Paper 2766-2018

Automation of Clinical Trial Result Posting to ClinicalTrials.gov and EudraCT

Sherry Meeh, Janssen - a pharmaceutical company of Johnson and Johnson

ABSTRACT

According to US Food and Drug Administration Amendments Act (FDAAA) and European Medicines Agency (EMA) regulations, companies are now required to disclose their applicable clinical trial's aggregated results to US ClinicalTrials.gov and to EMA European Clinical Trials Database (EudraCT) public websites: ClinicalTrials.gov (for US), and eudract.ema.europa.eu (for EMA). Currently in many pharmaceutical companies, the clinical trial aggregated results are prepared manually by a designated group based on the clinical study report. The manual process is time consuming and error-prone, with iterative back-and-forth steps and a lot of reviews among different groups. Therefore, an innovative and automated process is needed to proactively streamline the result posting process, improve the accuracy, efficiency, and consistency of the result posting, and reduce manual entry errors. At Johnson & Johnson, a biostatics programming group developed an innovative automated process to proactively streamline the result posting process, improve the accuracy, efficiency, and consistency of the result posting, and reduce manual entry errors by using SAS®.

This paper will provide data flow of automation process and automation method to create a required XML file from clinical trial SAS datasets.

INTRODUCTION

On September 27, 2007, U.S. Public Law 110-85 (Food and Drug Administration Amendments Act of 2007, or FDAAA), Title VIII, Section 801 mandates that a "responsible party" (i.e., the sponsor or designated principal investigator) register and report results of certain "applicable clinical trials", and "applicable clinical trials" must be registered in ClinicalTrials.gov. A clinical trials must be registered if it is a Phase 2 to 4 trial conducted for a product that must be approved in the United States (per policy), and has been initiated or ongoing as of 27-Sept-07 per FDAAA (2007). Starting December 26, 2007, at the time of submission of an application to FDA shall be accompanied by a certification, where available, such certification shall include the appropriate National Clinical Trial control numbers (NCT numbers). ClinicalTrials.gov uses a web based data entry system called the Protocol Registration System (PRS) for "responsible party" to register the clinical trials and submit result of certain "applicable clinical trials". Once the information is posted at ClinicalTrials.gov, the public can find registered clinical trial Information at http://www.clinicaltrials.gov. In general, the results disclosure at ClinicalTrials.gov registry system for applicable clinical trials must be posted with required information within 1 year of "primary completion date". If the product or indication is not yet marketed, or a blind has not yet been removed at this time, then a delayed submission must be filed.

On January 18, 2017, the final rule for Clinical Trials Registry and results Information Submission (42CFR Part 11) became effective. It requires that sponsors need to: post results for all applicable trials including for non-licensed products; disclose the full protocol, statistical analysis plan and all amendments (with redaction, as required) when results are posted. Here is the summary of how to apply Final Rules (i.e., effective 18Jan2017) and Statute (i.e., requirements prior to 18Jan2017):

(1) For Registration information: determined by Study Start Date

- Study Start Date on or after January 18, 2017: FINAL RULE
- Study Start before January 18, 2017: STATUTE
  - Study Start Date after September 27, 2007 but before January 18, 2017
  - Study Start Date before September 27, 2007, with Primary Completion Date after December 26, 2007 (i.e., ongoing study)

(2) For Reporting results: determined by Primary Completion Date

- Primary Completion Date on or after January 18, 2017: FINAL RULE
- Primary Completion Date before January 18, 2017: STATUTE

Realistically, this means that we will submit protocols under the new regime in January 2017, but results from January 2018 onwards by using FINAL RULE in ClinicalTrials.gov.

The requirements of result disclosure of ClinicalTrials.gov can be found in data elements definitions defined by ClinicalTrials.gov under:https://prsinfo.clinicaltrials.gov/results_definitions.html, and the requirements contains information of result disclosure required:

- Results Point of Contact
- Certain Agreements
- **Participant Flow**
- **Baseline Characteristics**
- Outcome Measures
- Overall Limitations and Caveats
- **Adverse Events**

Similar to US requirements to post clinical trial aggregated result in public domain, on July 21st 2014, the European Medicines Agency (EMA) has mandated the aggregated results disclosure for all interventional clinical trials conducted in the European Union (EU) and the European Economic Area (EEA), and all Pediatric studies irrespective of location. For the retrospective remediation, trials completed after 1 May2004 are in scope for results disclosure. These trial results are required to be uploaded to EudraCT database. EudraCT (European Union Drug Regulating Authorities Clinical Trials) is the European Clinical Trials Database of all interventional clinical trials of medical products commencing in the European Union from 1 May 2004 onwards. The EudraCT database has been established in accordance with Directive 2001/20/EC. Based on applicable business rules within EudraCT, Information of clinical trials in EudraCT database is posted to public via a website of EU Clinical Trials Register: www.clinicaltrialsregister.eu.

Per EudraCT, modalities of result disclosure are: Summary Attachments; Full Data Set. The “Summary Attachments” refers to CSR synopsis; and the “Full Data Set” refers to summary tables containing the following information:

- **Trial Information**
- **Subject Disposition**
- **Baseline Characteristics**
- End Points
- **Adverse Events**
- More Information

Detailed requirements of result disclosure can be found under: https://eudract.ema.europa.eu/result.html.

According to requirements from ClinicalTrials.gov and EudraCT, the clinical trial aggregated results contain, but not limit to, a set of defined summaries for:

- **Trial information**
- **Participant Flow (or Subject Disposition)**
  - total number of subjects completed / non-completed trial
  - reasons of discontinuations

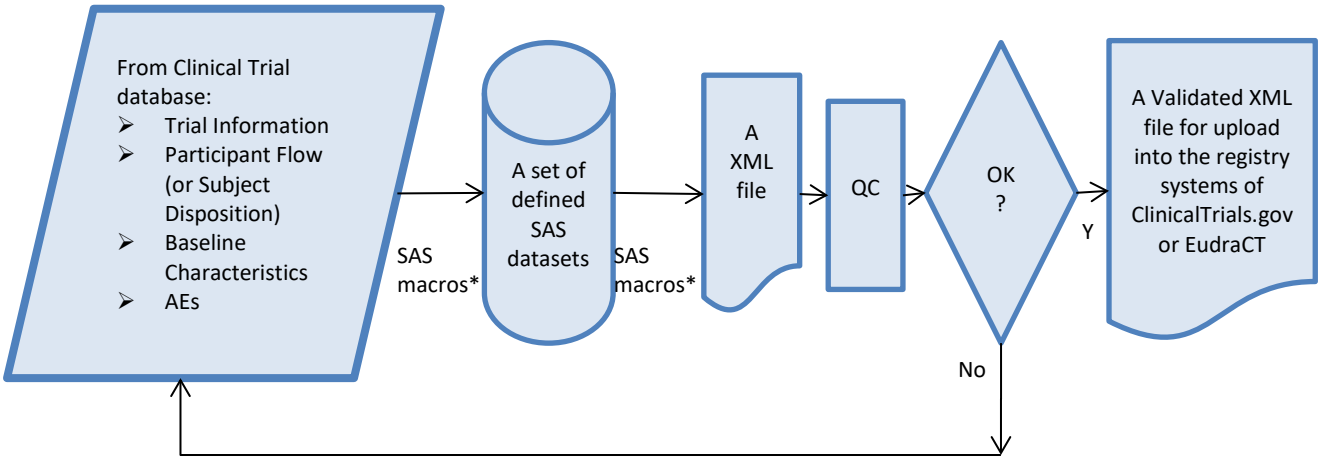
- **Baseline and Demographic Characteristics**
  - age
  - gender
  - other baseline characteristics.
- **Adverse Events**
  - Serious adverse events
  - Frequent adverse events (Non-serious adverse events with maximum 5% threshold)

These aggregated results can be obtained from an existing clinical trial database collected during the clinical trial and is also being used for clinical trial study report. Both ClinicalTrials.gov and EudraCT registry systems provide a partial upload option that allows companies to upload a XML file to ClinicalTrials.gov or EudraCT registry system based on the registry systems' requirements/schemas. Combining these two facts, we can tap into the existing clinical trial database, create a XML file that can be directly loaded into ClinicalTrials.gov or EudraCT registry system according to the registry systems' schemas, therefore automate the generation and posting of the trial results process to stream line the process to improve the accuracy, efficiency and consistency of the result disclosure across all clinical trials.

AUTOMATION METHOD

The automation process begins with a set of clinical trial SAS datasets used by many pharmaceutical companies, and ends with a validated XML file according to the schemas provided by the registry systems of ClinicalTrials.gov and EudraCT. SAS to XML automation are done through SAS macros. The following Figure 1 is the data flow of the automation process:

Figure 1: Data Flow of the Automation Process



*SAS macros are developed in two sets of macros: one set is based on the requirement of ClinicalTrials.gov; the other set is based on the requirement of EudraCT.

We defined a set of SAS datasets (please see table 1 below for an example of defined SAS dataset for Participant flow) based on the requirement of Data Element Definitions and the schemas from the registry system of ClinicalTrial.gov or EudraCT. We developed a set of SAS macros to: (1) create the set of defined SAS datasets from a clinical trial SAS database; (2) create a XML file from the defined SAS dataset according to the required schemas. We also developed a QC tool to validate the XML file according to the required schemas and contents of the XML file. Once a XML file is validated, it is ready for upload directly into the registry systems for posting in ClincialTrials.gov or EudraCT.

The set of SAS macros are designed to fit different types of clinical trial designs and various types of clinical trial data structure, including SDTM, and ADaM. The SAS macros are portable and stand-alone macros that can be adapted and integrated into any SAS system.

DEFINED SAS DATASETS

The following is an example of defined SAS datasets we defined based on ClinicalTrials.gov:

Table 1: Defined SAS dataset for Participant Flow of ClinicalTrials.gov:

Variable name	Variable length	Additional comments
PERIOD	$40	Discrete stages of a clinical trial during which numbers of participants at specific significant events or points of time are reported. If only one period, use "Overall Study".
PERIODNUM		Sorting order for PERIOD
REPORTINGGROUP	$62	Title of treatment arm. Minimum length is 4
REPORTINGGROUPID	$40	Code for REPORTINGGROUP. It contains value of: "ParticipantFlow-ParticipantFlowGroup.*x*" with x numerical code from 1-number of reporting groups
TOTAL		Total number of subjects in REPORTINGGROUP in a given period.
REASONTYPE	$30	Withdrawal reason. It contains the value of standard reasons that is defined by clinicaltrials.gov. The values are: Adverse Event Death Lack of Efficacy Lost to Follow-up Physician Decision Pregnancy Protocol Violation Withdrawal by Subject Other
OTHERREASONNAME	$40	Only available for non-standard reason when REASONTYPE="Other".
COUNT		Number of subjects with withdrawal reason per REPORTINGGROUP in a given period.
CRPF_DSGRP	$40	Working variable for SAS to XML process. Values are: startedMilestone completedMilestone dropwithdrawreason
SEQORDER		Display sequence order in output

Good to know

Standard discontinuation reasons per clinicaltrials.gov are

- Adverse Event

- Death
- Lack of Efficacy
- Lost to Follow-up
- Physician Decision
- Pregnancy
- Protocol Violation
- Withdrawal by Subject
- Other

Only these (case sensitive) are accepted in the xml element reasonType. In case other reasons are to be reported reasonType should be “Other” and the reason itself should be reported in otherReasonName.

Similarly, we defined the sets of defined SAS datasets for Trial Information, Participant Flow (or Subject Disposition), Baseline and Demographic characteristics, and Adverse Events for ClinicalTrials.gov and EudraCT.

SAS MACROS TO CREATE DEFINED SAS DATASETS

We developed a set of SAS macros to read in our SAS clinical datasets and then create the defined SAS dataset that contains aggregated summary result according to ClinialTrials.gov and EudraCT requirements. The following is an example of SAS codes to create of a defined SAS dataset called CRPF by calling our internal developed SAS macro %CRPF to create a defined SAS dataset for participant flow result of ClinicalTrials.gov:

```
proc format;
     invalue armordf   /* output display order based on the input data */
     'TRT1'    = 1
     'TRT2'    = 2
     'Placebo' = 3
     ;
    value $armf     /* format for treatment arm that will be displayed in
                       output based on the input data */
     'TRT1'    = 'Treatment 1'
     'TRT2'    = 'Treatment 2'
     'Placebo' = 'Placebo'
     ;
    value periodf
     1='Overall Study'
     ;
    invalue wdordf       /* display order of withdraw reasons */
    'ADVERSE EVENT'          = 1
    'DEATH'                  = 2
    'LACK OF EFFICACY'       = 3
    'LOST TO FOLLOW-UP'      = 4
    'PHYSICIAN DECISION'     = 5
    'PREGNANCY'              = 6
    'OTHER: EXCLUSION #9'    = 7
    'SUBJECT CHOICE'         = 8
    "SUBJECT NON-COMPLIANT"           = 91
    "SUBJECT REACHED A CERTAIN POINT" = 92
    "SPONSOR'S DECISION"              = 93
    OTHER                             = 99
```

```
     ;
/* input left side of format per study, please do not change the right side
of format that is highlighted in yellow !!! */
     value $withdrf
     'ADVERSE EVENT'       = 'Adverse Event'
     'DEATH'               = 'Death'
     'LACK OF EFFICACY'    = 'Lack of Efficacy'
     'LOST TO FOLLOW-UP'   = 'Lost to Follow-up'
     'PHYSICIAN DECISION'  = 'Physician Decision'
     'PREGNANCY'           = 'Pregnancy'
     'OTHER: EXCLUSION #9' = 'Protocol Violation'
     'SUBJECT CHOICE'      = 'Withdrawal by Subject'
      OTHER                = 'Other'
     ;
    run;


data crds;
  set a_in.adds;
  length xreasons $40 xstart $3;
  where saffl = 'Y'and dscat='DISPOSITION EVENT' and DSSCAT='TRIAL';
  xstart='YES';
  if dsdecod='COMPLETED' then xcomp='YES';
  else do;
    xcomp='NO';
    if compress(dsdecod) in (' ') then dedecod='OTHER';
    xreasons=dsdecod;
    xorder=input(dsdecod, wdordf.);
    /* to display in mixed-case */
    if xorder >90 then do;
       %crfcase(vars=xreasons, case=mixed);
    end;
  end;
run;

%crpf(indsn   =crds,     /* input data, one record per subject per
                            participant flow period */
      outlib  =a_out,    /* libname for output datasets */
      subj    =usubjid,  /* input var. for unique subject ID */
      arm     =trt01p,   /* arm/treatment group variable */
      armf    =$armf.,   /* format for treatment group */
      armordf =armordf., /* format for display order of treatment group */
      arm_n   =3,        /* number of arms */
      prdnum  =1,        /* input var.for period (num.),default: &period=1 */
      period  =periodf., /* format for period var.,default:1=Overall study */
      started =xstart,   /* input var. for start population: YES/NO */
      comp    =xcomp,    /* input var. for completed population: YES/NO */
      withdraw=xreasons, /* input var. for withdraw reasons */
      withdrf =$withdrf.,/* format for withdraw variable */
      wdordv  =xorder    /* input var. of display order for withdraw reasons */
      );
run;
```

Similarly, we developed a set of macros to create a set of defined SAS datasets that contain aggregated summary result for Trial Information, Participant Flow /Subject Disposition, Baseline and Demographic characteristics, and Adverse Events for ClinicalTrials.gov and EudraCT separately.

SAS MACROS TO CREATE A REQUIRED XML FILE FROM DEFINED SAS DATASETS

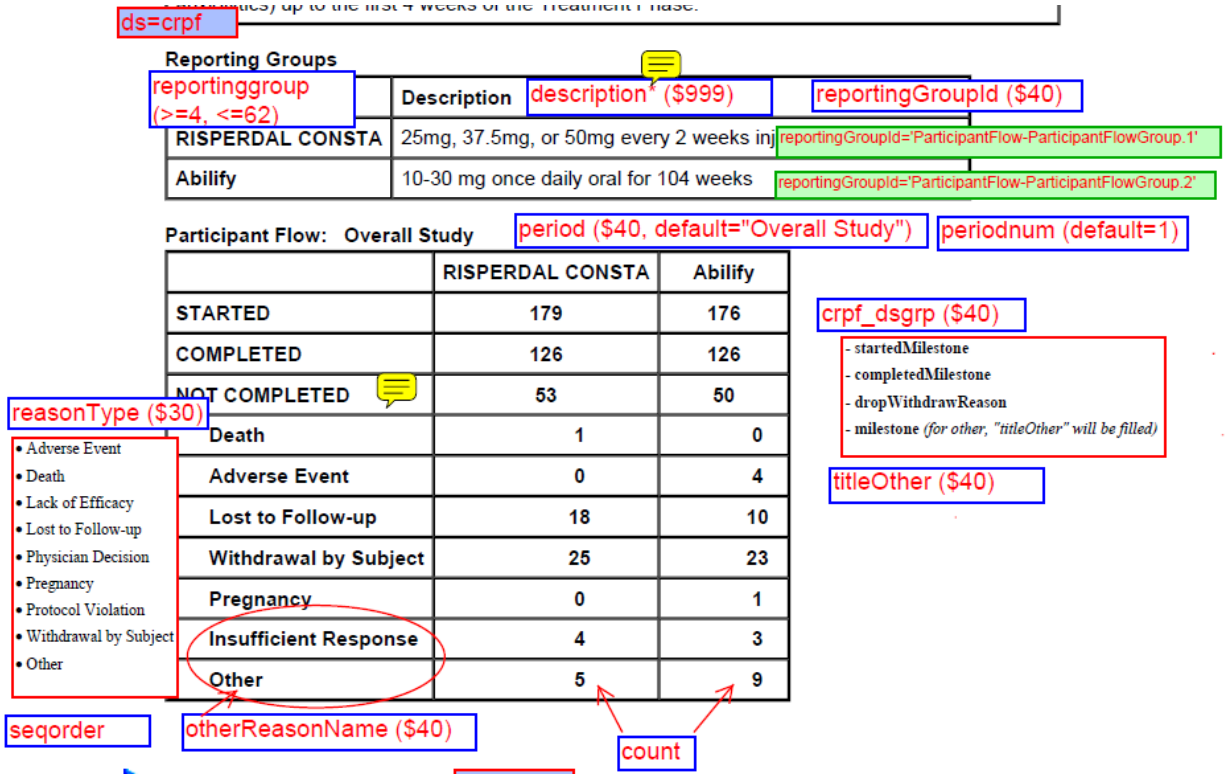
We developed a set of SAS macros to create required XML file that pull together with aggregated result data from our defined SAS datasets and other descriptive registration free-text information (such as registration IDs, and treatment descriptions). The following is an example of SAS codes to create a required XML file by calling our internal developed SAS macros %crgetfreetext and %crSAStoXML to get information from our defined SAS datasets that contains aggregated result for Trial Information, Participant Flow /Subject Disposition, Baseline and Demographic characteristics, and Adverse Events.

```
*=== combine prepared files with free text =============;
  %crgetfreetext(excelfile1           = &ipath\freetextfields_Study123.xls, /*an input free-text file */
                 outfile              = &opath\CRissues.xls, /*an output list of issues found in XML file */
                 baselineds           =  a_in.crbs, /* input defined SAS dataset of Baseline meas. */
                 participantflowds = a_in.crpf, /* input defined SAS dataset of Participant flow */
                 aeds                 = a_in.crae, /*input defined SAS dataset of AE */
                 );
*=== Join all Sections ============================;
  %crSAStoXML;
```

VALIDATION AND REVIEW OF XML FILE

As a XML file is different from normal SAS datasets/outputs that SAS programmers are used to, we developed validation tool that can validate the format of XML file along with a stylesheet so that the SAS programmers can review a XML with an user-friendly view that mimics the view in ClinicalTrials.gov and EudraCT, and validate the XML file according to CLinicalTrials.gov and EudraCT requirement. Below is a view of participant flow result disclosure in ClinicalTrials.gov along with our annotation of defined SAS dataset:

Figure 2: Screen shot and annotation of Participant Flow Result Disclosure in ClinicalTrials.gov



CONCLUSION

The automation process can generate a XML file that can be directly loaded into registry system of ClinicalTrials.gov and/or EudraCT. The source data may be from a clinical trial database with various types of trial designs and/or different data modules from different companies. The automation process provides accuracy, efficiency and consistency in aggregated result posting required by FDAAA and/or EMA.

The set of SAS macros, developed by the Biostatistics and Programming group of Johnson & Johnson Pharmaceutical R&D, reads the input source SAS clinical trial datasets that are used in many pharmaceutical companies, and generates a validated XML file that meets requirements of ClinicalTrials.gov and EudraCT.

Both ClinicalTrials.gov and EudraCT requirements contain certain standards of controlled terminologies, such as, reasons for discontinuations, age group, MedDRA coding for AE, and the specific rule for reporting frequent adverse events and AE reporting threshold. Therefore, it is better to build these requirements into the setup of the clinical trial database and analysis plan, it will then be more efficient to generate the aggregated results for result disclosure in ClinicalTrials.gov and EudraCT.

REFERENCES

ClinicalTrials.gov. Accessed March 8, 2018. http://www.clinicaltrials.gov

ClinicalTrials.gov Protocol Registration System (PRS) “ClinicalTrials.gov Results Data Element Definitions”. Accessed March 8, 2018. https://prsinfo.clinicaltrials.gov/results_definitions.html.

EU Clinical Trials Register. Accessed March 8, 2018. www.clinicaltrialsregister.eu.

EudraCT Results related documentation. Accessed March 8, 2018. https://eudract.ema.europa.eu/result.html

ACKNOWLEDGMENTS

I would like to thank the following colleagues who made positive valuable contributions in development and maintenance of automation process/macros/tools:

Denis Michel

Augustyns, Ilse

Lim, Layfong

Kang, Diye

Li, James

He, Yeming

Zackiewicz, Christina

Schirmer, Erica

Fu, Min

Arikala, Sivasankara

Yan, Kenneth

CONTACT INFORMATION

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

Sherry Meeh
Integrated Data Analytics and Reporting (IDAR),
Janssen, a pharmaceutical company of Johnson & Johnson.
1125 Trenton Harbourton Rd, Titusville, NJ 08560
Email: smeeh@its.jnj.com