

Paper TT07

The Future of the Digitized Data Flow

Silas McKee, Accenture, Pennsylvania, USA

ABSTRACT

The protocol provides the plan for a clinical trial, yet we often struggle to translate that plan into the systems and artifacts needed to meet the trial objectives. In this paper, we describe a future where the study objectives and design clearly prescribe the data collection, management, and analysis required to meet the trial's goals, and where machine-readable specifications enable the automated build of databases, collection tools, and study artefacts. We explore key gaps in the existing clinical trial landscape and discuss potential enablers for the new process.

INTRODUCTION

The protocol is arguably the most important artifact for the conduct of a clinical trial. Nevertheless, the process for protocol authoring has not fundamentally changed in decades. Most organizations' processes involve a clinical scientist developing a protocol within Microsoft Word. Once reviewed and approved, the protocol is distributed to study teams who perform various activities such as eCRF design and database build to operationalize the protocol. The narrative format of the protocol is vulnerable to human interpretation, and consequently there have been numerous efforts to structure the protocol and thereby promote re-use of common elements, reduce ambiguity, and enable automation of downstream processes.

A truly digital clinical trial dataflow requires modeling of certain components of the protocol. Key examples include objectives, end-points, inclusion / exclusion criteria, trial design elements, and the schedule of assessments. These components share a common thread in that when structured in a machine-readable form, they can be used to drive downstream efficiencies and automation. We'll spend the rest of this paper discussing the potential advantages gained from digitizing key aspects of protocol and linking them to clinical trial artifacts.

STRUCTURING THE PROTOCOL: EXAMPLES FROM INDUSTRY

TransCelerate's Common Protocol Template (CPT) provides perhaps the most widely known example a more structured approach to protocol authoring. The CPT initiative seeks to streamline the process for protocol development and has brought 18 biopharma companies together with the NIH and FDA to align on a harmonized template for clinical trial protocols. This has been very important to drive agreement on **what** needs to go into the protocol and **where** to put the information. The CPT also takes things a step further towards structuring protocol content by providing key protocol information in a semi-structured XML format for downstream consumption. There is more work to be done, but the TransCelerate CPT initiative is a great example of the industry coming together to collaborate and move towards a more structured clinical trial protocol (Alsumidaie, 2017).

Other efforts have also been taken in the industry. Merck has built upon the CPT's foundation to develop an in-house structured authoring tool (Allred, 2017). They reported key benefits in re-using standardized content and thereby increasing consistency and quality. Roche provides another example; they have developed a tool that provides a template-based approach to producing a schedule of assessments that promotes adoption and adherence to standards (Willson, 2013). Similarly, Bayer reported the development of a Protocol Designer tool that links standards to Protocol Activities to enable automated standards setup for a study (Velurethu & Pangritz, 2015).

The key point to takeaway is that there is a recognized need within the industry for more structured protocol content, and that a variety of efforts in this space are ongoing.

DIGITIZING THE DATA FLOW: ADVANTAGES

PROMOTE DECISION MAKING

Many elements of the protocol are implicitly related. For instance, objectives are measured through specific end points. These end points should be traceable to the CRF forms where data to support them is collected and to the analyses (tables, listings, and figures) where they are evaluated. Structuring each of these elements allows for explicit traceability between them and sets the foundation for exploring a variety of important questions such as: (1) Will all collected data be presented in at least one analysis? (2) Does the value of the data justify the cost of obtaining it?

These are important questions as clinical trial protocols grow more complex. The Tufts Center for the Study of Drug Development recently published a paper in *Nature Reviews Drug Discovery* that describes the rising complexity of clinical trials. The authors compared a set of protocols from 2001-2005 and from 2011-2015. They found that Phase III protocols saw an average increase of 70% in the number of total procedures and of 59% in the number of distinct procedures conducted in each protocol over that timeframe (Getz & Campo, 2017).

The rising complexity of clinical trials has implications both for the cost of running the trials as well as the time required to conduct them. Getz et. al. report in a separate paper that nearly 25 % of procedures performed in Phase III protocols and 18 % of those performed in Phase II protocols support secondary, tertiary or exploratory end-points. The authors estimate that the direct cost of performing these 'non-core' procedures is on the order of \$3.7 billion annually (Getz, et al., 2015).

Protocols are growing increasingly complex, and the number of procedures performed in each protocol has increased. Much of the collected data may not even make its way into a New Drug Application (NDA). Digitizing the protocol and enabling explicit traceability from the protocol to analyses brings this to the fore and will promote conversations surrounding the value of the data collected versus the cost to both the patient and the sponsor.

REUSABILITY, EFFICIENCIES, AND AUTOMATION

Certain tasks in a clinical trial require re-using / reformatting information that is already provided within the protocol. One common example includes providing trial metadata to clinical trial registries. The requisite information is already present in the protocol, albeit in an unstructured format. By structuring this information at the outset of the trial, you can reduce the need to re-enter the information when it is needed for additional purposes.

Another example lies in the generation of trial design domains. The information needed to create these domains comes directly from the protocol, but generally needs to be re-entered and perhaps interpreted by the statistical programmers generating the domains. Structuring the trial design information provided in the protocol to enable automated generation of the trial design domains improves both the efficiency and consistency with which these datasets can be produced. Indeed, Rho has implemented just such a system that models the trial design elements and supports automated generation of the SDTM trial design domains (Abolafia & Dilorio, 2016).

A structured representation of the protocol schedule of activities can also enable automation of the study specification. There is considerable overlap in the assessments used across studies. Significant progress has been made both across the industry and at specific sponsors in developing standards for data collection. These standards serve both to ensure consistency across studies and to promote re-use of common assets. Nevertheless, making efficient use of standards remains a challenge in many cases. Modeling protocol assessments and linking them to corresponding standards can simplify the use of standards by study teams.

For instance, consider the scenario where you need to create a schedule of assessments for a new protocol. In a future state process, a clinical scientist could select protocol assessments from an evolving library of assessments that are linked to one or more corresponding standard eCRF forms (Figure 1). Using the example provided in Figure 1, if the clinical scientist chooses to conduct a Vital Signs assessment for an Oncology trial, the system could deterministically predict the correct form (VS001) to use for this specific trial. The example is simple but generalizes well to more complex scenarios. Indeed, using new technologies it should be possible to build a recommender engine to suggest the appropriate standards to use for a given trial, or to identify gaps in existing standards where additional work must be done.

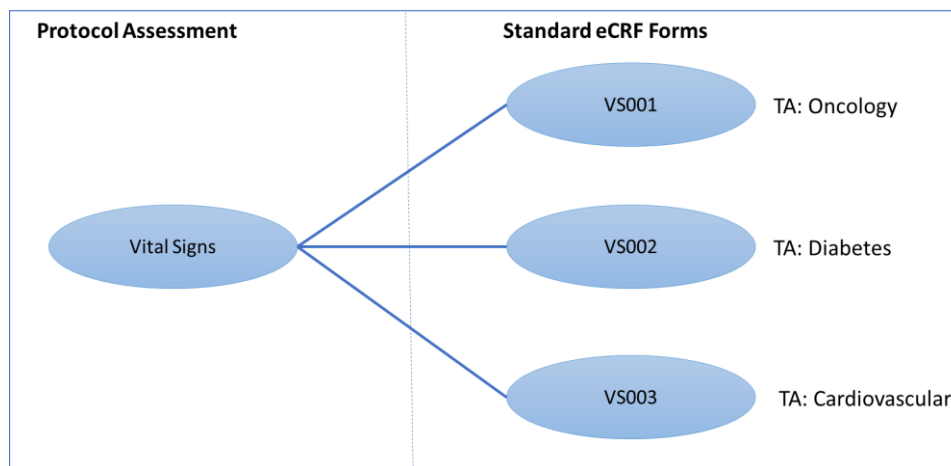


Figure 1: Linking Protocol Assessments to Standards

These examples provide evidence of the efficiencies that can be gained by modeling certain components of the protocol and using these structured components to drive efficiencies and automate manual processes.

CONCLUSION

A digital representation of the clinical trial protocol provides significant advantages in the setup of a clinical trial. It can improve decision-making by supporting explicit traceability between the protocol and downstream artifacts. Furthermore, a digital protocol would enable automation of tasks as disparate as creating the study specification for an EDC system and generating SDTM Trial Design domains. Streamlining the clinical data flow is an ongoing effort and will take concerted effort from the industry to make a reality. Moving towards a digital representation of the protocol is a key step in the direction of a more digital future in pharmaceutical research and development.

REFERENCES

- Abolafia, J., & Dilorio, F. (2016). *Protocol Representation: The Forgotten CDISC Model*. Retrieved from lexjansen: <https://www.lexjansen.com/phuse/2016/cd/CD01.pdf>
- Allred, M. H. (2017). Realizing the Benefits of Structured Content Management in a Stepwise Fashion. *Medical Writing Communication Conference*. Orlando.
- Alsumidaie, M. (2017, June 2017). *TransCelerate Establishes Technology Common Protocol Template and Collaborates with the FDA / NIH*. Retrieved from Applied Clinical Trials Online: <http://www.appliedclinicaltrialsonline.com/transcelerate-establishes-technology-common-protocol-template-and-collaborates-fdanih>
- Getz, K., & Campo, R. (2017). Trial watch: Trends in clinical trial design complexity. *Nature Reviews Drug Discovery*, 307.
- Getz, K., Stergiopoulos, S., Marlborough, M., Whitehill, J., Curran, M., & Kaitin, K. (2015). Quantifying the Magnitude and Cost of Collecting Extraneous Protocol Data. *American Journal of Therapeutics*, 117-124. doi:10.1097/MJT.0b013e31826fc4aa
- Velurethu, A., & Pangritz, S. (2015). "Connect and Create" Connect Standards and Enable Protocol Authoring. Retrieved from lexjansen: <https://www.lexjansen.com/phuse/2015/dh/DH05.pdf>
- Willson, A.-L. (2013). *Supporting End to End Standards in Study Set Up*. Retrieved from lexjansen: <https://www.lexjansen.com/phuse/2013/pp/PP15.pdf>

CONTACT INFORMATION

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

Silas McKee
Accenture – Pennsylvania, USA
Silas.a.mckee@accenture.com