-level dataset containing all participants to be reported on, whether or not they contribute any AEs.
- SUBJFLDR: the complete path to the location of the subject-level dataset.
- ADAEDS: the name of the coded AE dataset.
- ADAEFLDR: the complete path to the location of the coded AE dataset.
- RANDDS (optional): the name of the randomization dataset.
- RANDFLDR (optional): the complete path to the location of the randomization dataset.
- RAND: the name of the randomization variable. The randomization variable may exist in either the subject-level dataset or the randomization dataset.
- TRT: the name of the treatment group variable.
- SUBJID: the name of the ID variable in the subject-level dataset.
- AESUBJID: the name of the ID variable in the AE coded dataset.
- RANDSUBJID: the name of the ID variable in the randomization dataset.
- SERIOUS: the name of the SAE flag in the AE coded dataset.
- PT: the name of the preferred term variable in the AE coded dataset.
- SOC: the name of the system organ class variable in the AE coded dataset.
- FREQTHRESH: ClinicalTrials.gov typically only requires non-SAEs to be uploaded for types of events which occur with a certain frequency in any of the treatment groups. By default, this frequency threshold is set at 5% (.05).
- SOURCEVOCAB: The source vocabulary used for coding AEs.
- OUTPATH: the complete path to the desired location of the output.

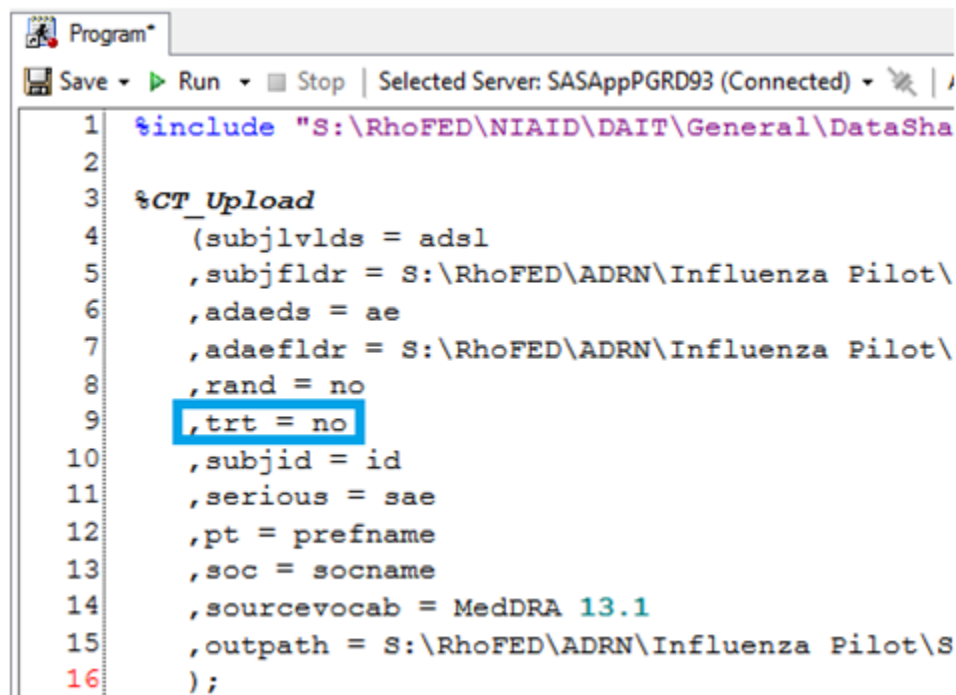
Once the macro is called, a dataset named "aefreq.sas7bdat" will be placed in the folder specified in OUTPATH. This dataset will resemble the one shown in Figure 1. If the user entered invalid values for the macro variables, the macro will terminate and give an error message in the log.

ADDITIONAL CAPABILITIES OF THE MACRO



```
1 %include "S:\RhoFED\NIAID\DAIT\General\DataSharin
2
3 %CT_Upload
4     (subjlvlids = adstart0
5     ,subjfldr = S:\RhoFED\CTOT-SACCC\CTOT\CTOT-11-
6     ,adaeds = adae1
7     ,adaefldr = S:\RhoFED\CTOT-SACCC\CTOT\CTOT-11-
8     ,rand = no
9     ,trt = trt
10    ,subjid = usubjid
11    ,aesubjid = usubjid
12    ,serious = aeserc
13    ,pt = aept
14    ,soc = aesoc
15    ,sourcevocab = MedDRA 13.1
16    ,outpath = I:\SHARE\USERS\anmoseby
17    ,freqthresh = 0
18    );
```

Figure 3. Setting RAND = NO prevents the macro from subsetting the population to randomized subjects. The user may instead choose to use RAND to subset on a positive value of another variable, such as a safety flag.



```
1 %include "S:\RhoFED\NIAID\DAIT\General\DataSha
2
3 %CT_Upload
4     (subjlvlids = adsl
5     ,subjfldr = S:\RhoFED\ADRN\Influenza Pilot\
6     ,adaeds = ae
7     ,adaefldr = S:\RhoFED\ADRN\Influenza Pilot\
8     ,rand = no
9     ,trt = no
10    ,subjid = id
11    ,serious = sae
12    ,pt = pefname
13    ,soc = socname
14    ,sourcevocab = MedDRA 13.1
15    ,outpath = S:\RhoFED\ADRN\Influenza Pilot\S
16    );
```

Figure 4. Set TRT = NO if there is only one treatment group.

```

1  libname LeapOn 'S:\RhoFED\ITN\AsthmaAllergy\ITN049AD_1';
2
3  %include 'S:\RhoFED\NIAID\DAIT\General\DataSharing\CT_Upload.sas';
4
5  data ae;
6      set LeapOn.adae1;
7      trt_new = compbl(stratum||trt);
8  run;
9
10 %CT_Upload
11     (subjlvl0 = adstart0
12     ,subjfldr = S:\RhoFED\ITN\AsthmaAllergy\ITN049AD_1
13     ,adaeds = ae
14     ,adaefldr = work
15     ,rand = safety
16     ,trt = trt_new
17     ,subjid = usubjid
18     ,aesubjid = usubjid
19     ,serious = aeser
20     ,pt = aept
21     ,soc = aesoc
22     ,sourcevocab = MedDRA 11.1
23     ,outpath = S:\RhoFED\ITN\AsthmaAllergy\ITN049AD_1
24     ,freqthresh = .05
25 );

```

Figure 5. If necessary, any of the input datasets can be modified in the work directory before being used by the macro. Setting SUBJFLDR, ADAEFLDR and/or RANDFLDR = WORK causes the macro to use the corresponding dataset(s) in the work directory.

COMPLETING THE AE SUBMISSION TO CLINICALTRIALS.GOV

All that needs to be done with the output dataset to make it ready for upload is converting to XML. At the time this paper was written, the authors plan to implement this as a built-in feature of the macro.

CONCLUSION

In this paper, we have provided a macro that can be used to facilitate the generation of an AE summary dataset to be uploaded to ClinicalTrials.gov. The macro takes into account study-specific variations such as number of treatment groups, presence or absence of a randomization dataset, and variable names, and outputs a dataset that just needs to be converted to XML before uploading to ClinicalTrials.gov. The macro itself can be found on Rho's GitHub page:

https://github.com/RhoInc/sas-ct-gov-aes/blob/master/CT_Upload.sas

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Andrew Moseby
Rho, Inc.
Andrew_Moseby@rhoworld.com

Maya Barton
Rho, Inc.
Maya_Barton@rhoworld.com