

# Extracting the value of standards: The role of CDISC in a pharmaceutical research strategy<sup>\*1</sup>

Frank W. Rockhold<sup>1)\*2</sup>      Simon Bishop<sup>2)\*3</sup>

1) Global Clinical Safety and Pharmacovigilance, GlaxoSmithKline

2) Data Stewardship Operations, GlaxoSmithKline

## Abstract

Regardless of your role in drug development (Clinical, Statistics, Data Management, Safety, etc.), the importance of information that is of high quality and open to immediate and reliable access has never been more apparent. The ultimate utility of the information is directly related to how it is collected, stored and able to be located. That is to say, it has little value if not retrievable. The need for clear, useable and available standards to accomplish this is essential. The true value of standards is only seen when they are absent and information cannot be used. Not only is information our most expensive and valuable resource, when patients are asked to sign an informed consent to enroll in one of our clinical trials, they are assured the information gathered will be used to better the health of future patients. If we cannot store, find and retrieve the information in a reliable way, not only are we wasting valuable time and money, we are not meeting our commitment to patients or supporting the research goals. CDISC (Clinical Data Interchange Standards Consortium) is an organization that was created and is tasked with being in the middle of solving this most critical piece of transforming the data from trials into a true information-based model that will allow subject matter experts to derive practical wisdom on behalf of patients. This paper outlines both at a strategic and tactical level how standards in general and CDISC in particular help us be more efficient in our handling of research data and sustain valuable contributions to science.

## Key words

clinical research, clinical trials, data standards, regulatory submissions, data exchange

---

\*1 Japanese translation of this article originally written in English is included in *Rinsho Hyoka (Clinical Evaluation)*, 40 (1): 91-6. English version is published only in the journal web-site ([http://homepage3.nifty.com/cont/40\\_1/p91-6eng.pdf](http://homepage3.nifty.com/cont/40_1/p91-6eng.pdf)).

\*2 Past Chair of the Board, CDISC

\*3 Member CDISC advisory Board

## 1. Introduction

Regardless of your role in drug development, (Clinical, Statistics, Data Management, Safety, etc) the importance of information that is of high quality and open to immediate and reliable access has never been more apparent. The ultimate utility of the information is directly related to how it is collected, stored and able to be retrieved. That is to say, it has little value if not retrievable. The true value of standards is only seen when they are absent and information cannot be used.

Many, if not most, pharmaceutical companies—including GlaxoSmithKline (GSK)—have engaged in, or are engaging in, overhauls of their data collection and storage systems. These all aim to reduce and standardize the myriad systems that exist in clinical development. By simplifying our systems landscape, we expect to greatly improve the speed at which data are available. Many are turning to “off the shelf” systems to reduce development time and compliance efforts. In fact, most vendors prefer to work with customers who employ commonly-accepted standards for the same reasons of efficiency and cost. That, in turn, will enhance the quality and consistency of data and documents, resulting in delivery of products that reach patients sooner, at a lower cost, with less effort involved for clinical teams.

Systems process improvement and standards development have shared goals of enabling companies to reliably find, understand, and use clinical trial data to deliver differentiated medicines of value and sustain them in the marketplace. This article intends to provide support to the importance of the concept of applying standards and the need for dedicated staff to develop and maintain them.

The data resource we build in the development

process is *the* major asset we build in R&D (research and development). It is the end product of every experiment and clinical study that we perform. We need it for decision making internally, achievement of regulatory approval and reimbursement, monitoring and ensuring the safety of our medicines in the market place. Ensuring the appropriate governance and integrity in our use of data is part of the commitment we make to the investigator and the patient when we undertake a study with them. We commit to the use of information to further benefit future patients in a standard informed consent form.

Our data also offer real opportunity, for example, to build better experiments and studies using existing data and to plan recruitment strategies. Competence in design, rigour and use of our data is both an essential for GSK and the source of potential productivity and innovation opportunities. In addition, the ability to do this is a critical part of simplification and trust and transparency initiatives in place across our industry. It is difficult to maintain trust without transparency, both inside and outside of an organization, and in turn it is hard to be transparent without reliable access to information (a.k.a. standards). Standards further enhance our, ability to in-license and out-license development compounds, outsource data collection and management, share patient level data, or engage in other data transfers.

## 2. Principle strategic focus of standards

There are three major areas we need to focus on to meet the challenges and take the opportunities: acquiring the right data, making sure we can consistently store and find them and, finally, understanding and being innovative (e.g., re-use) with the data resources that we have.

First, we have to ensure that we collect the right data. This sounds obvious, but various analyses over the past few years have demonstrated that 20 to 40 percent of data items collected may not have been required and were never used in drawing inferences about our trials. Much of the cost reduction in GSK's programs was achieved by reducing patient visits by one-third and eliminating a number of laboratory visits. Our information planning must follow on from medicines development strategy and development plans to build a coherent information resource based on the right science.

Second, we must ensure that we can find and understand our data. We have many examples in which 70 to 80 percent of the cost and effort of answering regulatory or safety queries has been spent finding, retrieving and shaping the data so they can be used (or re-used) to address the question. Standards are essential for industry to meet regulatory and other stakeholder demands and to reduce the cost of data ownership – in acquisition, retrieval and use of data. Standards and systems are necessary to achieve the goals above, but personal accountability and competence are also essential. All researchers must behave as though they personally own this valuable resource. The behavioural shift needed to make this work is not to be taken lightly. From the most senior manager to the front line scientists, the value and access to our information must be held sacrosanct. This takes skill in communication—all the way from the senior manager who must allocate the time and money to achieve the goal to the front-line scientist who must assume that someone, 5 – 10 years from now, will need a clear path to retrieve and reuse the information.

Finally, we need to make the best use of our data. We are using our data in regulatory applications and in support of our medicines reimbursement efforts, but this is just the beginning. We

already use our data for in-stream detection of safety signals. There are pockets of people using (or re-using) clinical trials data for many purposes ranging from planning clinical trials, to reviewing information within classes of compound to boost early development, to validating the designs in new toxicology screens to reduce the usage of animals in safety studies. New search technologies and improved data resources will offer many more opportunities. We have to know as much about our medicines as we possibly can to help our patients.

A global non-profit organization called the Clinical Data Interchange Standards Consortium (CDISC) was formed 15 years ago to create an environment to enable researchers to deliver on the need expressed above.

### 3. CDISC

The CDISC ([www.cdisc.org](http://www.cdisc.org)) mission as described on the website:

“CDISC is a global, open, multidisciplinary, non-profit organization that has established standards to support the acquisition, exchange, submission and archive of clinical research data and metadata. **The CDISC 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.** CDISC standards are vendor-neutral, platform-independent and freely available via the CDISC website”.

This organization is the sole Standards Development Organization devoted to clinical research. The CEO, management team and Board of Directors represent some of the brightest and most experienced minds involved in the standards business. They are the standards selected by regulators (e.g. FDA), health authorities (e.g. NCI) and professional societies to drive their research and health-

care agendas.

The Strategic Themes of CDISC are:

- Achieve significant progress in the use of CDISC standards to allow scientifically sound **data aggregation** and support secondary uses of research data for the purposes of scientific investigation and comparative effectiveness.
- Achieve significant progress in enabling **interoperability between clinical care and clinical research**, and explore expansion from bench to bedside (translational research); accelerate the cycle through which healthcare informs research and research informs clinical decisions.
- Develop CDISC SHARE, a global, accessible, Shared Health And Clinical Research Electronic library for CDISC content/**semantics** that will enable precise and standardized data element definitions and richer metadata that can be reused within applications and across studies to improve biomedical research and its link with healthcare.
- Expedite the development and rollout of new **therapeutic-area or specialty standards, while continuing to refine, support and educate on existing/foundational standards**, to ensure consistency in data capture and analysis related to efficacy in addition to patient safety.
- Leverage our global, non-profit, vendor-neutral, independent status to forge productive **collaborations** with and provide **value** to key stakeholder communities.

In order to accomplish this “higher” purpose, operational skill, experience and focus are essential to create a robust and sustainable product. What follows is an example of the time, intellectual investment, and discipline it takes to make this happen. This is taken from the efforts at GSK, but would be a good example of the process and

approach needed generally.

#### 4. The CDISC approach at GSK

While this may differ somewhat from what other companies have done, this represents one current approach to delivering on the strategy outlined above. Since its merger in 2001, GSK has valued the importance of the information model and rapidly evolved to driving a standards model which had begun to evolve in each of its two parent companies. Each employed the standards based on the CDISC Submission Data Standards (SDS) available at that time, or developed employing the principles that underlay the SDS standards. The GSK standards, like SDS, focused on data collected across multiple kinds of studies, primarily safety data. GSK therapeutic areas developed additional standards or collected data in one-off study-specific datasets. After the release of the Study Data Tabulation Model (SDTM) standards, new GSK standards were developed following the principles that underlay the SDTM standards. The standards comprised paper and electronic case report (CRF and eCRF) forms, collection datasets and in some cases analysis datasets and data displays (mainly tables and listings). GSK currently has about 60 “core” standards, standards for over 250 questionnaires and around 400 “therapeutic” standards.

FDA’s increasing desire for the provision of SDTM and Analysis Dataset Model (ADaM) datasets led GSK to examine how best to incorporate SDTM datasets into the clinical study process. We found, as others had, that there were two approaches available: the incorporation of “SDTM aligned” datasets into the overall data flow or a back-end process that mapped existing datasets to SDTM.

As part of tactical delivery of SDTM, we created mapping specifications from our standards to SDTM. This was time consuming and exposed the

limitations of both the GSK standards and the SDTM standards.

A review of GSK's 700+ datasets and 25,000+ variables revealed issues of duplication and ambiguity, and our mapping to SDTM demonstrated that different studies had used some of our variables in unintended/inconsistent ways. We also knew that although GSK's dataset-based standards did deliver benefits, they provided only limited opportunities for automation and aggregation of data across studies and across compounds was often difficult.

The limitations of SDTM as a standard include the absence of therapeutic data domains, the need for a 300-page implementation guide, and its broad unsuitability as an operational standard.

At this point, we realised that we want more than just successful implementation of SDTM, ADaM and their metadata (in the CDISC define.xml format). We want

- quicker study set-up and reporting
- easier aggregation and re-use of all data
- easier sharing of data with partners (e.g. development partners, CROs and FDA)
- the ability to implement new systems without completely redeveloping our standards
- greater capability for automation.

Individuals from Lilly, Novartis and GSK recognised the limitations of SDTM at much the same time and triggered the development of CDISC SHARE. SHARE is based on two industry standards, the Biomedical Research Integrated Domain Group (BRIDG) model and the ISO21090 data type standards) and in time will underpin SDTM and Clinical Data Acquisition Standards Harmonization (CDASH), and help speed the creation of future SDTM domains. GSK remains involved in and committed to the development of the CDISC SHARE.

Following GSK's decision to replace the whole clinical computing toolset with best-of-breed systems, we began development of new GSK standards

based on the same approach as that of CDISC SHARE, recognising that SHARE would not be available in time for launch.

The GSK approach is to separate the definition of clinical data from the creation of operational objects (e.g., eCRFs, datasets). There are four layers:

1. Definitions of clinical concepts (e.g., a systolic blood pressure observation; a height result) together with the identification of all component parts (e.g. date body position). It should be possible for these to be used across the whole pharmaceutical industry.
2. Terminology used for component parts (e.g., a set of valid values for the body position component of the systolic blood pressure observation). It is desirable that these be industry standard.
3. Groupings of individual clinical concepts (e.g., those comprising the set of vital signs). It may be difficult to agree this classification across the industry.
4. Standard definition of operational objects (e.g., as eCRFs, SDTM datasets, company specific datasets). Of these, only CDASH modules and SDTM datasets can be standardised across the industry.

Our expectation is that CDISC SHARE will define and hold the industry standard content.

The value we obtain from the two industry standards that underpin the approach is significant:

- The standards provide the basis for templates from which the data items for a particular clinical concept can be selected, eliminating duplication and ambiguity.
- The relationships between many data items are automatic (e.g., the relationship between a value and its unit or between a heart rate observation and the body position of the subject during the observation).

- Other relationships can be documented as part of the definition of a clinical concept (e.g., that a lab test is performed on a specimen, while a heart rate observation is performed on a person) and there is the facility to document relationships between clinical concepts at the study level (e.g. heart rate measurements were taken during an exercise test).
- Using a mapping between SDTM and BRIDG, SDTM datasets can be auto-generated, eliminating inconsistencies in the current processes. Mappings to future versions of SDTM will be accomplished through updates to the SDTM-BRIDG mapping.

Our target at GSK is to use the clinical concepts described above within our new eProtocol tool to start developing the study's time and events. This information will then be extracted and refined, and terminology added resulting in a detailed study specification independent of any of the GSK consuming systems but from which such specifications as an eCRF specification, a data management dataset specification (for data cleaning) and an SDTM specification will then be generated. ADaM datasets will use the SDTM datasets as their starting point. For re-use purposes, the study data will also be held in a "clinical concept aligned" form auto-generated using the detailed study specification.

This process will enable us to optimise study execution whilst providing the ability to deliver quality datasets (SDTM and ADaM) to FDA. Long term, cross-compound data aggregation and re-use will be made quick and simple. Using a largely metadata-driven approach will ensure that traceability is built into our processes rather than being an expensive afterthought. Ensuring that the operational objects can always be mapped back to the clinical concepts will be the key to success of the new GSK standards. Our expectation is that we will migrate from GSK SHARE-like content to

CDISC SHARE content when that becomes available.

## 5. Governance and maintenance of standards

In order for standards to retain their value over tens of years, the objects that make up those standards must have unambiguous definitions that do not change over time and must not be tied to any particular technological solution. New objects should be added when new indications or new scientific information are to be included in the standards and, when scientific knowledge advances, it may be necessary to "retire" out of date objects. Controls need to be in place to avoid the creation of multiple objects with essentially the same definition. Strong governance is necessary to ensure that the requirements detailed above are met, in order to protect the long term value of the information we collect from the patients who volunteer to be in our trials.

The inherent capability of the standards is also important. Clinical research is complex and the standards need to be able to represent that complexity in a usable form. The creation of independent clinical data elements is not sufficient. For example, understanding the relationships between separate pieces of information is important to ensure appropriate utilisation of that information.

To gain maximum scientific knowledge from the totality of data collected across the clinical research community, we need to eliminate "competing" standards. The ultimate vision of total interoperability and transparency will only be achieved if people collect and store data in a way that enables others to quickly access and utilise them. We must guard against directing time and rare critical resource onto multiple solutions, as this would add to the burden of researchers and ultimately impede

the help we can provide to patients.

The FDA's "55 in 5" initiative, i.e., developing standards for 55 indications in 5 years, which is being done in partnership between CDISC and the Critical Path Institute, has a laudable aim. CDISC SHARE has the potential to deliver higher value standards in a faster and less resource intensive manner than has been the case for other recently developed therapeutic standards such as those for Alzheimer's and Parkinson's.

## 6. Summary

Information is our most expensive and valuable resource. When patients are asked to sign an informed consent to enrol in one of our trials, they are assured we will use the information gathered to

better the health of future patients. If we cannot store, find and retrieve the information in a reliable way we are not meeting our commitment to patients or supporting the business of R&D and GSK.

CDISC is an organization that was created and is tasked with being in the middle of solving this most critical piece of transforming the data from trials into a true information-based model that will allow subject matter experts to derive practical wisdom on behalf of patients.

### Acknowledgements

The authors would like to thank Robbin Safron, and Drs. Suzanne Markel-Fox and Diane Wold for their help in preparing the manuscript. Their editorial suggestions and direction were invaluable in finishing this work.

\* \* \*