

Using Python to make Clinical Trial Result Posting automatically

Derek Li

INTRODUCTION

The public accessibility of clinical trial results information is required both in United States and the European Union.

- US: Final Rule for Clinical Trials Registration and Results Information Submission (42 Code of Federal Regulations Part 11) , effective on 18Jan2017
- EU: Clinical Trial Regulation No. 536/2014, effective on 16Jun2014

Most pharmaceutical companies have designated a group to do this work, using information from clinical study reports (CSR) with additional manual effort. This brings several challenges, such as maintaining consistency across publicly available data for a given trial, and the different requirements for data disclosure on ClinicalTrials.gov and European Union Drug Regulation Authorities Clinical Trials databases.

All required data from the CSR are in Tables, Listings, Figures (TLFs). Both regulator's websites have a portal to upload data with specific XML format. Since FDA's required result can be converted from EMA's version, we decide to develop an App to make this process smoother with higher accuracy and efficiency in an automated way.

INTRODUCTION

For FDA (https://prsinfo.clinicaltrials.gov/results_definitions.html) the required information includes:

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

```
importRecordCollection
└─ importRecord "xxxx"
  └─ resultsRegistrationInfo
    └─ result "true"
      └─ baseline
      └─ certainAgreement
      └─ limitationsAndCaveats
      └─ outcomeMeasures
      └─ participantFlow
        └─ participantFlowGroups
        └─ periods
          └─ preAssignmentDescription A total of 443 participants were scr
          └─ recruitmentDetails This study was conducted from 19-Nov-201
          └─ typeUnitsAnalyzed
        └─ pointOfContact clinical-trials-contact@bayer.com
        └─ reportedEvents
```

ClinicalTrials.gov Results Data Element Definitions for Interventional and Observational Studies

February 1, 2021

This document describes the definitions for results data elements submitted to ClinicalTrials.gov for interventional studies (clinical trials) and observational studies. These definitions are mostly adapted from [42 CFR Part 11](#).

Data element entries are annotated with symbols to indicate generally what information is required to be submitted and under which circumstances. The responsible party must ensure that the information provided complies with any applicable laws, regulations, or policies. For more information about various requirements and definitions of regulatory terms under 42 CFR Part 11, see [Support Materials](#).

Note: The term "clinical study" is used to refer to both interventional and observational studies. The term "participant" is used to refer to a human subject.

* Required

*§ Required if Primary Completion Date is on or after January 18, 2017

[*] Conditionally required

INTRODUCTION

For EMA(https://eudract.ema.europa.eu/multimedia_tutorials.html) the required information includes :

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

```
ns3:result "http://eudract.ema.europa.eu/schema/clinical_trial_result/adverse_event"
  trialInformation xxxx
  subjectDisposition Study was conducted at multiple centers in 7 countries
  baselineCharacteristics
    reportingModel BASELINE_REPORTING_MODEL.arms
    studyCategoricalCharacteristics
    studyContinuousCharacteristics
    ageContinuousCharacteristic true
    genderCategoricalCharacteristic true
    ageCategoricalCharacteristic false
    baselineReportingGroups
  endPoints
  trialChanges
  subjectAnalysisSets
  adverseEvents 0
```

Tutorials on posting results on EudraCT

In order to post results in EudraCT, please follow the steps below:

1. [Set up an EMA account](#) on the [EMA website](#) and use the new credentials to [register as results user](#) on the [EudraCT website](#) ([log in](#) using EMA credentials "become a results user", log out and [log in](#) again).
→ In case of login problems (inactive accounts/forgotten password): refer to [EMA account management](#)
2. [Request assignment to become the primary user for your trial](#): download the [form](#), and [fill it in](#), inserting your results user email as Requestor's email, [in](#), click on "your page" and "request assignment", insert your trial's details and upload the filled in form.
3. Ask another person to set up an EMA account and register as results user (see step 1) and [nominate him/her as back-up user](#)
4. Go to "your page", click on "edit" nearby your trial and **upload results** as:
 - o a **summary attachment** (e.g. pdf file), if the end of trial date of your trial is before 21st of July 2013 and your trial is not part of a PIP/Art 46 o your trial was prematurely ended, no subjects recruited, or only partial results are available.
 - o a **full data set**, in all other cases. See tutorials:
 - [Trial Information](#)
 - [Subject Disposition](#)
 - [Baseline Characteristics](#)
 - [End Points](#)
 - [Adverse Events](#)
 - [More Information](#)

INTRODUCTION

Common requirements from both FDA and EMA:

- Trial Information
 - Count of subject of each country
 - Population Age Group
- Subject Disposition
 - total number of subjects completed / non-completed trial
 - reasons of discontinuations
- Baseline Characteristics
 - Age
 - Gender
 - Categorical Baseline Characteristics
 - Continuous Baseline Characteristics
- End Points/Outcome Measures
- Adverse Events

INTRODUCTION

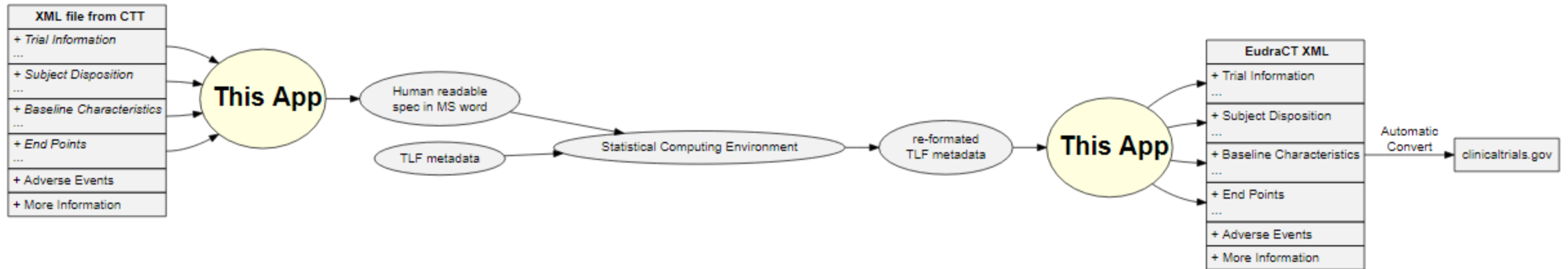
Examples of TLF metadata:

	text	col1 Drug A@N=2833(100%)	col2 Drug B@N=2841(100%)	col3 Total@N=5674 (100%)
1	Race			
2	WHITE	1777 (62.7%)	1815 (63.9%)	3592 (63.3%)
3	BLACK OR AF...	140 (4.9%)	124 (4.4%)	264 (4.7%)
4	ASIAN	717 (25.3%)	723 (25.4%)	1440 (25.4%)
5	AMERICAN IN...	78 (2.8%)	76 (2.7%)	154 (2.7%)
6	NATIVE HAWAI...	11 (0.4%)	7 (0.2%)	18 (0.3%)
7	NOT REPORTED	9 (0.3%)	10 (0.4%)	19 (0.3%)
8	MULTIPLE	101 (3.6%)	86 (3.0%)	187 (3.3%)
9	Sex			
0	Male	1953 (68.9%)	2030 (71.5%)	3983 (70.2%)
1	Female	880 (31.1%)	811 (28.5%)	1691 (29.8%)
2	Age (YEARS)			
3	n	2833	2841	5674
4	Mean	65.44	65.67	65.56
5	SD	8.94	9.16	9.05
6	Min	32.0	28.0	28.0
7	Q1	60.00	60.00	60.00
8	Median	66.00	66.00	66.00
9	Q3	72.00	72.00	72.00
0	Max	90.0	97.0	97.0

	name	_COL1_ Drug A@N=2866(100%)	_COL2_ Drug B@N=2868(100%)	_COL3_ Total@N=5734 (100%)
1	Completed epoch	2011 (70.2%)	2040 (71.1%)	4051 (70.6%)
2	DEATH	130 (4.5%)	157 (5.5%)	287 (5.0%)
3	WITHDRAWAL BY SUBJECT	159 (5.5%)	172 (6.0%)	331 (5.8%)
4	LOST TO FOLLOW-UP	5 (0.2%)	4 (0.1%)	9 (0.2%)
5	NON-COMPLIANCE WITH STUDY DRUG	19 (0.7%)	7 (0.2%)	26 (0.5%)
6	PHYSICIAN DECISION	148 (5.2%)	109 (3.8%)	257 (4.5%)
7	ADVERSE EVENT NOT ASSOCIATED WITH CL...	0	2 (<0.1%)	2 (<0.1%)
8	SPONSOR DECISION	6 (0.2%)	2 (<0.1%)	8 (0.1%)

INTRODUCTION

- All of these information can be found from the CSR tables
- Both FDA and EMA website have portal to upload XML file
- FDA required XML file can be converted from EMA XML file automatically



CTT = Clinical Trial Transparency

NOTE: Adverse Events section is not included as we already have a separate standard work process on it.

Step 01: Create specs from XML file

As XML file format is not human readable, the first step is to create a human readable specification with Microsoft Word format (.DOCX), which is just like a TLF specification for our TLF programming during our general work as a statistical analyst.

Go To

- ☒ Create Specification
- ☐ Update the XML
- ☐ Check result XML

Create SPEC

This page is used to create the Spec base on the XML file from web posting team

Upload XML file



Drag and drop file here
Limit 200MB per file

Browse files

Create spec

Step 01: Create specs from XML file

From this specification, we know:

- Which kind of data are needed
- In what kind of format

Periods and discontinuation reasons

Shells

phase	period	periodid	categoryid	details	other	group0	group1	group2	group3
0	1	PostAssignmentPeriod-46431	CompletedMilestone-113150	complete treatment	xx	xx	xx	xx	xx
0	1	PostAssignmentPeriod-46431	StartedMilestone-113151	start treatment	xx	xx	xx	xx	xx
999	1		ReasonNotCompleted-126704	physicianDecision		9	9	9	9
999	1		ReasonNotCompleted-126705	other	Private reason	9	9	9	9

period: flag variable of each period, mandatory

periodid: ID of each period, extract from input xml, only for reference

phase: flag variable of not complete reason of each period, only need to populate 999 for all withdraw reasons, mandatory

categoryid: ID of each category, extract from input xml, mandatory

details/other: detail wording of each category, extract from xml and for reference

groupxx: Treatment group variables

NOTE: Only exist reasons in input dataset will be kept

Step 02: Update the input XML file

Example SAS programs about how to extract, re-format the needed data from TLF metadata also provided:

```
data DISP;
  set TLFMETA.T_14_1_2_ADSL_DISPOVERALL_1;
  period=1;
  if _order_ = int(_order_) then phase=0;
  else phase=999;
  if _order_ = 4 then categoryid="CompletedMilestone-119604";
  if _order_ = 3 then categoryid="StartedMilestone-119605";
  if _order_ = 7.001 then categoryid="ReasonNotCompleted-132869";
  if _order_ = 6.003 then categoryid="ReasonNotCompleted-132867";
  if _order_ = 6.001 then categoryid="ReasonNotCompleted-132865";
  if _order_ = 6.002 then categoryid="ReasonNotCompleted-132866";
  group0=_col1_;
  group1=_col2_;
  group2=_col3_;
  group3=_col4_;
  if not missing(categoryid);
  keep phase period categoryid details group0 group1 group2 gro
run;
```

[illegible]

Step 02: Update the input XML file

As shown in this snapshot, each required section from the XML file has a sheet:

- COUNTRY: Count of subject of each country, mandatory
- AGE: Population Age Group, mandatory
- DISP: Population disposition, the category may not consist with the data shown in CSR, mandatory
- AGE_N: Age summary as in CSR, optional
- AGE_C: Age category analysis as in CSR, optional
- GENDER: Gender Baseline summary as in CSR, optional
- DEMO_C: Categorical Baseline Characteristics, optional
- DEMO_N: Continuous Baseline Characteristics, optional
- ENDPOINTS: End Points/Outcome Measures, mandatory
- STATS: The statistical analysis for each end pints, optional

Step 02: Update the input XML file

Go To

- ☐ Create Specification
- ☒ Update the XML
- ☐ Check result XML

Upload XML file



Drag and drop file here

Limit 200MB per file

Browse files

Upload Eucat contry excel file



Drag and drop file here

Limit 200MB per file • XLSX

Browse files

Country:

country

The Country EuCAT ID is:

	identifier	term	short_name	version	abv2
0	100000000315	Islamic Republic of Af...	Afghanistan	14	Al
1	100000000316	Åland Islands	Åland Islands	14	A:
2	100000000317	Republic of Albania	Albania	14	Al

Step 02: Accuracy Check

XML→XLSX, then compare the XLSX file with the input XLSX file from Step 2

[illegible]

CONCLUSION

- The XML file can be upload to EMA website smoothly, without worry about the system requirements/schemas, categoryIDs, etc.
- FDA's XML file can convert from EMA's XML automatically.
- High accuracy and more consistency with the data from CSR while decreasing time to prepare by reducing manual effort from clinical trial posting team and reducing validation effort from study team.
- Take advantage of a web application to make sure high accessibility.
- Outsourced study can also be processed if Vender can provide needed EXCEL file.
- The biggest pain-point is extracting required data from TLF metadata, so highly standardized TLF metadata is preferred.

CONTACT INFORMATION

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

Derek Li
Industry Regulatory Interaction-Statistical Programming Analyst,
Bayer HealthCare
Email: derek.li1@bayer.com