PharmaSUG China 2016-15

From Data Collection to Regulatory Submission – a Journey

Todd Case

Vertex Pharmaceuticals, Boston, USA

ABSTRACT

An NDA/BLA is the end result of an enormous effort to submit and present summaries of clinical trial results and data to the FDA (and almost certainly other geographies will soon be requiring actual data). But just *how* big of an effort is this? From First Patient/First Visit until Last Patient/Last Visit *hundreds of millions* of data points are collected, derived, summarized and written up in a CSR, etc. For example, a recent analysis for a 'routine' drug program; *for a single study*, had ***48,970,722*** data points that were either collected or created. It doesn't take a great imagination to calculate how much data will be created if multiple studies are included in a submission. This presentation will focus on flow of clinical trial data throughout its journey from visit/collection to CSR to NDA/BLA, provide useful metrics to show how much time is saved by using data standards and provide some recommendations about how to further improve efficient data flow.

INTRODUCTION

The purpose of this paper is to demonstrate how data standards (e.g., CDISC data standards such as SDTM and ADaM) streamline the flow and shorten the timeline of data from patient collection at the site through regulatory submission. It will also emphasize how uses of data standards facilitate/accommodate inevitable changes to a study such as protocol amendment, CRF modification or change to the Statistical Analysis Plan (SAP). As such, this presentation will focus on the flow of data throughout its journey from collection to CSR to NDA/BLA. It will also provide useful metrics to show how much effort is saved by using standards and provide some recommendations about how to further improve more efficient data flow and, ultimately, shorter time from Database Lock (DBL) to First Look/Top Line/Key Results announcement.

HOW IS 'A LOT' OF DATA DEFINED?

As mentioned in the abstract, there can be around 50 million or more data points (the number of rows times the number of columns for each dataset) for a single study. It's not uncommon to have at least two pivotal Phase III studies, a Phase II study and several Phase 1 studies as well as an Integrated Summary of Safety and Efficacy (which generate analyses based on pooled data) in a submission. In addition to this there are documents that describe the submission data (e.g., Define.xml and the Reviewer's Guide for both SDTM and ADaM). Based on this there are not only *hundreds of millions of data points* but also a lot of upstream/downstream impact if the protocol or CRF change and/or standards are not used or are employed in the late stage as opposed to the earliest stage (e.g., the raw data).

INITIAL CDISC ESTIMATES OF RETURN ON INVESTMENT USING DATA STANDARDS

When CDISC data standards were initially announced by the FDA in a 2004 Memo[1], there was a lot of enthusiasm, excitement and even fear by those who thought standards would make work magically (and instantly via automation) happen. After a couple of years of sponsors (or sponsors working with CROs) working with CDISC data the PhRM A-Gartner r-CDISC Project[2] estimated that Return on Investment (ROI) if using existing standards is 70%-90% for Start-up stage, ~40% for Study Conduct, ~50 for Analysis and Reporting and an Overall Savings of ~60%. What data standards would also mean (not necessarily spelled out in the paper) is that an entirely new set of skills and perhaps even more work (at least up front/in the initial implementation stages) would be needed than moving along without data standards.

While these findings suggest more research and validation is need to support the claims above and that these metrics can be defined and calculated in many ways it was clear in that report way back in 2009 that use of existing standards would expedite the flow of data and reporting in clinical trials once standards had been implemented.

[1] http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2004/ucm108330.htm

[2] http://www.cdisc.org/system/files/all/article/application/pdf/businesscasesummarywebmar09.pdf

ACTUAL RETURN ON INVESTMENT (ROI)

What this paper found was that, for a recent study, there were 95 raw datasets, 33 SDTM datasets and 7 ADaM datasets to support the required analysis. The number of data points decreased along with the number of data sets as well – from 36,705,621 (raw) to 16,011,253 (SDTM) to *4,624,216* (ADaM). And since all Tables/Listings/Figures (T/L/Fs) outputs are generated by ADaM datasets that means 8% of the data generated 100% of the analysis. If one measured ROI as percentage of the actual ADaM data used to report T/L/Fs as ROI this would indicate an ROI of 92%. There are undoubtedly other ways to calculate efficiency but it's worth noting just how much data gets collected and, by implementation of standards, how much easier and clear it is to generate analysis once we convert from raw→SDTM→ADaM. See Figure 1, below for a visual in terms of how much data is collected and how much of an impact implementation of Standards (both SDTM and ADaM) have on the process – enabling 8% of the actual data (as a percentage of all the data either collected, mapped or derived) to generate all of the analysis (T/L/Fs).

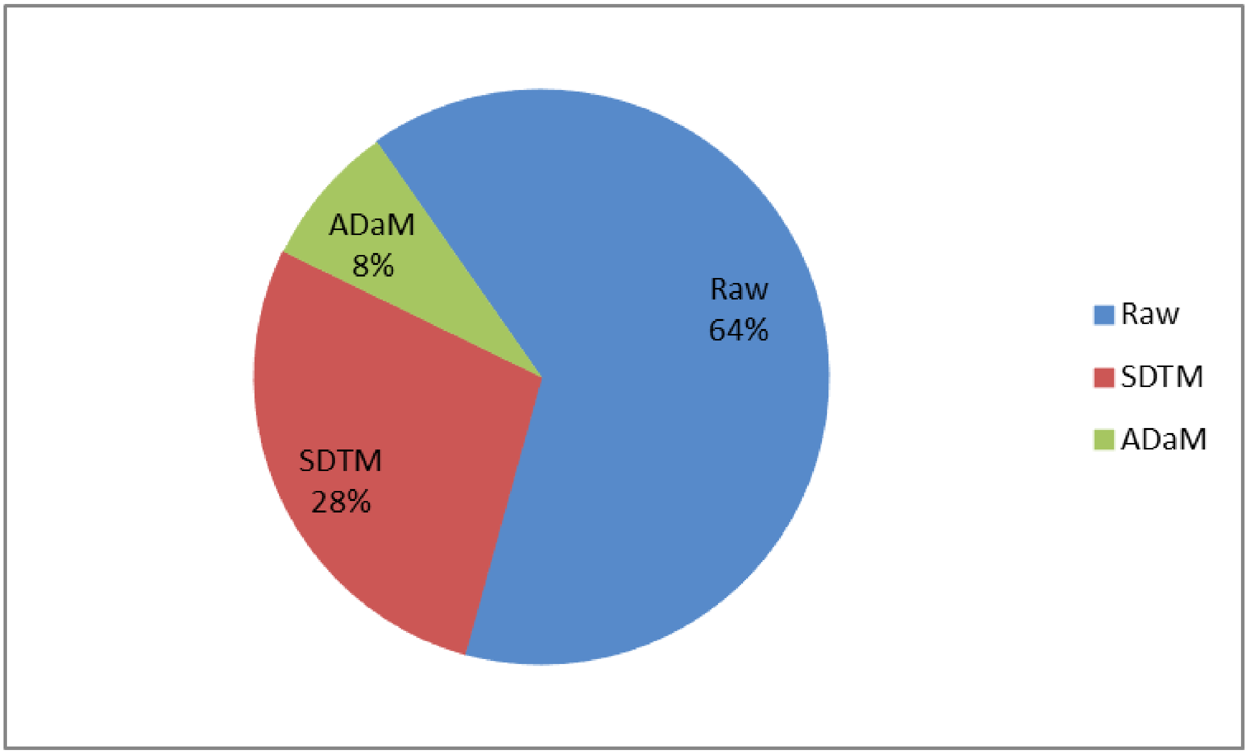

Figure 1. Percentage of Data, by Type

To summarize the amount of effort saved by standardization of data primarily by implementing the SDTM And ADaM models (which allow us to have a very specific list of variables and a flexible set of values for those variables) we see that having 10-20 ADaM datasets (for a submission with an ISS/ISE) ADaM datasets can allow us to generate all the required T/L/Fs – sometimes upwards of 1000 for a Pivotal Phase III/Registration study. Lastly, the FDA is mandating SDTM for all data by December 17, 2016[3] to submit SDTM for the vast majority of all studies.

FLOW OF CLINICAL TRIAL AND CLINICAL TRIAL DATA

While the first section of this paper describes at a high level the ROI for implementation of standards this section of the paper will focus on the flow of a clinical trial, particularly the flow of the data, with the goal of showing that any change to the overall design of the study (e.g., CRF, Protocol Amendment, non-adherence or misinterpretation of standards) has on the trial, particularly re-work. While it's much more difficult to actually calculate the cost it should be evident by taking a look at the flow of the data and the scope of a submission (if a study is fortunate enough to be included in a submission) how much time is saved by thinking more up front about Protocol and CRF development (e.g., the upstream activities).

[3] http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM455270.pdf

CLINICAL STUDY MILESTONES

Proof of Concept, Protocol, CRF, Database Build, First Patient First Visit (FPFV), Last Patient, Last Visit (LPLV) Database Lock, the Key Results Memo and hopefully a regulatory submission - these are usually what gets all the attention and headlines (rightly so as they are all major breakthrough dates either from a research, operational or financial and always patient perspective). However, there are numerous more steps in the process from the time proof of concept is established to a regulatory submission - for purposes of this paper the focus is on one way the data can flow (see Figure 2, below).

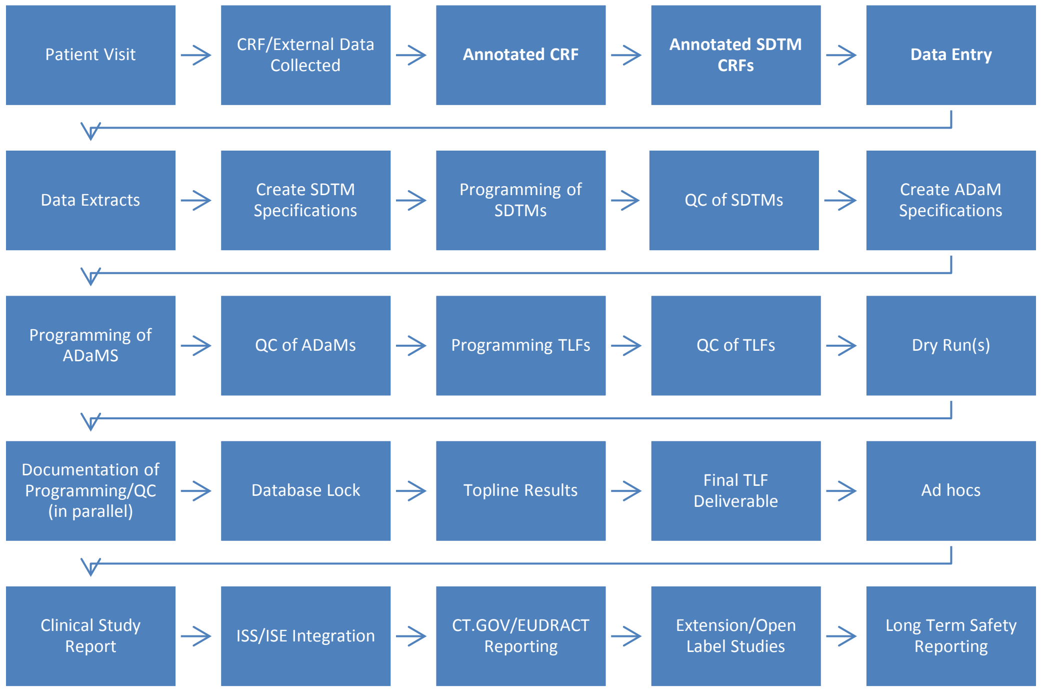

Figure 2. Example Flow of Clinical Trial Data

So, while a protocol amendment or a decision to not implement a current version of the standards is always made with necessity and often a lot of thought, if there is a change in direction the consequences downstream can be astronomical (less so with a protocol amendment but as you can see if it involves data there are at least 25 distinct time points in the flow of a clinical trial where this data will need to be entered, manipulated, changed, etc.). The more patients and/or visits that are impacted the more times you can multiply this process and realize how 'scope' creep doesn't take long to enter into the process – no matter how well-defined the process is.

This paper focused on Protocol Amendments as having a large impact on data flow and timelines, but even if there is a change that does not impact the protocol (e.g., CRF page is modified or SAP is updated), the impact goes back all the way to the 1st box (Patient Visit) in Figure 2, above. All data from that point on undergo the same process of programming, QC programming, documenting and reporting.

CONCLUSION

Clinical studies are very, very large activities (e.g., involving hundreds of millions of data points) that require a lot of process, documents, laws, organization, standards and share one goal: to get life-saving/changing therapies to patients as soon as possible. In order for that to happen ideally the Protocol and/or study design would be set in stone, patient recruitment would go just as projected and there would be no problems with data. But we live in reality where none of these three are likely to occur and, in addition, there could many other reasons for something which impacts the data to change during the course of the clinical trial.

As a result, the following are suggestions to help teams understand and work around proposed changes:

- Educate/remind teams of magnitude of data (*hundreds of millions of data points*) that exists for a typical trial and the effort any changes to the study design or standards have on data and corresponding timelines
- Understand the flow of the *entire* study and data

  - Understand how impact changes in study design (e.g., Protocol), documents (e.g., CRF), and/or data have on both upstream and downstream processes
- Educate teams that a change/addition/ can have an impact on millions of data records, maybe more:
  - Fixing raw data (~70% of all data)
  - Adding or re-mapping SDTM data (~25% of all data)
  - Modifying ADaM datasets (~5% of all data)
- Have a single source of Raw data (if possible), as opposed to multiple different Electronic Data Capture (EDC) or paper systems
- Have a library of standard CRFs
- Have a library of standard SDTM specs
- Have a library of template SAS SDTM template programs
- Dry Runs, Dry Runs, Dry Runs!

The last bullet in this section is extremely important and cannot be over-stated. Dry Runs are critical for practicing the flow and going through the motions that you and your entire team will go through right before and after a DBL. As patients, their family and caregivers, the industry and shareholders are expecting the results of clinical trials within days of a DBL this means that hundreds of millions of data points, thousands of SAS programs and outputs, required documentation and a process to unblind data (if it's a blinded trial) need to be executed flawlessly at DBL. This cannot be done without a process in place and practice trying to anticipate and address any and all issues that can arise.

REFERENCES

- Food and Drug Administration. "FDA Announces Standard Format That Drug Sponsors Can Use to Submit Human Drug Clinical Trial Data". FDA. July 21, 2004. Available at http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2004/ucm108330.htm
- Clinical Data Interchange Standards Consortium. Rozwell, Carol; Kush, Rebecca Daniels; Helton, Ed; Newby, Frank; Mason, Tanyss. "Business Case for CDISC Standards:Summary. PhRMA-Gartner-CDISC Project, September 2006". September, 2006. Available at http://www.cdisc.org/system/files/all/article/application/pdf/businesscasesummarywebmar09.pdf
- Food and Drug Administration. "Data Standards Strategy, 2015-2017. FDA Center for Drug Evaluation and Research (CDER), Food and Drug Administration (FDA). July 15, 2015. Available at http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM455270.pdf

ACKNOWLEDGMENTS

I'd like to acknowledge all my colleagues and management at Vertex Pharmaceuticals for their dedication to standards and quality.

CONTACT INFORMATION

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

Name: Todd Case
Enterprise: Vertex Pharmaceuticals
Address: 50 Northern Avenue
City, State ZIP: Boston, MA, 02210
Work Phone: +1 617 961 7907
E-mail: todd.case@vrtx.com