PharmaSUG 2017 - Paper PO07

ClinicalTrials.gov Results: an End of Study Deliverable That Should Be Considered at Study Startup

Maya Barton and Elizabeth Paynter, Rho, Inc.

ABSTRACT

Results submission to ClinicalTrials.gov is a deliverable required by law for most clinical trials. The deadline occurs one year from reaching the primary endpoint Last Patient Last Visit (LPLV) or the study LPLV. If the primary endpoint LPLV occurs prior to the study LPLV, this would lead to two submissions to ClinicalTrials.gov. Since this milestone will be different for different studies, the optimum time point to start working on this deliverable needs to be considered. Additionally, the submission typically pulls in data from multiple resources which are populated at different time points during the study lifecycle. We propose using early planning and standardization at study startup so that with the press of a button results will be generated for ClinicalTrials.gov for review and upload to the website. With this in mind we created a template program using SAS Enterprise Guide 7.1® that will enable the study team to map the data from multiple sources into one. Using the published standard template from ClinicalTrials.gov, we developed generic and standard data sets and tables to incorporate the fields and formats specified in the ClinicalTrials.gov template. This will eliminate manual entry of study protocol information that may not be used as titles, headers, or footnotes in everyday data cleaning and bio-statistical analysis data sets and displays. When working on multiple studies, setup of email alert programs for upcoming ClinicalTrials.gov milestones is also essential. This monitors the clinical data status and helps in planning the generation and review for correct and timely submission.

INTRODUCTION

Submitting results to ClinicalTrials.gov can be a complex deliverable during the end stage of a clinical trial. Many in this industry work with multiple clinical trials at once, necessitating the use of planning and standards to efficiently manage deliverables across trials. During the course of this paper we hope to provide the following insights into managing ClinicalTrials.gov results development and submission processes:

- Understand the time points at which either 1) information needs to be collected or 2) work needs to be performed in relation to results development for ClinicalTrials.gov
- Understand where all of the information needed for ClinicalTrials.gov results submissions will come from
- Develop a standardized template program that can be built at study start-up. This program will perform the following roles:
    - Create a standard across multiple trials for which ClinicalTrials.gov results submission is required – reduces work per trial
    - Majority of work on this program performed during study start-up rather than end of study, ensuring deadlines are met
    - Consolidates deliverable into one step, while still leaving room for review and validation as needed.

CLINICALTRIALS.GOV RESULTS DEVELOPMENT AND SUBMISSION

ClinicalTrials.gov is the government's protocol registration (PRS) and basic results database for clinical trials. The purpose of this database is to provide the public with a record of clinical study protocols and basic study results and create clinical trial transparency. ClinicalTrials.gov has two major goals, 1) reduce publication bias, and 2) reduce outcome reporting bias.

All applicable clinical trials (as described in Section 801 of the Food and Drug Administration Amendments Act [FDAAA 801 {Clinical Trial Databases, 2007}] and clarified in the Final Rule for Clinical Trials Registration and Results Information Submission [42 CFR Part 11 {Clinical Trials Registration and Results Information Submission, 2016}]) must be registered on ClinicalTrials.gov and submit basic results.

The responsible party for completing these requirements is typically the sponsor or principal investigator. However this task can be delegated to other parties, such as a contract research organization (CRO). For the purposes of this paper, we will be presenting management of this deliverable from the point of view of a CRO that manages many clinical trials from different sponsors.

CLINICALTRIALS.GOV REQUIREMENTS

For applicable clinical trials, basic results must be posted within 12 months of the Last Patient Last Visit (LPLV) for the primary endpoint. If the primary endpoint occurs before the end of the study, then basic results must be posted twice, once within 12 months of the primary endpoint LPLV and again within 12 months of the study LPLV.

Below are the categories of information required for ClinicalTrials.gov basic results submission:

- Participant flow
- Baseline results
- Primary endpoint outcomes
- Secondary endpoint outcomes
- Adverse events

There are two methods for submitting basic results to ClinicalTrials.gov:

1. Manual data entry
2. Upload results via eXtensible Markup Language (XML)

Manual Data Entry

ClinicalTrials.gov requires specific information with only minor flexibility to allow for search-ability and standardization across its vast database. The most common way to enter this information is to log in to the PRS and enter the data directly online for the study's record within ClinicalTrials.gov.

This method is time consuming, even if you have planned ahead and developed a template to gather all necessary information beforehand.

XML Upload

ClinicalTrials.gov provides a schema to allow for the upload of results via XML. With this schema it is possible to programmatically pull necessary information to populate the results. From the outset, this seems like the obvious choice for many reasons. A little work up front and then at the press of a button, all the work for an end of study deliverable is complete.

This 'press of a button' programmatic results output is the end goal of our proposal for management of ClinicalTrials.gov results submission. However, this lofty goal comes with several hurdles to jump.

1. Information required for ClinicalTrials.gov may not be in data sets
2. The program has to be flexible enough to accommodate the needs of different types of clinical trials
3. Steep learning curve to populating results for ClinicalTrials.gov, compounded with different people working on each clinical trial
4. A standard program template needs to exist so that the wheel does not have to be reinvented for each new trial that needs ClinicalTrials.gov results development

The biggest hurdle to getting the program up and running to create a schema compliant XML for submission to ClinicalTrials.gov is understanding the lifespan of a clinical trial as it relates to where data points are created and can be drawn on later for results creation. In the next section we will discuss the life span of a clinical trial and why we need to focus on early development on this program to create results for ClinicalTrials.gov submission.

CLINICAL TRIAL LIFESPAN AND CLINICALTRIALS.GOV RESULTS

In order to understand the full impact of ClinicalTrials.gov results submission, it is important to review the lifespan of the entire clinical trial. In the next few sections we will review three main phases during the life of a clinical trial, study start-up, maintenance, and end of study. At each point within the life of clinical trial we will discuss what information is needed for ClinicalTrials.gov results development. See Figure 1 for a graphical view of the clinical trial lifespan.

Figure 1 is the lifespan of a clinical trial.

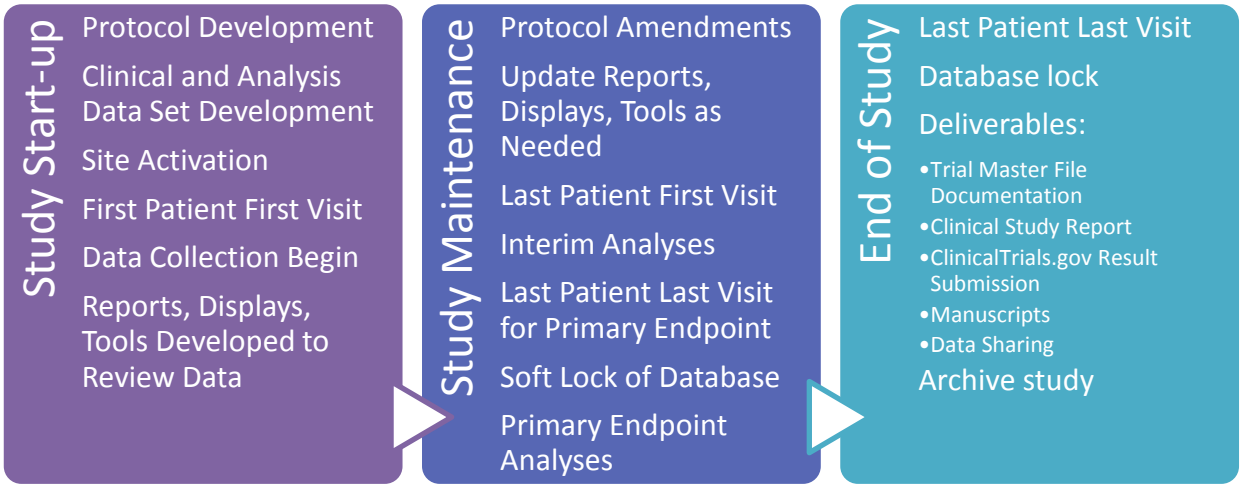

Figure 1. The lifespan of a clinical trial broken into three main phases: study start-up, study maintenance, and end of study

STUDY START-UP

Study start up is the time in a clinical trial's life where there is a flurry of activity. During study start-up the study team plans everything that they want to happen during the trial. This starts with the protocol development, followed by development of how data will be captured. Once that is known, programming can begin on a multitude of study related items. This includes, but is not limited to: electronic data capture systems, data sets, displays, and reports.

We feel that this is the most important phase of the study as it relates to ClinicalTrials.gov results submission. This phase of a clinical trial's life is the point at which all planning is made to set up the data storage for the remainder of the clinical trial. If the study team considers the needs of the clinical trial during maintenance and end of study phases as well as the immediate needs during study start-up, much efficiency can be made.

Standardization across clinical trials is the key to efficiency in the programmatic effort. In Table 1 we review what information is required for ClinicalTrials.gov and an example of where that data might be

placed for a clinical trial. If a CRO puts the same type of data for every trial they manage in the same place, then a template program for ClinicalTrials.gov results submission is possible.

Required Sections for ClinicalTrials.gov	What kind of Data is Required	Example of Data Storage Location
Participant flow	Screening/Enrollment	Enrollment data set – standardized (CDISC) same name/formats/drive locations
	Treatment arms	Protocol – stored in a Protocol Reference Model (PRM) type database
	Disposition	Subject level data set – standardized (CDISC) same name/formats/drive locations
Baseline results	Demographics	Subject level data set – standardized (CDISC) same name/formats/drive locations
	Baseline measures that directly correspond to endpoint outcomes	Various disease specific data sets - standardized (CDISC) same name/formats/drive locations --Map out specifics (which data sets this data will be in and variables needed)
		Measure descriptions – stored in a PRM type database
Primary endpoint outcomes	Primary endpoint data	Derived data sets – these may not be standard, but every effort should be made towards standardization --Map out specifics (which data set(s) this data will be in and variables needed)
		Measure descriptions – stored in a PRM type database
Secondary endpoint outcomes	Secondary endpoint outcomes	Derived data sets – these may not be standard, but every effort should be made towards standardization --Map out specifics (which data sets this data will be in and variables needed)
		Measure descriptions – stored in a PRM type database
Adverse events	Safety data	Safety data set – standardized (CDISC) same name/formats/drive locations
		Safety data descriptions – stored in a PRM type database

Table 1. Example of Mapping Data for ClinicalTrials.gov Results Submission

Study start-up is when data sets are designed, and endpoint analyses are considered. Since these considerations are directly related to deliverables at the end of study, the study team should be aware of requirements for ClinicalTrials.gov. Once these data elements are in place, this is a good time to start mapping out the program for the ClinicalTrials.gov XML creation. These locations and data elements should not change during the study, though the data itself will not exist until the study is up and running.

STUDY MAINTENANCE

During the study maintenance phase of the clinical trial life cycle there will not be nearly as much work needed for ClinicalTrials.gov results development. By the time a clinical trial has begun enrolling subjects, all of the study structure and the majority of the programmatic tasks should have been completed.

This phase of the trial allows the study team the opportunity to review the initial setup of the data capture of the trial to ensure that things are as standard as possible across other trials and to make sure all known activity required for end of study is prepared for as much as possible. To this end, while the ClinicalTrials.gov standard template program should have been already mapped out during study start-up, testing during the study maintenance phase is best practice. During study maintenance there will be data available in most of the locations that your program will be pointing to. This will allow you to confirm that there are no errors within the program that could cause delays at end of study. Another activity that should be ensured is that the PRM type database is maintained and updated per any protocol amendments or other changes in study.

Another item to consider during study maintenance is when the LPLV for the primary endpoint will occur. If this milestone is reached during the study maintenance phase, it is a good idea to set up an automatic email reminder program to alert the study team that this milestone has been reached. At this point the program must be run and the resulting XML output uploaded to the ClinicalTrials.gov PRS (see Section - Develop a Standard Template Program),

END OF STUDY

The end of study phase in the life cycle of a clinical trial can be the busiest time for the study team. There are many deliverables that can compete for priority. ClinicalTrials.gov results submission can sometimes be over shadowed by the big ticket deliverables like NDA submissions and manuscripts. However, there is a firm federally guided deadline of submission within 12 months of LPLV.

If the standard template required for ClinicalTrials.gov results gets created at study start-up, at study database lock, with a press of a button all of the output you need is created. This will save a significant amount of time and resources, as compared to assigning someone on the team to fill out a document manually. This manual process involves compiling information from multiple sources and requesting unanticipated analyses not done during study maintenance. Once the document is filled out it must then be reviewed by all applicable parties before it can be submitted to ClinicalTrials.gov. This demands considerable time and resources, and negatively affects a timely deliverable. Once the document is approved, it will require a copy and paste of the information into the fields within the PRS of ClinicalTrials.gov.

With a standard template program pulling from data sets that are standardized across studies you will be able to cut out the document creation process. When the program is run, one output would be available to send out to responsible parties for review. Once approved, the second output is the XML file that can be uploaded to the ClinicalTrials.gov PRS.

DEVELOP A STANDARD TEMPLATE PROGRAM

At this point we hope that we have convinced you that creating a standard template program is the optimal solution to the ClinicalTrials.gov results submission problem. We acknowledge that there are many moving pieces to clinical trials and not every study is going to be the same. However we feel that many things can be standardized, such as data sets, naming conventions, and locations. From the point of view of CRO, this is crucial to making the results development more efficient across different studies. Once that structure is in place, we propose a standard template program that pulls data from multiple sources into one data set and then outputs two documents. The first document is a user friendly file that shows the data that will be uploaded to ClinicalTrials.gov so responsible parties can review it. The second file is an XML file that can be uploaded to the ClinicalTrials.gov PRS.

Below are some examples of how a template program could work. We propose that the template start with the bare bones of requirement. Then each template program could be modified to meet the needs of each individual study.

ADVERSE EVENT MACRO

Before undertaking a larger task of creating a full study macro, we tested the functionality and compatibility of a smaller section of a study's clinical data, the adverse events. The reason for choosing adverse event data over any other data section required for ClinicalTrials.gov is that adverse event data is more standardized and therefore less variable across studies. This is a key feature for implementing a macro successfully.

An adverse events analysis data set is a standard requirement during study start-up. A PROC CONTENTS of a sample data set is shown below in Figure 2. This data set would be populated as automated periodic reports are run, resulting in active capture of adverse events as the study progresses. At database lock a final run of the program creating this data set would have complete clean information of all adverse events that occurred in the study.

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Label
52	AEACT	Char	2	$2.	Action Taken
53	AEACTC	Char	31	$31.	Action Taken (Char)
44	AEDUR	Num	8	8.	Duration of Adverse Event
41	AEENDT	Num	8	DATE9.	Stop Date of Adverse Event
42	AEENDTC	Char	9	$9.	Stop Date of Adverse Event (Char)
43	AEENDY	Num	8	8.	Adverse Event Study Day - End
59	AEICU	Char	2	$2.	Admitted to ICU
60	AEICUC	Char	3	$3.	Admitted to ICU
45	AEONG	Char	3	$3.	AE Ongoing?
46	AEOUT	Char	2	$2.	Outcome of Adverse Event
47	AEOUTC	Char	22	$22.	Outcome of Adverse Event (Char)
62	AEPT	Char	200	$200.	Preferred Term
54	AEREL	Char	2	$2.	Causality
55	AERELC	Char	15	$15.	Causality (Char)
56	AERELPRC	Char	15	$15.	Causality for periodic reports (Char)
57	AESER	Char	3	$3.	Serious Adverse Event
58	AESERC	Char	3	$3.	Serious Adverse Event (Char)
50	AESEV	Char	2	$2.	Severity/Intensity of Adverse Event
51	AESEVC	Char	16	$16.	Severity/Intensity of AE (Char)
61	AESOC	Char	200	$200.	System Organ Class
37	AESTDT	Num	8	DATE9.	Start Date of Adverse Event

Figure 2. Partial Screen Capture from PROC CONTENTS of the Adverse Events Data Set.

A generic macro to generate the required 'Adverse Events Summary Table' XML output is developed as one of the first building blocks of automation. This macro requires standardized variable names and formats across studies both in clinical and analysis data sets related to adverse events. These standardizations help to shorten the analysis data set specification process. A sample of the generic macro call that we have developed requires the following variables:

```
%macro AdverseEvents_for_ClinicalTrialsGov
  (SUBJLVLDS =      Subject level data set, # of treatment groups, # of subjects in each
                    treatment group.
  ,ADAEDS =         Adverse event data set.
```

,TRT =	Treatment group variable - if multiple treatment arms.
,SUBJID =	ID variable in &SUBJLVLDS.
,SERIOUS =	Serious adverse event (SAE) flag variable in &ADAEDS. Example: Yes/No, Y/N, 1/0
,PT =	Adverse event preferred term variable
,SOC =	Adverse event system organ class variable
,FREQTHRESH =	Frequency threshold for non SAEs. Non SAEs occurring among a smaller proportion of subjects than &FREQTHRESH in every treatment group are not included in the output. Default is 5% (.05).
,SOURCEVOCAB =	Source vocabulary, (Sample value: MedDRA 13.1)
,OUTPATH =	Complete path to desired output folder
);	

The required schema published by ClinicalTrials.gov can be used by this macro to generate an XML output of the 'Adverse Event Summary Table' that can then be uploaded to the ClinicalTrials.gov PRS.

Figure 3 is a sample partial screen capture of the 'Adverse Event Summary Table' XML output.

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 3.2 Final//EN">
<HTML>
   <BODY>
      <TABLE border="1" width="100%">
         <TBODY>
            <TR>
               <TD> False </TD>
               <TD> Systematic Assessment </TD>
               <TD> MedDRA V8.0 </TD>
               <TD>   </TD>
               <TD>   </TD>
               <TD> ReportedEvents-InterventionGroup.1 </TD>
               <TD> numEvents </TD>
               <TD> 27 </TD>
               <TD> 1 </TD>
            </TR>
            <TR>
               <TD> False </TD>
               <TD> Systematic Assessment </TD>
               <TD> MedDRA V8.0 </TD>
               <TD>   </TD>
               <TD>   </TD>
               <TD> ReportedEvents-InterventionGroup.1 </TD>
               <TD> numSubjectsAffected </TD>
               <TD> 10 </TD>
               <TD> 2 </TD>
            </TR>
            <TR>
               <TD> False </TD>
               <TD> Systematic Assessment </TD>
               <TD> MedDRA V8.0 </TD>
               <TD>   </TD>
```

Figure 3. Partial Screen Capture of the 'Adverse Events Summary Table' XML Output.

This macro has been tested and successfully used across multiple studies. The functionality of this program eliminates hand entry of every adverse event (which can number in the hundreds depending on the study indication and adverse event collection rules) and allows this section of required data entry in ClinicalTrials.gov to be almost completely automated.

DEVELOPING FULL RESULTS OUTPUT

Now that we, as a CRO, are dedicating ourselves to implementing greater standardization across our clinical trials, we feel that we can begin development of a larger scale template program that will automate each of the five data sections required for ClinicalTrials.gov results submission (see Table 1).

With the schema in hand from ClinicalTrials.gov, we will use similar macro language as used for the adverse event section to create a template program which pulls from each data element necessary to complete each section within ClinicalTrials.gov.

During study start-up all of the planning, and most of the programming, for analysis data sets will occur. We propose that this template program should actually be an analysis data set, ensuring that 1) it will get lumped into the programming that is necessarily done at study start-up and 2) the data will be stored in a way that will be logical and easy to find and compare later.

One of the harder problems to solve in order for this full results template program to work is pulling in data that may not live in a typical clinical or analysis data set. The following types of information are necessary for basic results submission to ClinicalTrials.gov but they don't live in a clinical dataset or analysis dataset:

- Study Metadata – Information related to the trial that is outside of what may be captured within a clinical database. This mostly includes protocol related information such as the study endpoints and treatment arms and treatment arm descriptions
- Endpoint Measure Descriptions – This information is specific to ClinicalTrials.gov in many cases. While the protocol will have a lot of this information, it must be modified to be understood by a lay person as well as contain certain elements that may or may not have been included in the protocol

In order to tackle the two data problems listed below we have worked over the last few years to develop a PRM type data base to track this type of information. This database holds a variety of study level metadata which is structured in such a way as to be programmatically usable. This data base must be maintained throughout the life of a study to ensure that all protocol amendments and study changes are captured.

With regards to the second problem, the PRM type database has to include fillable data points specific to ClinicalTrials.gov outcome measure descriptions. Some manual work will be needed to populate these fields within the PRM type database. We have spent the last few years tracking ClinicalTrials.gov approved endpoint measure descriptions within a database that the study can draw upon. This database of endpoint measure descriptions can be searched for similar endpoint measures and the descriptions recycled for each new study.

This analysis data set program, which we have dubbed ADCTGOV, would include the following outputs:

- Analysis Data set – Data set that contains every data point needed for ClinicalTrials.gov results submission
- RTF File – A document with tabular outputs displaying the data in a way that is similar to how it will appear on ClinicalTrials.gov so that responsible parties can review the results prior to submission to the ClinicalTrials.gov PRS
- XML File – XML file output per the schema available on ClinicalTrials.gov that can be uploaded into the ClinicalTrials.gov PRS

An added benefit to the program design is that if it is set up during study start-up, all that will be needed after LPLV for the primary endpoint (if it occurs before the final study LPLV) is to push a button and the baseline results and primary endpoint results will be output. Then after LPLV at the end of study, again,

all you would need to do is push the button again, and this time all of the results would be output. This is because the program will pull from other analysis data sets that are run on different time schedules, with some being populated with data throughout the study, and others only being populated after certain events (such as LPLV for the primary endpoint or LPLV).

When we implemented the adverse event macro and eliminated manual entry of adverse events into ClinicalTrials.gov, we noticed that the manpower needed dropped from ~10 hours per study to <1 hour per study. We are hopeful that successful implementation of ADCTGOV will significantly reduce manpower during the end of study phase of a clinical trial. We estimate it could save as much as 100 hours of work. While significant investment will be needed upfront to develop the initial program, with standardization, only minor effort will be needed to fit it to each study.

CONCLUSION

This paper presents an overview of ClinicalTrials.gov results submission and how to make this deliverable efficient when working with many different clinical trials. We proposed that a study team should use a standardized template program customized to the study during the study start-up phase of a clinical trial. This has the benefits of reducing work during a very resource scarce time at the end of study. It also allows the study team to create and submit results twice if needed due to different timing for the LPLV for primary endpoint and final study LPLV.

We would like to point out that while we have proposed this solution, we have only implemented a small scale example of this solution on multiple clinical trials. We were very successful with the first test, and we hope that we will be able to show you our outcomes in the near future.

REFERENCES

Clinical Trial Databases, Title 8: Food and Drug Administration Amendment Act §§ 801 (2007).

Clinical Trials Registration and Results Information Submission, Rule: National Institutes of Health, Department of Health and Human Services (2016).

CONTACT INFORMATION

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

Elizabeth Paynter
Elizabeth_paynter@rhoworld.com

Maya Barton
Maya_barton@rhoworld.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.