

## COMMENTARY

# Global Standards to Expedite Learning From Medical Research Data

Lynn D. Hudson<sup>1,\*</sup>, Rebecca D. Kush<sup>2</sup>, Eileen Navarro Almario<sup>3</sup>, Nathalie Seigneuret<sup>4</sup>, Tammy Jackson<sup>5</sup>, Barbara Jauregui<sup>6</sup>, David Jordan<sup>7</sup>, Ronald Fitzmartin<sup>3</sup>, F. Liz Zhou<sup>8</sup>, James K. Malone<sup>9</sup>, Jose Galvez<sup>10</sup> and Lauren B. Becnel<sup>11,12</sup>

## INTRODUCTION

Opportunities for meaningful data sharing and maximizing the return on investment of medical research rely on broad adoption of global data standards. Standards for collecting and exchanging data are an often overlooked “infrastructure” aspect of medical research. The value of standardization has been substantiated for data submitted to regulatory agencies to support approval of new therapies. However, inadequate adoption of standards by researchers at the start of a study continues to negatively impact data sharing.

### Standards Provide a Common Language for Medical Research

“Standards are documented agreements containing technical specifications or other precise criteria to be used consistently as rules, guidelines, or definitions of characteristics to ensure that materials, products, processes, and services are fit for their purpose.”<sup>1</sup> Standards for medical research provide a collective knowledge that defines not only how we exchange data, but also how we collect, understand, and use data for a common purpose: to find new therapies for patients. Through standards, we attain a higher level of data quality, achieve consistency and a common understanding of a disease area that is accepted and useful among the medical research industry, regulatory agencies, the healthcare community, and patients.

There are different types and levels of standards for medical research (**Supporting Table S1**). They extend from the level of defining terms (terminology) to the data element level or field of a case report form (CRF) for an individual patient to standards for metadata (data about data) and representation of tables and statistical analyses of data from multiple patients. Research standards from the Clinical Data Interchange Standards Consortium (CDISC), for example (**Figure 1**), begin with the protocol and the study design stage and encompass all research processes. They also cover data exchange standards that carry the metadata for representing an audit trail or provenance information.

Robust standards take time to develop, and there is not typically a “right or wrong”; rather, consensus-based standards with broad input will be more readily adopted and thus more valuable. Ideally, standards should be i) global (not local); ii) open and freely available (not proprietary); iii) based on consensus, as much as possible; iv) developed through an appropriate standards development process; v) authorized by a standards developing organization (SDO); vi) unique and not redundant; vii) adopted widely and endorsed by key stakeholders and endusers; and viii) fit for purpose.

### How Standards for Clinical Research Are Developed

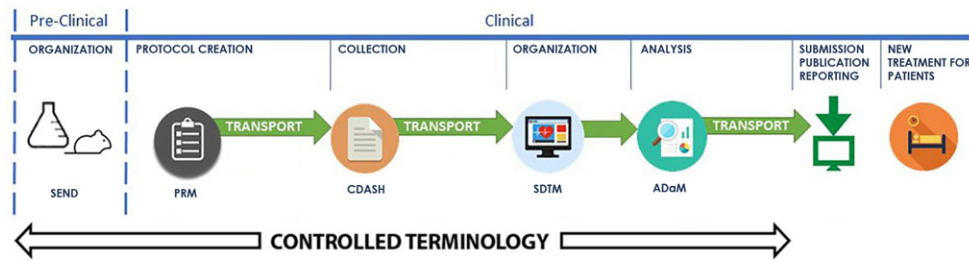
Over the past 3 decades there have been concerted efforts to develop global guidelines and standards for developing new therapies. Specifically, the International Council on Harmonization (ICH) developed technical guidelines and requirements for pharmaceutical product development and registration. To complement the ICH work, CDISC developed global data standards for individual patient data from clinical research studies. CDISC standards support the gamut of medical research (including nutrition, public health, epidemiology, outcomes, and interventional research) and apply to protocol information (study summary), study design, data collection, tabulation, analysis data sets and results reporting, in addition to data exchange. Health Level Seven (HL7) focuses on healthcare standards. Recognizing the overlap between clinical research and healthcare, CDISC and HL7 collaborate on specific projects as well as with other SDOs through a Joint Initiative Council (JIC)<sup>2</sup> to avoid duplicating efforts. Synergistic standards developed through cooperation among SDOs for different purposes can eliminate redundancy and optimize resources.

### How Standards Are Employed From Beginning to End

The maximum value of standards applied to clinical research is realized when standards are applied from beginning to end of a study. Using standards in the planning stages of research will pave the way to meeting data sharing and/or aggregation needs or regulatory requirements, and minimizing the time-consuming mapping at the project's end. This allows

<sup>1</sup>Critical Path Institute, Tucson, Arizona, USA; <sup>2</sup>Catalysis Research, Austin, Texas, USA; <sup>3</sup>Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, Maryland, USA; <sup>4</sup>Innovative Medicines Initiative, Brussels, Belgium; <sup>5</sup>Wilmington, North Carolina, USA; <sup>6</sup>Pan American Health Organization/World Health Organization, Washington, DC, USA; <sup>7</sup>TransCelerate BioPharma, Emeritus, Libertyville, Illinois, USA; <sup>8</sup>Sanofi, Bridgewater, New Jersey, USA; <sup>9</sup>Eli Lilly and Company, Indianapolis, Indiana, USA; <sup>10</sup>National Institutes of Health, Clinical Center, Bethesda, Maryland, USA; <sup>11</sup>Clinical Data Interchange Standards Consortium (CDISC), Austin, Texas, USA; <sup>12</sup>Baylor College of Medicine, Dan L. Duncan Comprehensive Cancer Center, Houston, Texas, USA. \*Correspondence: Lynn D. Hudson ([LHudson@c-path.org](mailto:LHudson@c-path.org))  
Received 20 February 2018; accepted 19 March 2018; published online on 26 April 2018. doi:10.1111/cts.12556

### CDISC Standards in the Clinical Research Process



**Figure 1** CDISC standards provide a common language for clinical and translational research. Research starts with a protocol or experimental plan, though it may be informed by preclinical research results (SEND, far left). Data are collected (CDASH), organized (SDTM), and analyzed (ADaM, right) with reports and summaries generated. Regulated clinical trials also require electronic submission of data to regulatory agencies (Define.xml, SEND, SDTM, and ADaM). CDISC has standards for each of these steps (at bottom) and also a set of data transport and exchange standards (green arrows) to support data flow between different research databases. Ultimately, CDISC’s common language helps researchers make discoveries to improve human health (far right).

for increased return on investment (ROI), more streamlined staff training, decreased opportunities for error, and reduced overall research timeframe for all types of research. In the case of disease outbreaks, for example, having predesigned case report forms can save substantial response time.<sup>3</sup>

#### How Standards Play a Critical Role in Data Sharing

There are several pitfalls to conducting research without standards or with “proprietary standards.” First, the studies cannot be readily aggregated or compared; each study is a one-off—a “silo.” Additional time and resources must be allocated to map the data into a common format such that comparisons can be made; in turn, this mapping step can compromise the data quality, trustworthiness, integrity, and traceability. Interpretation of the initial intent of the data collection can be difficult if not impossible without the original investigator’s input.

The Cancer Moonshot called for sharing data to break down barriers between institutions and maximize the benefits of this knowledge for patients.<sup>4</sup> Yet tapping into the “treasure trove” of data that have already been collected has been hampered by the lack of standards for collecting, organizing, and analyzing cancer data. Efforts such as Project Data Sphere LLC (PDS), a not-for-profit initiative whose platform hosts control arm data from historical cancer trials ([www.projectdatasphere.org/projectdatasphere](http://www.projectdatasphere.org/projectdatasphere)), encountered the challenges of hosting data sets that were constructed with differing standards. Significant resources are needed to remap individual data sets to one standard, and even the most sophisticated algorithms are unable to compensate for missing data or definitions and concepts that are not aligned. Similar challenges were encountered by TransCelerate (<http://www.transceleratebiopharmainc.com/>) for a “data harmonization” project using the CDISC Study Data Tabulation Model (SDTM) (**Figure 1**) to convert and harmonize placebo data and standard of care data from hundreds of studies across many therapeutic areas to create one of the largest historical control databases of its kind.<sup>5</sup>

Repurposed data have contributed to other advances in clinical research. Working with CDISC, the Consortium for Prevention of Alzheimer’s Disease (CPAD; formerly CAMD) developed an Alzheimer’s disease data standard and

mapped the data from nine different organizations into this standard to create an openly available database containing individual records of 6,500 Alzheimer patients who were enrolled in the placebo arms of clinical trials.<sup>6</sup> This initiative revealed shortcomings of mapping legacy data, including the loss of data that could not be interpreted, and emphasized both the value of prospectively applying standards and the need to train researchers to carry out performance tests in a standardized manner. CPAD used the database to develop a clinical trial simulation tool that was qualified by the European Medicines Agency (EMA) and endorsed by the US Food and Drug Administration (FDA).<sup>7</sup> Scientists worldwide (312 researchers to date) draw on the database and apply the clinical trial simulation tool (58 researchers), which could lead to new therapies. Now, when data are collected using the Alzheimer’s data standard, that data can be readily compared with the data in the CPAD database.

#### How the Global Research Community Can Align on the Use of Standards for Research

For many clinical researchers and principal investigators, the value of standards may be clear; yet there is a perception that a dizzying array of available “standards” exists and it is difficult to know which one(s) to choose. To help navigate the standards landscape, the Innovative Medicines Initiative project eTranslational Research Information Knowledge System (eTRIKS) developed a stewardship guide of standards called the eTRIKS Standards Starter Pack ([www.etriks.org/standards-starter-pack](http://www.etriks.org/standards-starter-pack)) to facilitate and increase data reusability, reproducibility, and preservation.

Regulators in the US (FDA)<sup>8</sup> and Japan’s Pharmaceutical and Medical Devices Agency (PMDA)<sup>9</sup> took the key first step in aligning global standards by requiring CDISC standards (e.g., SDTM and ADaM (Analysis Dataset Model), define.xml and therapeutic area extensions) for regulatory submissions. Certain National Institutes of Health (NIH) institutes/centers are tying the use of CDISC standards to funding and to their own internal research systems,<sup>10</sup> and the Innovative Medicines Initiative (IMI2) requests the use of well-established standards such as CDISC. These organizations have found that they can rely on standards to improve process efficiency and/or to enable the use

of various data analytic tools. They have engaged constructively in the consensus-based standards development process employed to develop the global suite of standards for medical research.

The regulatory requirements for CDISC standards ensure that the majority of therapeutic trials conducted worldwide will provide a rich source of reusable data. But numerous academic research studies still apply or create disparate data models or standards, which is particularly problematic in this era of precision medicine. Aggregation of multiple studies is needed to discern differences in response in relevant sub-populations, such as differential responses driven by pharmacogenomics differences.

Progress in personalized medicine will be accelerated through aggregation of relevant trials harmonized to a standard. Patients are both the contributors and the beneficiaries of this global collaboration on data sharing and standardization. The “voice of the patient,” heard through individual patients, caregivers, and disease foundations, has driven the recent surge in data sharing, reinforced by policies adopted by the International Committee of Medical Journal Editors and research funders.<sup>11</sup>

While the benefits of using standards from beginning to end are many and valuable, the costs of nonstandardization are high. Apart from the negative economic impact of the additional time and resources required to evaluate data that are not collected in a standardized fashion, the errors introduced by the failure to standardize, including those errors introduced by waiting until the end of a clinical trial before data are mapped to enable aggregation and analysis, can be considerable. To maximize data sharing and the value gained by data sharing while minimizing the public and private resources needed to support clinical trials, the use of standards should be required as a condition of funding support. Otherwise, the funders pay twice: once to generate the data, and then to remap the data into a format suitable for sharing. Even clinical studies that are not being submitted to the regulatory authorities should adopt data standards, because data on disease progression can then be readily combined with therapeutic intervention data on that disease population, as exemplified by the Alzheimer's disease clinical trial simulator.<sup>7</sup> Both the tools and the policy directives to apply those standardization tools will shape our data-sharing future. Sharing data is key to making progress on cures for the world's major healthcare challenges—cancer, Alzheimer's disease, diabetes, infectious disease, and more—and standardization will accelerate that progress.

**Acknowledgments.** The authors thank Alana St. Clair for organizational support. The authors represent various organizations on the Scientific Advisory Committee (SAC) of the Coalition For Advancing Standards and Therapies (CFAST), an initiative of CDISC and the Critical Path Institute, collaborating with the FDA, EMA, Japan's PMDA, TransCelerate Biopharma, National Cancer Institute Enterprise Vocabulary Services NCI EVS, Europe's Innovative Medicines Initiative (IMI), Association of Clinical Research Organizations (ACRO), and the Clinical Center at the US NIH.

**Funding.** This study was supported in part by the Intramural Research Program of the NIH, Clinical Center, and in part by 1U01FD005855 awarded to the Critical Path Institute by the FDA to develop therapeutic area data standards through the Coalition for Advancing Standards and Therapies (CFAST).

**Disclaimer.** The views expressed in this commentary do not reflect the official policies of the US Food and Drug Administration, or the Department of Health and Human Services; nor does any mention of trade names, commercial practices, or organization imply endorsement by the United States Government. The views expressed in this commentary do not necessarily reflect the positions and opinions of the European Commission or the European Federation of Pharmaceutical Industries and Associations.

**Author Contributions.** L.D.H. and R.D.K. wrote article; E.N.A, N.S., T.J., B.J., D.J., R.F., F.L.Z., J.G. and LB contributed to discussions, Figure 1, and wrote sections of the article.

**Author Disclosures.** F.L.Z. is an employee of Sanofi; J.K.M. is an employee of Eli Lilly; T.J. is an employee of PPD.

1. Standards [Internet]. [cited 2018 Feb 16]. Available from: <https://www.iso.org/standards.html>
2. Welcome to Joint Initiative Council [Internet]. [cited 2018 Feb 16]. Available from: <http://www.jointinitiativecouncil.org/council.asp>
3. IDDO hopes CDISC training will open doors in disease research | Worldwide Antimalarial Resistance Network [Internet]. [cited 2018 Feb 16]. Available from: <http://www.wwarn.org/news/news-articles/iddo-hopes-cdisc-training-will-open-doors-disease-research>
4. Biden (Archives) J. Inspiring a New Generation to Defy the Bounds of Innovation: A Moonshot to Cure Cancer [Internet]. Medium. 2016 [cited 2018 Feb 16]. Available from: <https://medium.com/cancer-moonshot/inspiring-a-new-generation-to-defy-the-bounds-of-innovation-a-moonshot-to-cure-cancer-fbdf71d01c2e>
5. Placebo and Standard of Care Data Sharing [Internet]. Transcelerate. [cited 2018 Feb 16]. Available from: <http://www.transceleratebiopharmainc.com/assets/placebo-and-standard-of-care-data-sharing/>
6. Neville, J. et al. Development of a unified clinical trial database for Alzheimer's disease. *Alzheimers Dement.* **11**, 1212–1221 (2015).
7. Romero, K. et al. The future is now: model-based clinical trial design for Alzheimer's disease. *Clin. Pharmacol. Ther.* **97**, 210–214 (2015).
8. Electronic Regulatory Submission and Review [Internet]. [cited 2018 Feb 16]. Available from: <https://www.fda.gov/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/default.htm>
9. Reviews | Pharmaceuticals and Medical Devices Agency [Internet]. [cited 2018 Feb 16]. Available from: <http://www.pmda.go.jp/english/review-services/reviews/0001.html>
10. Using a data warehouse model based on CDISC SDTM for centralized data and safety reporting for clinical trials | NIH Research Festival [Internet]. [cited 2018 Feb 16]. Available from: <https://researchfestival.nih.gov/2016/posters/using-data-warehouse-model-based-cdisc-sdtm-centralized-data-and-safety>.
11. Taichman, D.B. et al. Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. *Ann. Intern. Med.* **167**, 63–65 (2017).

© 2018 The Authors. Clinical and Translational Science published by Wiley Periodicals, Inc. on behalf of American Society for Clinical Pharmacology and Therapeutics. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Supplementary information accompanies this paper on the *Clinical and Translational Science* website. ([www.cts-journal.com](http://www.cts-journal.com))