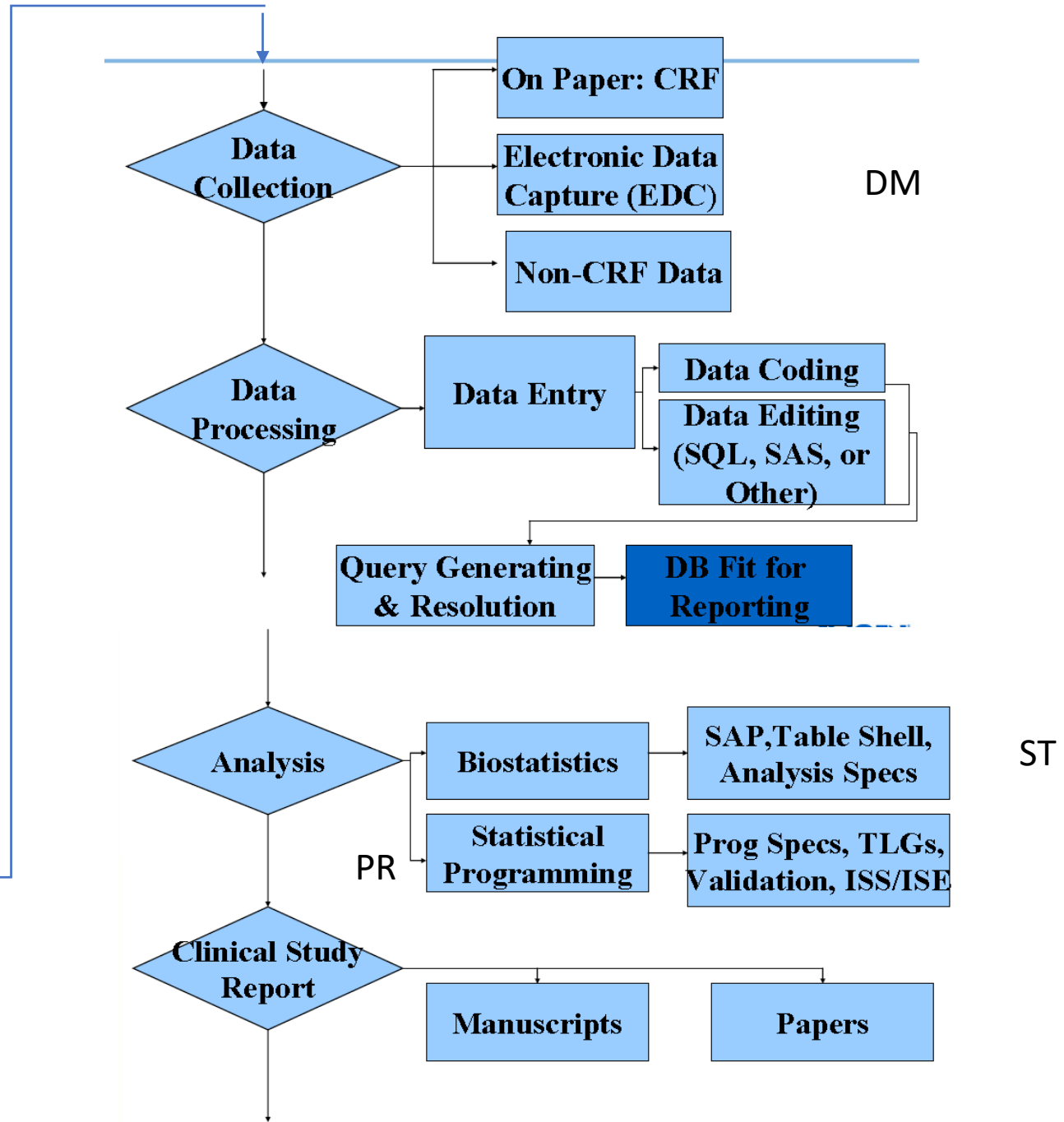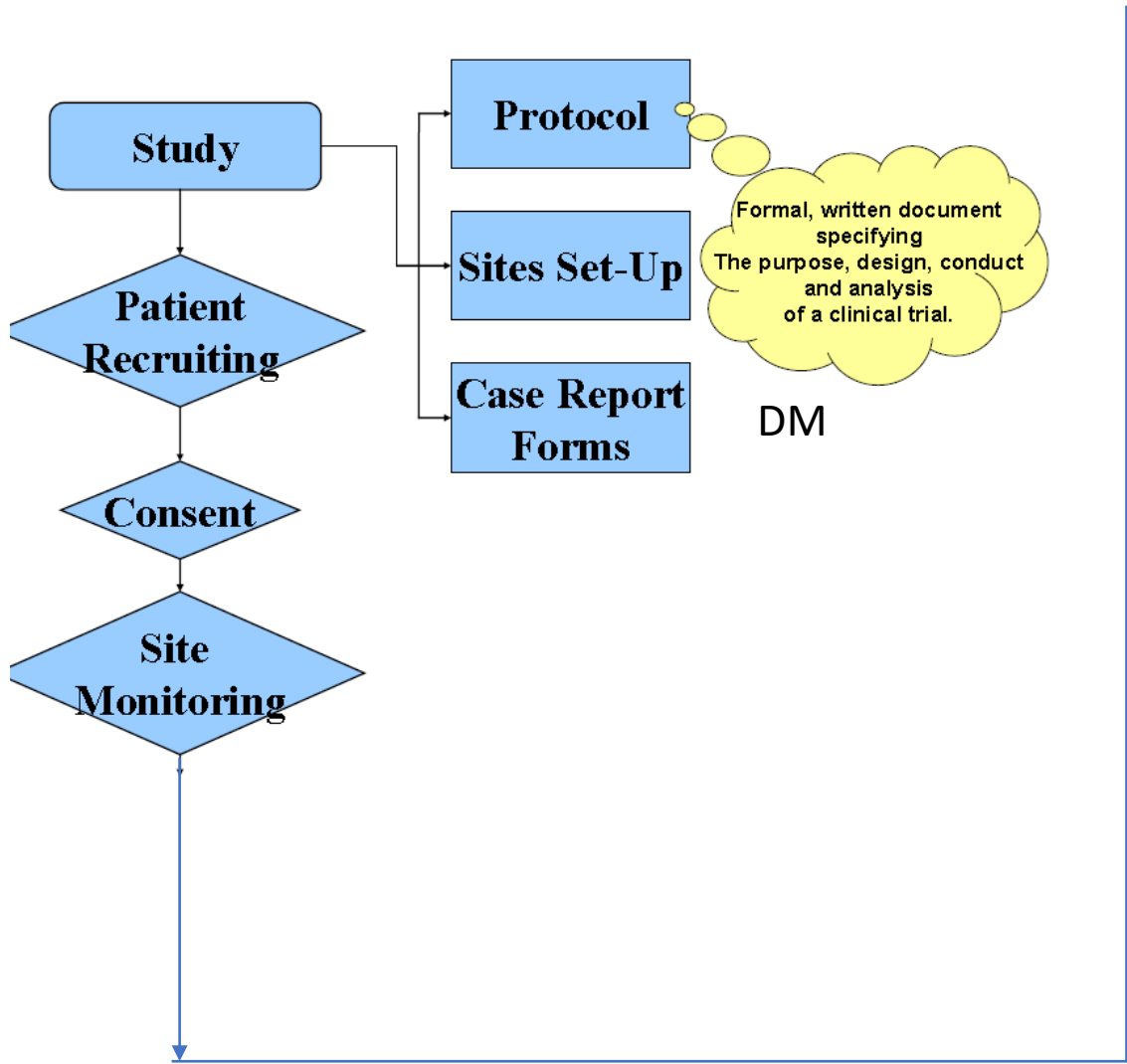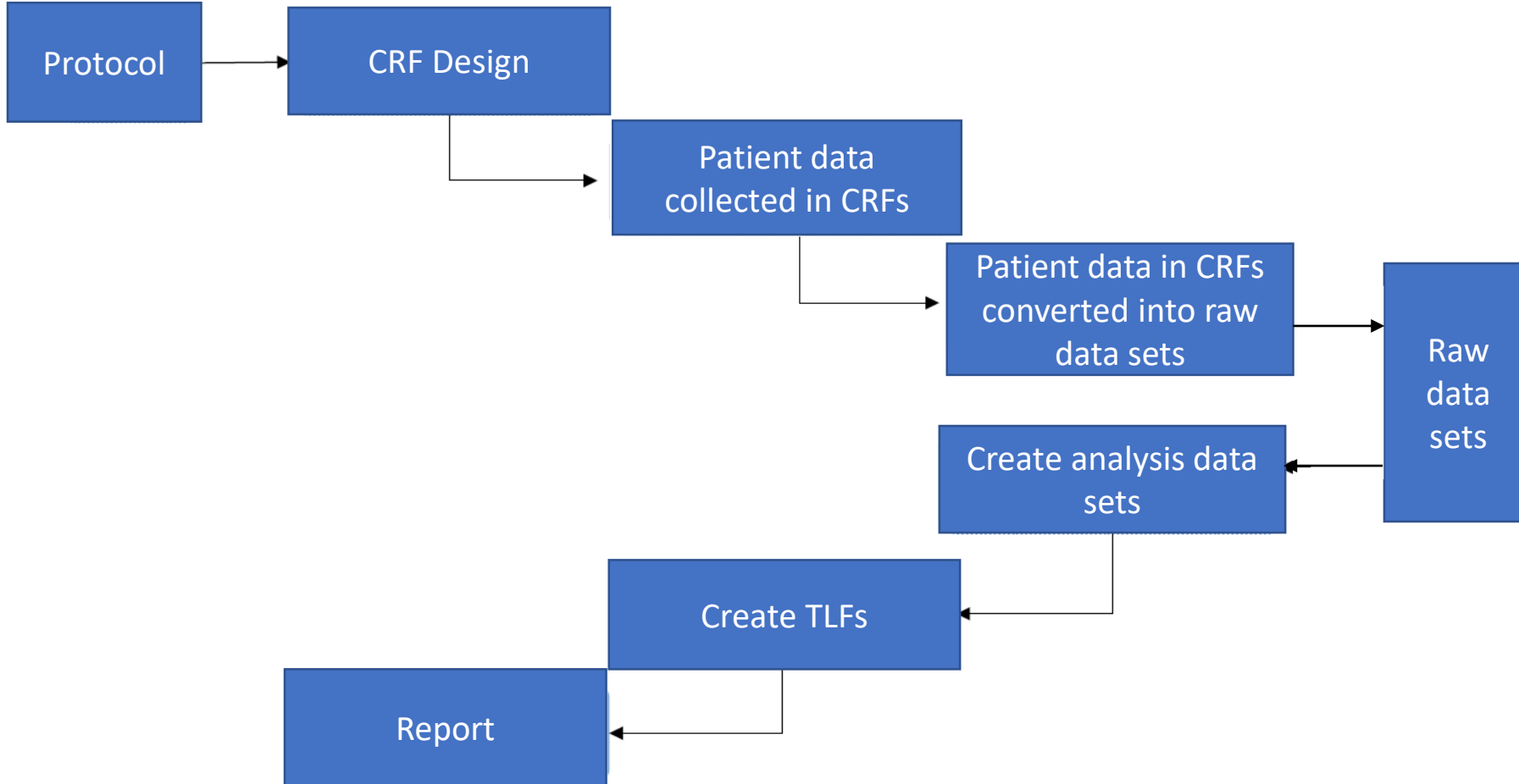# Flow of data in clinical trials

March, 2020

From Data Collection to Regulatory Submission – a Journey

- An New drug Application (NDA) is the end result of an enormous effort to submit and present summaries of clinical trial results and data to the Regulatory Authorities e.g. the FDA in the US.

- 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 e.g. CSR (clinical study report).

- 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. More often than not, multiple studies are included in a submission.

- We describe the 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.

Partially based on an article by: Todd Case, Vertex Pharmaceuticals, Boston, USA

Simplified flow of data

Clinical data management (CDM)

- **Clinical data management** is a critical phase in clinical research, which leads to generation of high-quality, reliable, and statistically sound data from clinical trials.

- Clinical data management assures collection, integration and availability (transfer) of data.

- It also supports the conduct, management and analysis of studies.

Subject Data Collection

- Data is collected on case report forms (CRF)
- Much of clinical data is taken from the subjects medical record (source documents)
- Pharmaceutical and device trials, data is verified by multiple players
- A good protocol leads to a simple, clear CRF, built from a standard template: Restrictive

Can I trust the data generated?

- Accuracy
- Reliability
- Integrity – can't be modified
- Availability – for staff and inspectors
- Authenticity – is it from the correct person

- Does the system do what it is meant to do?

Standrads: CDISK

- The **Clinical Data Interchange Standards Consortium** leads the development of global, system independent, data standards now commonly used as the underlying data structures for clinical trial data.

- CDISC data standards were initially announced by the FDA in 2004,

- This also meant that entirely new skills and perhaps even more work at least up front/in the initial implementation stages.

- After a couple of years of working with CDISC data an Overall Saving of ~60% was estimated.

- It is clear that use of standards expedites the flow of data and reporting in clinical trials once standards had been implemented.

Non-profit organization
in 2000

Clinical Data Interchange
Standards Consortium
(CDISC)

1. Development of industry standards to support the
   a) electronic acquisition
   b) exchange
   c) submission and
   d) archiving
   of clinical trials data
2. Development of global, platform independent standards
5. Improve data quality and accelerate product development

hbp://www.cdisc.org/sdtm/

Using Standrads

- Examples of CDISC data standards are SDTM and ADaM,

- These help streamline the flow and shorten the timeline of data from patient collection at the site through regulatory submission.

- Using data standards facilitates changes to a study such as protocol amendment, CRF modification or change to the Statistical Analysis Plan (SAP)

- This leads to more efficient data flow and, ultimately, shorter time from Database Lock (DBL) to Resuts.

Study Data Tabulation Model (SDTM)

- SDTM provides a standard for organizing and formatting clinical data to streamline processes in collection, management, analysis and reporting.

- SDTM is built around the concept of observations collected about subjects who participated in a clinical study.

- Each observation can be described by a series of variables, corresponding to a row in a dataset or table.

- SDTM is also used in non-clinical data (SEND), medical devices and pharmacogenomics/genetics studies.

SDTM

- Each variable can be classified according to its Role:
  - Identifier variables, which identify the study, subject of the observation, the domain, and the sequence number of the record
  - Topic variables, which specify the focus of the observation (such as the name of a lab test)
  - Timing variables, which describe the timing of the observation (such as start date and end date)
  - Qualifier variables, which include additional illustrative text, or numeric values that describe the results or additional traits of the observation (such as units or descriptive adjectives).
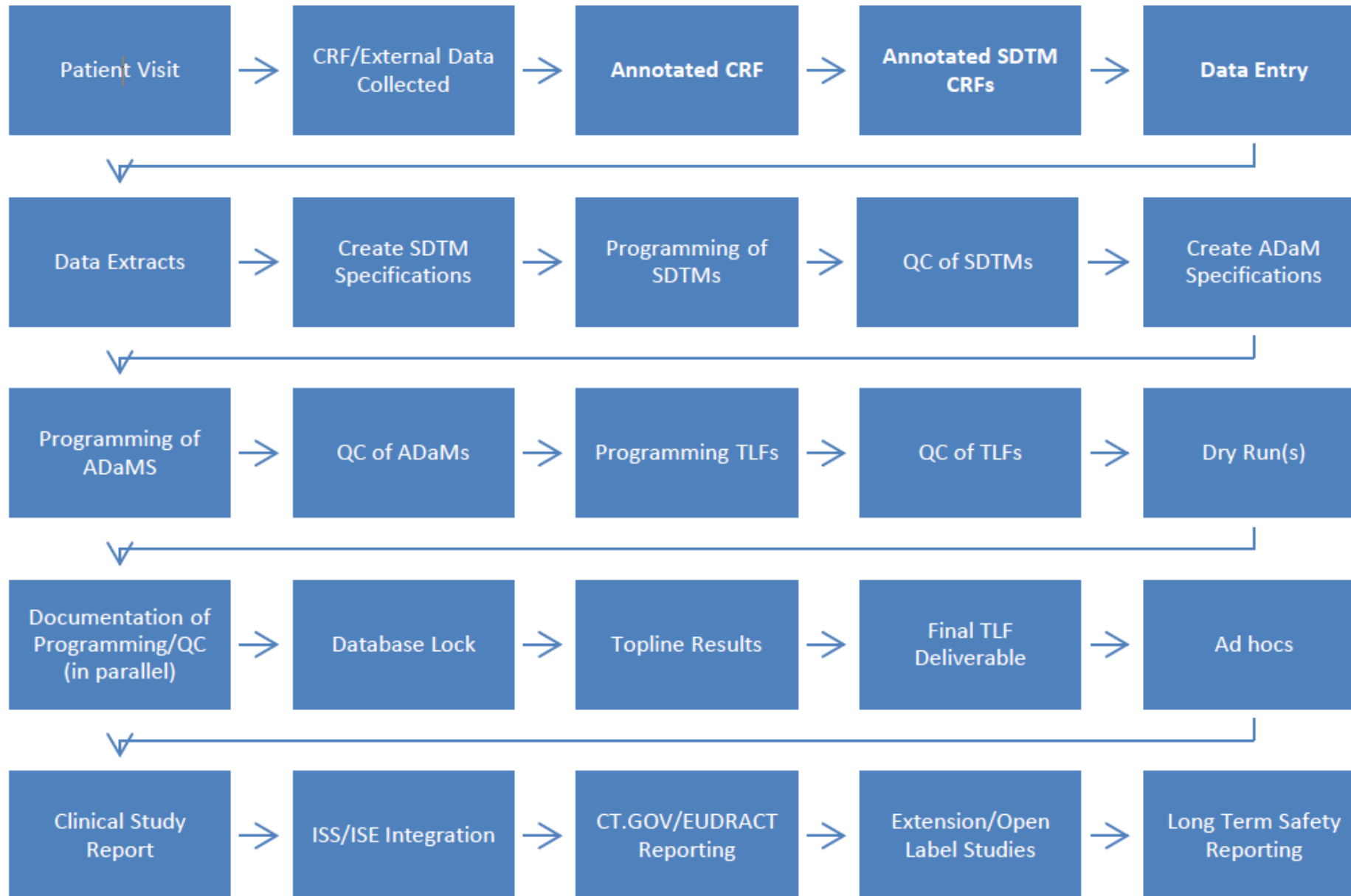
An example

- By implementation of standards, it is much easier to generate analysis once we convert from ==raw -> SDTM -> ADaM==.

- In 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 from ADaM datasets that means 8% of the data generated 100% of the analysis.

- Figure 1, below illustrates how much of an impact implementation of Standards has on the process – enabling 8% of the actual data to generate all of the analysis (T/L/Fs).
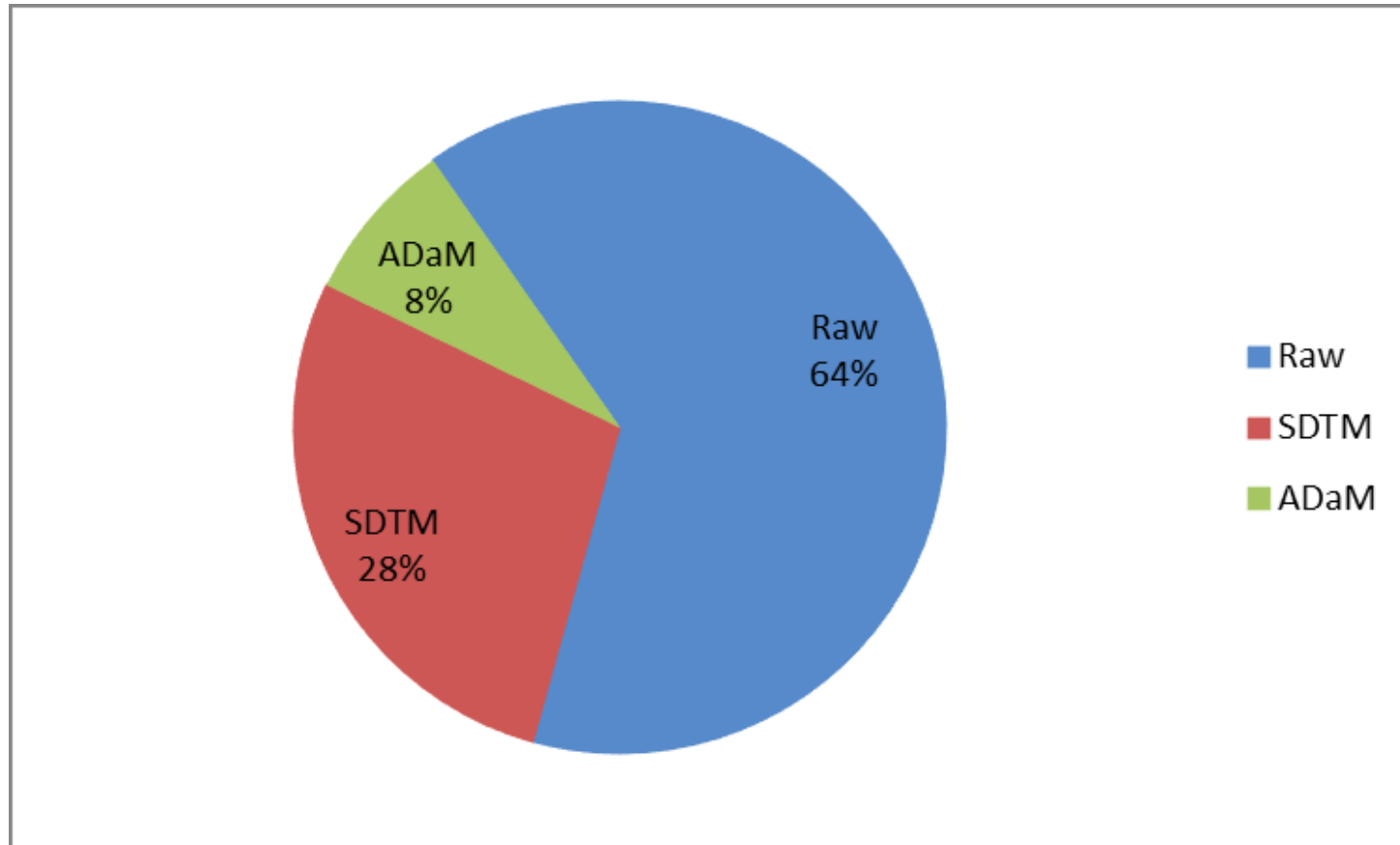
To summarize the amount of effort saved by standardization of data primarily by implementing the SDTM And ADaM models (which allow us to have a very specific list of variables and a flexible set of values for those variables) we see that having 10-20 ADaM datasets (for a submission with an ISS/ISE) ADaM 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, 20163 to submit SDTM for the vast majority of all studies.

Example Flow of Clinical Trial Data

Patient Visit	→ CRF/External Data Collected	→ **Annotated CRF**	→ **Annotated SDTM CRFs**	→ **Data Entry**

Data Extracts	→ Create SDTM Specifications	→ Programming of SDTMs	→ QC of SDTMs	→ Create ADaM Specifications

Programming of ADaMS	→ QC of ADaMs	→ Programming TLFs	→ QC of TLFs	→ Dry Run(s)

Documentation of Programming/QC (in parallel)	→ Database Lock	→ Topline Results	→ Final TLF Deliverable	→ Ad hocs

Clinical Study Report	→ ISS/ISE Integration	→ CT.GOV/EUDRACT Reporting	→ Extension/Open Label Studies	→ Long Term Safety Reporting

Dry Run?

- A dry Run (also called **Blind Delivery Review)** is conducted to allow the evaluation of the reporting datasets (SDTM/RDB) and statistical output tables, listings, and figures (TLFs) prior to database lock, to ensure that all outputs are prepared as expected as per agreement

- Dummy TLF outputs based on live SDTM the reporting datasets are checked whilst the data is still blinded

- Cross functional representation will be required to undertake review of the outputs, and at the DR Comments Resolution Meeting (CRM), to ensure the success of the DR

What is the purpose of DR?

The purpose of DR is primarily for verification of the assumptions of the analysis methods detailed in the Clinical Study Protocol (CSP) and/or Statistical Analysis Plan (SAP) and to also include an assessment of:

- *Recruitment, i.e. to ensure the correct subject population is being recruited*

- *Review of protocol deviations*

- *Investigate any potential safety issues*

As a result of DR**, the following changes may be needed:**

- Additional supporting tables may be considered (associated additional costs may apply)

- Analysis set rules may be reviewed, changes in rules for analysis sets documented in SAP amendment/addendum (optional format)

- Analysis set rules may include updated rules related to protocol deviations

• Planned analysis may need to change

Conclusions

- Clinical studies are very, very large activities (e.g., involving hundreds of millions of data points) that require a lot of processes, documents, laws, organization, standards and share one goal: to get new 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 in reality, there could many other reasons for the data to change during the course of the clinical trial.

- Therefore, it is good to be aware 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
  - The flow of the *entire* study and data
  - Understand how any changes in study design (e.g., Protocol), documents (e.g., CRF), and/or data can have big consequences on upstream

Conclusions

- 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)
- Use a single source of Raw data (if possible), as opposed to multiple different Electronic Data Capture (EDC) or paper systems
- Use a library of standard CRFs
- Use a library of standard SDTM specs
- Use a library of template programs
- Dry Runs, Dry Runs, Dry Runs!

Conclusions

- As patients, their family and caregivers and the industry 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