

"THE PROTOCOL IS AT THE HEART OF EVERY CLINICAL TRIAL. IT IS THE PLAN; IT IS A CRITICAL DOCUMENT FOR EVERYONE INVOLVED IN THE CONDUCT OF THE TRIAL"

— Rebecca Kush, President of CDISC, asks 'can clinical trial protocols be standardized?'

The protocol is used in selecting investigative sites and designing the data collection tools; it describes how to treat and evaluate the trial subjects; it serves as a reference for monitoring and auditing trial conduct, and it conveys the plan for analyzing the data when the study is complete. Institutional Review Boards (IRBs) or Ethics Committees use the protocol as the basis for approving whether a trial can be initiated. A well-constructed protocol can ensure common understanding of the study objectives and procedures to be implemented, thereby improving quality and saving time and effort for those using it. Clearly, it is one of the most important documents used in clinical research.

However, the development of a protocol can consume significant company resources and time, particularly when the review group is large or the review process is complex. Leveraging technology can streamline aspects of this process and/or be used to evaluate

the integrity within a protocol before it is finalized. However, to develop such an application requires that at least certain portions of the protocol be 'machine-readable' as well as 'human-readable' and implies at least some commonality across protocols.

The value of being able to leverage technology in the clinical protocol arena and to be able to reuse sections of the protocol for such purposes as tracking databases, end of study results and reporting, regulatory submissions and other related trial documentation has been recognized, particularly by medical communication experts and others in the project team. This prompted the initiation of a project within healthcare and clinical research standards development and research organizations to develop a protocol representation standard. Over the past three years, a collaborative, multidisciplinary project group has made significant inroads towards this end.

PROTOCOL: A DEFINITION

The International Conference on Harmonization defines a protocol as: "a document that describes the objective(s), design, methodology, statistical considerations and organization of a trial. The protocol usually also gives the background and rationale for the trial, but these could be provided in other protocol referenced documents. Throughout the ICH GCP Guideline, the term protocol refers to protocol and protocol amendments."

The goal: a 'standard' protocol representation

The majority of protocols are unique. This is the nature of clinical research. This may give the perception that standardizing protocols is impossible or that this would compromise the science behind the research. However, a review of multiple protocols (even across companies) demonstrates that there are certain sections and informational content found in all protocols. 'Standardizing' the protocol refers to identifying these common elements across protocols, defining them carefully and creating a list or a 'library' that provides for reuse from one protocol to the next or from the protocol to other documents such as the final clinical study reports. It does not imply that the trials are alike. In addition, there is no desire to restrict the trial design or the science employed; the intent is to provide a standard way to represent protocols so that a machine can identify certain common elements.

Primary concepts that form the basis for achieving a standard protocol representation include the following:

- a) There is a base set of common elements across protocols and these can be clearly identified and defined such that they can form a 'data layer' that can populate a database.
- b) The protocol represents the plan for a trial (not the actual results, although the results can also populate the same or

another database).

- c) The protocol is not only the common Word or pdf file that can be printed out in a paper format, but also an analogous computerized representation that is 'machine-readable', but not necessarily ready for human use.

The protocol representation group

The protocol representation standard was initiated as a project within the Health Level Seven (HL7) Regulated Clinical Research Information Management (RCRIM) Technical Committee with leaders from the US Food and Drug Administration (FDA – Kathy Hollinger) and the Clinical Data Interchange Standards Consortium (CDISC – Rebecca Kush). Domain experts, including medical communication specialists, statisticians and project managers, were recruited from CDISC member companies to augment the technologists in the group and to provide direct experience with protocol development for regulated clinical trials. (Note that 'protocol' in the healthcare setting often refers to a treatment plan, which is substantially different from a clinical trial protocol.)

The resulting group, the Protocol Representation Group (PR Group), is currently both an HL7 RCRIM project team and a CDISC team. It now includes representatives from the National Cancer Institute and 'observers' from FDA and EMEA, in addition to representatives of HL7 and CDISC. It is a multidisciplinary, representing the major types of stakeholders in clinical trials. The initial scope statement of the group was to "identify standard elements of a clinical trial protocol that can be further elucidated and codified to facilitate study design, regulatory compliance, project management, trial conduct and data interchange among consumers and systems." However, this scope statement has now been expanded "to develop a standard structured protocol representation that supports the entire lifecycle of clinical research protocols to achieve semantic interoperability (the exchange of content and meaning) amongst systems and stakeholders."

Progress and next steps

After much exploration and debate, the standardization of protocol representation was finally tackled by developing a set of decisions on the approach and a set of assumptions on what the resulting model should be.

The decisions:

- Development of the model/standard should concentrate on content first and implementation second.
- Elements must be defined in a glossary, since the industry uses multiple definitions for the majority of protocol elements.
- Identify a core set of elements initially, and expand with further detail, as needed.
- The initial set of elements will be based on ICH E6 and ICH E3 documents, which focus on efficacy and safety trials, but can be applied to other types of studies.



Assumptions included (as examples):

- Structure and content of the model/standard should be intuitive and clearly understandable to industry stakeholders familiar with clinical trial data and should have straightforward and easy to follow rules.
- The model/standard should be sufficiently flexible that it could be applied to any clinical trial.
- The model/standard should allow some degree of flexibility in the way that some information may be represented to support differing preferences within the industry.
- The model/standard should not be limited to any one specific implementation and so risk rejection by industry stakeholders.

A spreadsheet of elements was created, with the section headers reflecting those from the ICH E6, under which were added sub-sections and then elements. Each element was further elucidated with a glossary definition, source of the element (e.g. ICH, EudraCT, suggested codelists and attributes, cardinality, use case application and other relevant information. This spreadsheet was used to develop an extensive glossary for the protocol and also for an initial modeling attempt to develop an HL7 Clinical Document Architecture (CDA) for a Structured Clinical Trial Protocol (SCTP). This modeling was an education for the Protocol Group, which generally had experience in protocol development but not in the development of HL7 models or messages. Through this effort, the PR Group learned:

- a) It was not yet clear at that point how the protocol should best be modeled for a standard (CDAs were typically for smaller

ABOUT CDISC

The Clinical Data Interchange Standards Consortium (CDISC) is a global, non-profit organization that has developed standards to support the acquisition, exchange, submission and archive of clinical trial data and metadata to streamline clinical trials and improve data quality and patient safety. Its mission is to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare.

Initiated as a volunteer group in 1997 and incorporated in 2000, CDISC currently has over 170 corporate member companies globally, including biopharmaceutical companies, technology and service providers, academic research institutions, and others. CDISC standards are platform-independent, vendor-neutral and openly available at no cost to all (www.cdisc.org). CDISC achievements are due primarily to the concerted efforts of collaborative and multidisciplinary volunteer groups.

CDISC standards (SDTM and define.xml) are endorsed by the FDA as specifications in Guidance for eSubmissions and support their implementation of the ICH eCommon Technical Document (eCTD). CDISC has production versions of the Study Data Tabulation Model (SDTM), define.xml, Operational Data Model (ODM) and clinical laboratory (LAB) models, and there is a General Considerations document for submission of analysis datasets and examples of analysis dataset models (ADaM). The ODM defines format; it is an XML schema. Define.xml defines common metadata and SDTM, ADaM and LAB define content and metadata.

There is a CDISC-HL7 standard in progress for protocol representation (associated article) and there are concerted efforts underway to define common terminology across the CDISC standards. The CDISC Roadmap defines the means to create one CDISC standard from these existing component models. The vision is to map the CDISC standard to the BRIDG model as the portal to healthcare and the means to achieve the mission of interoperability among medical research and healthcare systems.

documents, the RIM was not well-understood by domain experts, and a new version of the CDA – perhaps more appropriate for protocols – was in development).

- b) The protocol group must initially focus on the protocol as the plan only, not the eventual instantiations and permutations (e.g. amendments).
- c) There must be a prioritization of use cases.

The BRIDG model

Concurrent with the initial model development efforts described above, the CDISC organization had initiated the development of an overarching model that would represent the clinical research domain. Following the HL7 Development Framework methodology, which uses unified modeling language (UML), and with the assistance of an HL7 expert, the vision behind this

domain analysis model was to harmonize clinical research (e.g. CDISC) standards among each other and to harmonize the clinical research standards with those of healthcare. The HL7 Development Framework provides a means of eliciting domain expertise from those who may not fully comprehend the HL7 Reference Information Model and representing it in a UML diagram using verbiage that the domain experts do comprehend. The National Cancer Institute (NCI) became instrumental in collaborating with CDISC to progress this vision.

The clinical research domain analysis model was eventually named the Biomedical Research Integrated Domain Group (BRIDG) model because it will not only bridge various standards within the clinical research domain, but it is already bridging organizations. It has now been adopted as the HL7 RCRIM domain analysis model and has been a truly collaborative project among CDISC, NCI, HL7, FDA and others. It is an open model that is now being used in numerous implementation projects by CDISC, NCI, FDA and HL7 RCRIM. (See www.bridgproject.com.)

At the very first BRIDG modeling session, CDISC participants (Directors from the Board) realized that the protocol is at the very central point of clinical research and that is where this domain analysis modeling began. Since the Protocol Representation Group was in a quandary as to the best way to model the protocol, it was decided that the group would focus their efforts on furthering progress for the BRIDG model. Their initial activity was to represent each element from the common elements spreadsheet they had developed in the BRIDG model. However, they also realized that this modeling forced them to capture the relationships among the different elements. These relationships are important to articulate and to build into the model when developing a machine-readable version of the protocol.

Priority use cases and implementations

With the domain analysis modeling efforts under way through CDISC and NCI, the PR Group also prioritized their use cases and focused on the protocol as the plan. Of nearly a dozen use cases, they identified three top priorities:

1. Represent the CDISC Study Data Tabulation Model (SDTM), including the Trial Design, Inclusion/Exclusion Criteria, Planned Assessments, Planned Interventions and Statistical Analysis. Rationale: The SDTM is a published CDISC standard that has been referenced in FDA Guidance for eSubmissions. If one is going to submit data/information to regulatory authorities using SDTM, a standard protocol should reflect this information in an analogous fashion. The FDA and NCI are implementing a cross-trial data warehouse that will include planned vs actual, or protocol and results.
2. Develop standards for the trial tracking/summary/registry sections of a clinical trial. Rationale: This should encourage a harmonization opportunity for many currently separate activities that are requiring collection and reporting of similar information content, e.g. EudraCT, WHO International Clinical Trial Registry Platform, clintrials.gov, SDTM Study Summary. This content is

typically contained as common elements within the protocol.

3. Develop a common representation for the machine-readable protocol document. Rationale: This is the ultimate goal of the PR effort – to have an automated way to use information within the protocol in databases to reduce re-entry and rework while improving quality.

The protocol group has focused thus far on providing their domain expertise through the identification of common protocol elements, developing definitions of these elements, and providing modeling information. These activities fold into the HL7 Development Framework. The next step will be to develop an XML schema/HL7 specification and message, initially for the trial tracking/registry use case. Eventually, the entire protocol will be modeled to provide outcomes such as the re-use of protocol elements in study reports or other clinical documents, population of management tools, population of cross-trial data warehouses such as the one that NCI and FDA are implementing, and to develop case report forms or data collection tools. However, a structured protocol will not be limited to these potential uses; the opportunities are numerous when considering opportunities for streamlining a clinical trial with a machine-readable protocol.

The protocol representation group has spent the past several months representing their elements in the BRIDG model. They are now focusing on specific areas and actually developing implementations based upon the BRIDG model, including one for the NIH-NCI caBIG (cancer Biomedical Informatics Grid), for the Department of AIDS (NIH-NIAID) and certain proprietary applications being developed by biopharmaceutical companies and technology providers.

Looking forward

Significant progress has been made to develop a standard for protocol representation. Certain companies are already developing technologies using the current BRIDG model and the elements in the PR Group's protocol element spreadsheet to develop applications for national academic research organizations and proprietary uses.

Work remains to be done, however, to develop an open, industry-wide, global standard for a structured clinical trial protocol. The BRIDG should continue to be refined and expanded, the trial registry/tracking specifications will go through an open review and consensus-building process (CDISC), after which a message can be developed and balloted through the HL7 process. The entire set of specifications for the protocol needs to be modeled and balloted.

Benefits are already being observed, with a common glossary available (www.cdisc.org/glossary/index.html) and the basic BRIDG model ready for initial implementations. The anticipated future opportunities and benefits, along with progress to date, have brought together a strong and talented group of experts and leaders from CDISC, NCI, HL7 and other organizations who have devoted substantial time and have contributed tirelessly to this open, collaborative initiative. ■