Citation: Clin Transl Sci (2021) 14, 214–221; doi:10.1111/cts.12845

ARTICLE

Standardized Data Structures in Rare Diseases: CDISC User Guides for Duchenne Muscular Dystrophy and Huntington's Disease

Ariana P. Mullin[1], Diane Corey[1], Emily C. Turner[1], Richard Liwski[1], Daniel Olson[1], Jackson Burton[1], Sudhir Sivakumaran[1], Lynn D. Hudson[1], Klaus Romero[1], Diane T. Stephenson[1] and Jane Larkindale[1,*]

Interest in drug development for rare diseases has expanded dramatically since the Orphan Drug Act was passed in 1983, with 40% of new drug approvals in 2019 targeting orphan indications. However, limited quantitative understanding of natural history and disease progression hinders progress and increases the risks associated with rare disease drug development. Use of international data standards can assist in data harmonization and enable data exchange, integration into larger datasets, and a quantitative understanding of disease natural history. The US Food and Drug Administration (FDA) requires the use of Clinical Data Interchange Consortium (CDISC) Standards in new drug submissions to help the agency efficiently and effectively receive, process, review, and archive submissions, as well as to help integrate data to answer research questions. Such databases have been at the core of biomarker qualification efforts and fit-for-purpose models endorsed by the regulators. We describe the development of CDISC therapeutic area user guides for Duchenne muscular dystrophy and Huntington's disease through Critical Path Institute consortia. These guides describe formalized data structures and controlled terminology to map and integrate data from different sources. This will result in increased standardization of data collection and allow integration and comparison of data from multiple studies. Integration of multiple data sets enables a quantitative understanding of disease progression, which can help overcome common challenges in clinical trial design in these and other rare diseases. Ultimately, clinical data standardization will lead to a faster path to regulatory approval of urgently needed new therapies for patients.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
☑ Foundational Clinical Data Interchange Consortium (CDISC) standards describe formalized data structures and controlled terminology for elements of clinical data in many clinical trials, but no such standards have been defined for Duchenne muscular dystrophy (DMD) or Huntington's disease (HD), two rare diseases where there is significant, ongoing drug development.

WHAT QUESTION DID THIS STUDY ADDRESS?
☑ This paper addresses the key question of how to efficiently document and integrate clinical research data in DMD and HD. To address this, the paper describes the development of the CDISC therapeutic area user guides for DMD and HD and the use of these data standards in the development of integrated databases containing patient level clinical data.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?
☑ The development of these data structures encourages the standardization of data collection and will allow us to integrate data for rare diseases where little may be known about disease progression. Standardized terminology defined in the user guides for DMD and HD may also be applied or adapted to similar concepts collected in other disease areas, expanding the utility of these guides. In addition to the specific benefit of publicly available standardized terminology for DMD and HD, clinical programs will greatly benefit from the ability to compare, contrast, and aggregate data in these areas.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?
☑ Use of standardized data structures and terminology allow integration of datasets and the ability to compare and contrast datasets, which enhances our understanding of disease natural history, and aids in the development of disease progression models and drug-disease interaction models to optimize clinical trial design and accelerate regulatory review.

[1]Critical Path Institute, Tucson, Arizona, USA. *Correspondence: Jane Larkindale (jlarkindale@c-path.org)
Received: April 9, 2020; accepted: June 14, 2020. doi:10.1111/cts.12845

The ability to readily understand the meaning of clinical data and use standard analytical tools to review such data was the impetus for the establishment of the Clinical Data Interchange Standards Consortium (CDISC) in 1997. The CDISC develops controlled terminology to optimize data capture and quality, and defines standards for structure of the data in a database. This is of particular importance in clinical research and drug development for rare diseases where data are limited and where the wide adoption of standardized data terminology, collection, and tabulation can enable dataset comparison and accelerate regulatory review. Use of such standards can also help integrate small data collections into larger, more informative datasets, which may be needed to detect variance in a small disease population. The CDISC does not dictate to sponsors what data to capture or end points to select in a given study, nor which end points might be acceptable to regulators, but rather provides standard terminology to harmonize the language and format of the data collected and subsequent analysis. The value of CDISC standards is reflected in the requirement for their use in electronic regulatory submissions by the US Food and Drug Administration (FDA) and the Japanese Pharmaceutical and Medical Devices Agency, and preference for their use in submissions to China's National Medical Products Administration.

Orphan diseases offer specific challenges to drug development inherent in the limited numbers of affected individuals. Many rare diseases also include multiple genetic subtypes that may give rise to variability within that disease phenotype. Progression rates may also be affected by external and internal factors, such as age, concomitant therapies, and sex. Due to the variable manifestations and progression rates of diseases, a quantitative understanding of their natural history and factors that affect their progression is paramount to the development of informed clinical trial protocols. Efficient clinical trials are essential to the development of effective treatments for rare diseases, in a reasonable period of time, and at a reasonable cost.

In most rare diseases, there are limited longitudinal data on clinical end points or biomarkers. As a result, trials frequently fail to give definitive answers on the effectiveness of a potential new therapy due to the suboptimal selection of clinical outcome assessments, or suboptimal selection of trial design parameters (e.g., study population, trial duration, frequency of observations, or control arm operationalization[1–4]). To optimize clinical trial design, and allow shorter, smaller trials to be conducted without compromising statistical significance,[5] it is important to utilize the totality of the available data to develop a clearer understanding of sources of variability and the effect on disease progression.

Given the limited numbers of individuals with each rare disease and the few treatments available, natural history data, if available, are limited due to small study cohorts and/or length of data collection. Despite these challenges, multiple small collections of data exist for various rare diseases that have been consented for research use. Together, these have the potential to be integrated into larger datasets that better reflect the broader disease community.[6] Such databases have a much greater aggregate sample size than each individual dataset, which may increase signal-to-noise ratio for analyses and provide orthogonal perspectives to generate quantitative models of the disease continuum. Further, such databases enable the population-level understanding of the relationship between disease variables across disease stages and help interpret changes in outcome measures and biomarkers that reflect the experience of the broader affected population.

Data integration is challenged by the fact that within one disease different studies may or may not include the same measurements, and similar measurements may not be recorded consistently between studies. In order to integrate data in a meaningful manner across natural history and observational studies, clinical trials, and registries, it is essential that the data are in a common format or structure, such that only like elements are combined, and similar elements are kept discrete. This may be achieved ideally by collection of data in a standardized format, or by mapping of existing data to such a format (with respect to both structure and terminology) after data collection.

The database structure and terminology described by the CDISC Study Data Tabulation Model (SDTM) are ideal for integration of datasets.[7–9] Where possible, new data collected should be collected using CDISC Clinical Data Acquisition Standards Harmonization (CDASH) standard format, which describes how to design case report forms that include the needed metadata associated with each data element. If data are collected using the CDASH standards, they may be more readily mapped into CDISC SDTM format. Although data in other formats can be mapped to SDTM, this is a lengthy and costly process and interpretation is required. Accordingly, initial data collection using CDASH with the controlled terminology managed by the National Cancer Institute-Enterprise Vocabulary Service (NCI-EVS) is more efficient than mapping to the structure at a later timepoint. In either case, CDISC SDTM describes the structure of data and relationships between domains and variables, which is essential to determine whether data elements can be combined. For example, the structure allows one to not only identify a measurement that was taken on a sample or subject, but also how a sample or measurement was collected and processed and how the measurement was taken. This allows the user to understand which data elements are comparable between clinical datasets, and which are not. This intra-operability means that data from multiple clinical studies can be compared, contrasted, and integrated in a meaningful way.

The initial CDISC standards (Foundational Standards) address the data elements that are common across the majority of clinical research studies (e.g., medical history, medication history, concomitant medications, and adverse events). The CDISC and its partners have developed CDISC Therapeutic Area User Guides (TAUGs) that describe CDISC data standards specific to various therapeutic areas.[10] The therapeutic area specific standards address those elements that augment the core/foundational data for that specific area, which often includes the documentation of efficacy data for a clinical research study. The CDISC TAUG describes how to document a measurement, but does not prescribe end point or outcome measure selection.

Here, we describe the development of TAUGs for two rare diseases, Duchenne muscular dystrophy (DMD) and Huntington's disease (HD). The use of these standards in the development of integrated databases for the development of drug development tools is described. We make a case for a future streamlined process for development of therapeutic area specific standards for other disease areas by leveraging the work that has already been done, and the need for adoption of such standard terminology for the acceleration of drug development. This is especially needed for rare diseases but is of general value for any therapeutic area for which standards have not yet been developed.

METHODS

CDISC standards for DMD and HD were developed through collaborations between the CDISC and consortia at the Critical Path Institute (C-Path), Duchenne Regulatory Science Consortium (D-RSC), and Huntington's Disease Regulatory Science Consortium (HD-RSC), respectively, according to the methods described in other diseases[7,8,11] and summarized in **Figure 1**. In each case, controlled terminology was developed for predefined measurements commonly collected in clinical research by engaging subject matter experts. Detailed concept maps were developed as needed for the measures and these concept maps are included in the final user guide.[10] Dataset examples of biomedical concepts were incorporated to show representations of concept-related trial information for end-users.

Development of the Duchenne muscular dystrophy Therapeutic Area User Guide

The development of the Duchenne Muscular Dystrophy-Therapeutic Area Unser Guide (DMD-TAUG) incorporated the expert advice of consortium members, which include industry representatives, clinicians, physical therapists, patients, patient advocacy groups, and other disease and drug development experts. D-RSC referenced the common data elements published for DMD by the National Institute for Neurological Diseases and Stroke (NINDS).[12] Common data elements define relevant concepts and their possible values, ranges, and units as applicable, but, unlike CDISC standards, do not offer formal terminology or rules to structure and organize the data. The consortium experts also reviewed the variables collected in the Cooperative International Neuromuscular Research Group Duchenne Natural History Study and in Phase II and III clinical trials of tadalafil, idebenone, drisapersen, eteplirsen, and ataluren. Variables collected in these studies were extracted from case report forms or from publicly available sources, such as clinicaltrials.gov and published literature.

Due to the multisystemic nature of the disease, it was necessary to include elements that described functional changes in the cardiac, respiratory, and skeletal muscle systems, as well as strength measures and some disease-specific biomarkers (dystrophin measurement). D-RSC held weekly teleconferences with experts in each area—neurology, physical therapy, cardiology, and pulmonology—to determine which elements should be included. Standards were developed for core concepts in each of these areas, as shown in **Table 1**. Many different functional measures have been used in trials, and D-RSC took a comprehensive approach to including these. Cardiac measures have not been broadly validated as efficacy end points in DMD so a broad selection of commonly used measurements was included. In contrast, consensus methodologies for muscle magnetic resonance imaging measures are still in development for DMD and were not included at this time.

D-RSC identified 10 major elements that were considered appropriate for inclusion in the TAUG. Of these, five

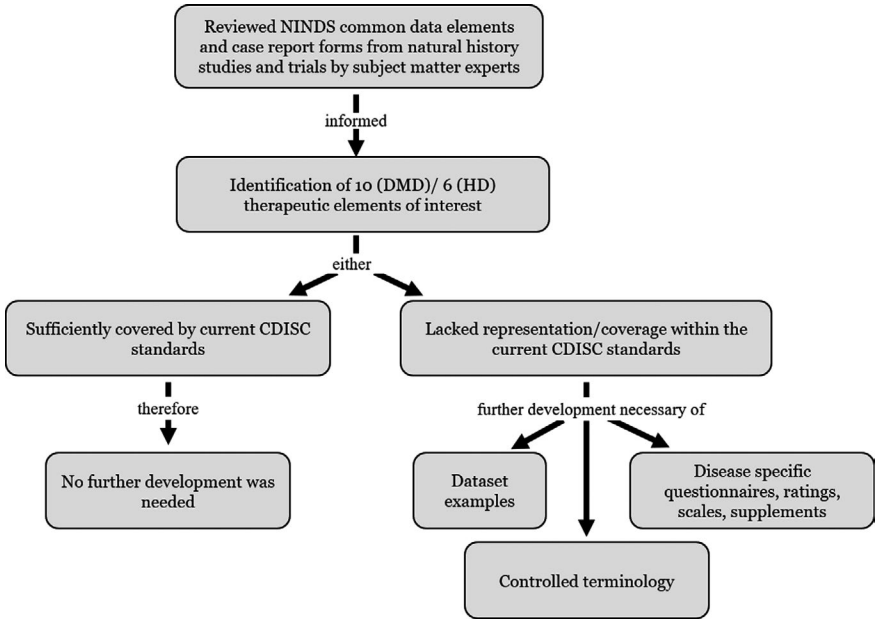

Figure 1 Flow chart showing the development of the Clinical Data Interchange Consortium (CDISC) standards for Duchenne muscular dystrophy (DMD) and Huntington's disease (HD). NINDS, National Institute for Neurological Diseases and Stroke.

Table 1 The concepts included in the Duchenne Muscular Dystrophy-Therapeutic Area User Guide[13]

Concept	Comments
Genetics of DMD	Types of genetic assessments
Assistive devices	Types of assistive device, number of hours per day, number of days per week, age at full-time use
Loss of ambulation	Loss of ambulation date
Cardiac assessments	Left ventricular mass (derived from either a transthoracic echocardiography or cardiac MRI), left ventricular volume, left ventricular internal diameter end diastole and systole, left ventricular fractional shortening, left ventricular ejection fraction, left ventricular end diastolic/systolic diameter, left ventricular posterior wall thickness, septal thickness, cross-sectional thickness end ventricular diastole, tricuspid/mitral/aortic/pulmonic, degree of tricuspid, cardiac valvular regurgitation, degree of mitral regurgitation, cardiac valvular regurgitation, degree of pulmonic regurgitation status, cardiac valvular regurgitation severity, degree of aortic regurgitation, cardiac valvular regurgitation indicator, and left ventricle end diastolic/systolic volume
Imaging of regional and whole-body composition	Dual-energy x-ray absorptiometry scan for bone mineral density, total body fat percentage, and ratio of lean body mass to total body mass
Muscle biopsy	Dystrophin measurement
Musculoskeletal assessments	Grip strength, pinch strength, manual muscle testing, and range of motion
Pulmonary function assessments	FVC, FVC% predicted, PEF, PEF% predicted, FEV1, FEV1% predicted, Maximum Inspiratory Pressure, Maximum Expiratory Pressure, and peak cough flow
Assisted ventilation	Cough assist and assisted ventilation
Rehabilitation therapy	Physical therapy, stretching, gentle exercise, and speech therapy
Questionnaires, ratings, and scales	North Star Ambulatory, Performance of Upper Limb Scale, Brooke Upper Extremity Rating Scale, Vignos Lower Extremity Rating Scale, Griffiths Scale of Mental Development, Bayley III Scales of Infant and Toddler Development, Pediatric Outcome Data Collection Instrument, Egen Klassifikation Scale, PedsQL 3.0 Neuromuscular Module, Quality of Life in Neurological Disorders, and Hammersmith Functional Motor Scale

DMD, Duchenne muscular dystrophy; FEV, forced expiratory volume; FVC, forced vital capacity; MRI, magnetic resonance imaging; PEF, peak expiratory force.

elements had existing terminology housed in the NCI-EVS, hence further development was not required, and the remaining elements all fit into existing SDTM domains. Where elements were complex (i.e., several things may vary within the concept, such as details of methodology that need to be documented), D-RSC developed concept maps that describe how the concepts are linked.

The draft guide then went through a formal review process through CDISC as a Standards Development Organization. First, internal CDISC terminology experts reviewed the TAUG and made updates to the terminology and data structure. Next, the draft user guide was made available for public comment for a 60-day period ending in July of 2017. The document was edited to accommodate the suggested changes and reviewed again by D-RSC and CDISC. The final guide was approved by the CDISC committee (Global Governance Group) in September of 2017.

Development of the Huntington's Disease Therapeutic Area User Guide

The development of the Huntington's Disease-Therapeutic Area User Guide (HD-TAUG) occurred in parallel to the formation of C-Path's HD-RSC. Representatives from the HD community, including subject matter experts, researchers, clinicians, HD gene expansion carriers, nonprofit advocacy groups, and industry members were consulted throughout the development process to outline common data elements, variables, and concepts for inclusion in the HD standards, and to align the standards with clinical and research practice. Measurements were identified and selected for inclusion in the TAUG early in the process with prioritization based on use in drug development. For HD, the measures that were selected for CDISC standards development included genetics, imaging biomarkers, and biofluid biomarkers. Except for magnetic resonance spectroscopy (MRS), discussed below, which used pre-existing data structures but required the creation of two new parameters, the biomarker domains had pre-existing data standards covered in other neurodegenerative therapeutic areas,[7] which served to minimize the workflow required for development. In most cases, even when pre-existing concept maps existed, subject matter experts reviewed the biomarker applications to assure alignment with use in HD clinical research (e.g., phosphodiesterase 10 positron emission tomography, mutant huntingtin in cerebral spinal fluid).

The draft TAUG was made available for a public comment period of 60 days ending in November of 2017, and comments, including those from regulators, were incorporated. In response to received comments that emphasized the increased use of MRS in clinical studies, the decision was made to include MRS as a concept of interest. MRS had originally been excluded due to known variability in measurement between study sites and machine vendors. However, following this input, additional subject matter experts were identified to assist in outlining the variables to capture in this new concept. Following these updates, the new TAUG was made available for an additional 60 days of public comment. The document was edited to address new comments, and the HD-TAUG was published October 1, 2018.

RESULTS

The TAUGs that describe the common data elements and how they should be mapped to SDTM are available online.[13,14] These standards provide consensus vocabulary and guidelines for the organization, structure, and format of

DMD and HD elements that can be used to collect clinical data as well as to map existing data to common terminology. These TAUGs are intended to be implemented for clinical research studies, along with the relevant CDISC Foundational Standards. Furthermore, elements from these (and other CDISC) guides can be used to map elements common to other diseases that use the same or very similar elements.

The concepts included in the DMD-TAUG and HD-TAUG's are shown in **Tables 1** and **2**, respectively. Although this terminology and structure is designed to standardize data reporting, the standard is not intended to dictate or limit the data that can be collected, nor should it dictate the methods that should be used to collect such data or the outcome measures to use in a given clinical program. This terminology and structure should, however, standardize the recording of variables and allow comparison between datasets of similar measurements. CDISC TAUGs are not static entities; additional elements or concepts may be added as new biomarkers or outcome measures are developed and come into common use.

To date, D-RSC has used the DMD-TAUG to map 15 clinical datasets into a common format, which were then integrated and stored in a database that is structured according to CDISC SDTM domains and variables. The individual clinical datasets each contain between 14 and 440 subjects, resulting in an aggregate database of over 4,500 subjects, covering a wide spectrum of the disease in subjects over 4 years old. A subset of the data in the standardized, aggregate DMD database are available to consortium members for analyses and have been accessed by at least seven entities to date. D-RSC members are using a subset of the data (1,100 subjects where relevant end points and covariate information were collected) to develop a series of integrated disease progression models. These models will be used to develop a clinical trial simulation platform that will help optimize trial design by informing inclusion criteria and endpoint selection, and to inform how to optimize clinical trial protocols based on trial duration, numbers of subjects, predicted drug effect, and outcome assessments used.[15]

An HD database with formats based upon the CDISC HD-TAUG is currently under development. It is anticipated to house data from over 10 registries, observational studies, and interventional clinical trials, resulting in an aggregate database of nearly 20,000 participants and covering the continuum of the disease in individuals over 18 years old. In contrast to the DMD database, the datasets to be contained in the HD database range from small observational cohorts of 80–100 subjects, to relatively large interventional clinical trials averaging 400 subjects, to a large observational study and clinical research platform of 15,000 participants. The HD-RSC plans to develop integrated disease progression models of multiple outcome measures that capture clinically meaningful aspects of disease, such as motor, cognitive, and psychological domains. The integrated models will inform quantitative solutions to optimize clinical trials in HD. The disease progression model will incorporate relevant sources of variability, including cytosine-adenine-guanine length, sex, age, and continuous indicators of baseline disease severity, as well as baseline and longitudinal biomarkers. Similar to the DMD database, the goal is to make this database available to external researchers once sufficient data have been integrated.

Table 2 The concepts included in the Huntington's Disease-Therapeutic Area User Guide[14]

Concept	Comments
Cytosine-adenine-guanine repeat length	Focuses on reusability of existing CDISC genetic concepts, including use of Human Genome Variation Society nomenclature
Family history of HD	Covers discussion only as it pertains to genetics of HD
Volumetric MRI	Globus pallidus, striatum, basal ganglia, and caudate nucleus (left and right)
MRS	NAA, total NAA compounds, myo-inositol, glutamate and glutamine, total choline compounds, and total creatine compounds
PET, PET-computed tomography	Fluorodeoxyglucose-PET and phosphodiesterase 10 standard uptake value or standard uptake value ratio, scan parameters, and procedure details (fasting status, radiotracer administration, etc.)
Biofluid sampling/ biomarkers (mutant huntingtin, tau, Aβ, neurofilament)	Procedure details (location of lumbar puncture, time of day, needle gauge, etc.), testing conditions, storage tube composition, and freeze/thaw cycles
Questionnaires, ratings, and scales	Unified Huntington's Disease Rating Scale, Huntington's Disease Cognitive Assessment Battery

CDISC, Clinical Data Interchange Consortium; HD, Huntington's disease; MRI, magnetic resonance imaging; MRS, magnetic resonance spectroscopy; NAA, N-acetyl aspartate; PET, positron emission tomography.

DISCUSSION

The standardization of data elements and data structure provides the ability to compare, contrast, and integrate data from different sources. Structured, standardized data can be used to compare clinical outcomes across different sites, compare different standards of care, and to build integrated databases for the generation of drug development tools. This is particularly valuable in rare diseases, where there are relatively few affected individuals spread across broad geographic regions and data may be collected sporadically and stored in small, independent repositories, in different locations often with different end-goals in mind.

Specific TAUGs are unlikely to be developed for every rare and ultra-rare disease, but the CDISC controlled terminology hosted in the NCI-EVS and the existing domains and general structure of CDISC SDTM are broadly applicable. Thus, the CDISC structure can be used across related diseases because similar, if not identical, data elements may be included in clinical studies for each. For example, C-Path is using guidance from the multiple sclerosis-TAUG to map data for Friedreich's ataxia (FA), a rare disease where no CDISC TAUG has been developed. This is possible because many concepts that are measured in multiple sclerosis are also used in FA, making the development of a new user guide unnecessary. FA-specific elements, such as the Friedreich's Ataxia Rating Scale, have been modeled based on scales of similar nature from other diseases. Similarly,

many of the elements defined in the DMD-TAUG may be common to other muscular dystrophies, so it is likely that only a few unique disease-specific data concepts may be needed in order to map data from these related diseases. Thus, the CDISC data structure can be used widely, with new elements being developed as needed, to structure data even in very rare diseases.

The newly launched Rare Disease Cures Accelerator-Data and Analytics Platform (RDCA-DAP), a partnership between C-Path and the National Organization for Rare Disorders (NORD) through a collaborative grant from the FDA, will utilize this approach across rare diseases. As the platform expands to include data from more rare diseases, new CDISC data concepts and controlled terminology that is not already found through the NCI-EVS can be developed as necessary. Over time, additional data standards will be needed to describe data from new measurements, such as those captured through digital sensors collected from both clinical studies and real-world sources, or newly developed biomarkers or clinical outcome assessments. The CDISC process includes regular updates to the SDTM Foundational Standards, which focus on the core principles for defining data standards and include models, domains, and specifications for data representation, as well as updates to disease-specific TAUGs, where new concepts and data elements will be standardized and included in later versions. This allows for rapid standardization, uptake, and maximum reproducible value of these new measurements.

CDISC standards have been used by C-Path to build integrated databases of clinical data across several diseases,[8,11,16] which have helped develop tools to accelerate drug development in these indications.[17] For example, through C-Path's Critical Path for Alzheimer's Disease (CPAD) consortium, an integrated Alzheimer's disease (AD) database was used to develop a clinical trial simulation platform for trials for mild-to-moderate AD, which has been endorsed by both the FDA and the European Medicines Agency (EMA).[18,19] The continuous expansion of CPAD's AD database to include trials in predementia led to the development of a clinical trial enrichment platform for trials in this stage of the AD continuum, which has received a letter of support from the EMA.[20] These tools have been used both to help explain why previous AD trials did not predict later outcomes and to help design more recent AD trials.[18,19,21–26] Clinical trial design optimization through the implementation of these tools will decrease uncertainty and increase efficiency (reduction of cost, time, and subject burden) in future AD trials. Based on the success of these models, the consortium is now further expanding the database to develop even more comprehensive models across as much of the disease continuum as the data allow. Moreover, these quantitative drug development tools are made possible because the disparate data sources that underlie the analyses were mapped to CDISC standards and thus able to be integrated together.

Whereas developing standards and integrating data in AD is critical to informing clinical trials, such work is perhaps even more essential in rare diseases, where limited numbers of affected individuals make large datasets difficult to collect. Integration of smaller datasets may be essential to put together enough data to create models or other tools that represent the disease in its entirety. For example, the polycystic kidney disease outcomes consortium used CDISC SDTM and the polycystic kidney disease-TAUG to map data from three academic registries and two natural history studies into an integrated database. Subsequently, those data were used to develop a joint biomarker dynamics and disease progression model to demonstrate the relationship between total kidney volume and 30% loss of kidney function.[11,27,28] This has resulted in qualification of total kidney volume as a prognostic biomarker for polycystic kidney disease trials by both the EMA[29] and the FDA,[30] and the marker is now considered a reasonably likely surrogate end point in trials (as defined by the Biomarkers Endpoints and other Tools lexicon[31,32]), enabling an accelerated approval of therapies for this disease. There was one potential therapy in clinical trials prior to this work in 2005. Tolvaptan was approved by the both the FDA and the EMA for the autosomal dominant form of the disease, and lixivaptan has received orphan drug designation; 11 more therapeutic candidates are in the pipeline (https://pkdcure.org/what-is-pkd/latest-research/pipeline/).

The D-RSC database holds data from over 4,500 subjects across disease stages mapped to CDISC SDTM. The curated, integrated database has been used to develop a series of six disease progression models showing the variance in progression in different clinical outcome assessments used in DMD. These models provide the basis for a clinical trial simulation platform that is in development. The models and platform have been accepted into review programs by the FDA and EMA for potential regulatory endorsement.[15] By developing the CDISC TAUG for DMD, a process and structure has been laid out for standardized collection of prospective data across studies, and through which results of different studies can be compared. This will lead to valuable new insights in both treating disease and developing new therapies.

Effective therapies are desperately needed in both DMD and HD, as current therapies at best only slow disease progression or aim to manage symptoms. The therapeutic landscape is ripe with promising new technologies and therapeutic approaches that target the underlying pathobiology. The combination of genetic diagnosis with a robust and standardized clinical database positions the DMD and HD fields to see success with some of the emerging therapies in the clinical pipeline. Further, a better understanding of DMD will likely inform the therapeutic development for related neuromuscular diseases. Advances in HD research may similarly inform drug development in other neurodegenerative diseases, such as Parkinson's disease and AD,[18] and provide learnings and insight for other related rare genetic disorders.

Individuals with rare diseases will benefit from tools that can accelerate the drug development process. Many rare diseases have no approved treatments, and clinical trials are uncommon. However, individuals affected by rare diseases are often very motivated to take part in research and are willing for their data to be shared and used by researchers to accelerate the discovery of new therapies.[22,23] CDISC standards allow data that have been collected to

be standardized and integrated, and for future data to be collected in a standardized format from the start. This will allow the data to be more readily integrated, compared, and contrasted, and for maximization of the utility of every datapoint. Researchers can thus gain new understandings of such diseases and develop smaller, shorter, and more informative clinical trials resulting in a faster progression to new therapies for these diseases. Although rare, these diseases affect many across the globe and represent a substantial unmet need.

Acknowledgments. The authors gratefully acknowledge the efforts of all involved in the development of the TAUGs, listed below. The authors acknowledge critical reviewers of this manuscript, which includes Emily Gantman, Rebecca Kush, Eileen Neacy, and Simon Noble.

DMD-TAUG team: Diane Corey, Laura Butte, Emily Hartley, Richard Abresch, Adrienne Arrieta, Tina Duong, Douglass Chapman, and the support of the Duchenne Regulatory Science Consortium members as a group.

HD-TAUG team: Maryam Abaei, Dana Booth, Laura Butte, Marc Ciosi, Cheryl Fitzer-Attas, Darren Freeman, Rebecca Fuller, Emily Gantman, Emily Hartley, Derek Hill, Beth Hoffman, Torsten Illman, Bernhard Landwehrmeyer, Doug Langbehn, Blair Leavitt, David Lythgoe, Alex MacKay, Darren Monckton, Erin Muhlbradt, Eileen Neacy, John Neville, Gerald Podskalny, Dorian Pustina, David Russell, Bretta Russell-Schulz, Cristina Sampaio, and Julie Stout.

D-RSC Database: Data used in this work was generously provided by members of the DMD community and integrated with funds and experience contributed to the Duchenne Regulatory Science Consortium. This work would not have been possible without those contributions and the willingness of patients with DMD and families to take part in such studies and share the data. D-RSC is funded by member fees, a grant from Parent Project Muscular Dystrophy, and FDA grant U18 FDA005320.

HD-RSC Database: Data used in the preparation of this research were obtained from the C-Path Integrated Huntington's Disease Database. The C-Path Integrated Huntington's Disease Database is funded by the CHDI Foundation, a nonprofit biomedical research organization exclusively dedicated to collaboratively developing therapeutics for Huntington's disease. The C-Path Integrated Huntington's Disease Database would not be possible without the vital contribution of the research participants and their families.

Funding. Duchenne muscular dystrophy: This work was supported by a grant from Parent Project Muscular Dystrophy and member fees from the industry Duchenne Regulatory Science Consortium (D-RSC) members.

Huntington's disease: The CHDI Foundation, a non-profit biomedical research organization exclusively dedicated to collaboratively developing therapeutics that substantially improve the lives of those affected by Huntington's disease, supported this study.

Funding for this publication was made possible, in part, by the US Food and Drug Administration through grant (U18 FD 005320). Views expressed in written materials or publications and by speakers and moderators do not necessarily reflect the official policies of 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.

Conflict of Interest. The authors declared no competing interests for this work.

Author Contributions. A.P.M, D.C., E.C.T., and J.L. wrote the manuscript. A.P.M., R.L., J.B., S.S., L.D.H., K.R., D.T.S., and J.L. designed the research. D.C. and D.O. performed the research. D.C. and D.O. analyzed the data.

1. Meier, T., Perlman, S.L., Rummey, C., Coppard, N.J. & Lynch, D.R. Assessment of neurological efficacy of idebenone in pediatric patients with Friedreich's ataxia: data from a 6-month controlled study followed by a 12-month open-label extension study. *J. Neurol.* **259**, 284–291 (2012).
2. Goemans, N. *et al.* A randomized placebo-controlled phase 3 trial of an antisense oligonucleotide, drisapersen. in Duchenne muscular dystrophy. *Neuromuscul. Disord.* **28**, 4–15 (2018).
3. McDonald, C.M. *et al.* Ataluren in patients with nonsense mutation Duchenne muscular dystrophy (ACT DMD): a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet* **390**, 1489–1498 (2017).
4. Edaravone (MCI-186) ALS 16 Study Group. A post-hoc subgroup analysis of outcomes in the first phase III clinical study of edaravone (MCI-186) in amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler. Frontotemporal Degener.* **18**, 11–19 (2017).
5. Arnerić, S.P., Kern, V.D. & Stephenson, D.T. Regulatory-accepted drug development tools are needed to accelerate innovative CNS disease treatments. *Biochem. Pharmacol.* **151**, 291–306 (2018).
6. Lochmüller, H. *et al.* Position statement: sharing of clinical research data in spinal muscular atrophy to accelerate research and improve outcomes for patients. *J. Neuromuscul. Dis.* **5**, 131–133 (2018).
7. Neville, J. *et al.* Accelerating drug development for Alzheimer's disease through the use of data standards. *Alzheimers Dement. (N Y)* **3**, 273–283 (2017).
8. LaRocca, N.G. *et al.* The MSOAC approach to developing performance outcomes to measure and monitor multiple sclerosis disability. *Mult. Scler.* **24**, 1469–1484 (2018).
9. US Food and Drug Administration (FDA). Electronic Source Data in Clinical Investigations. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/electronic-source-data-clinical-investigations> (2013).
10. Clinical Data Interchange Consortium (CDISC). CDISC Therapeutic Areas. <https://www.cdisc.org/standards/therapeutic-areas>.
11. Perrone, R.D. *et al.* Therapeutic area data standards for autosomal dominant polycystic kidney disease: a report from the Polycystic Kidney Disease Outcomes Consortium (PKDOC). *Am. J. Kidney Dis.* **66**, 583–590 (2015).
12. National Institute for Neurological Diseases and Stroke (NINDS) Common Data Elements. Duchenne muscular dystrophy/Becker muscular dystrophy (data standards). <https://www.commondataelements.ninds.nih.gov/Duchenne%20Muscular%20Dystrophy/Becker%20Muscular%20Dystrophy> (2012).
13. Duchenne Muscular Dystrophy v1.0. CDISC. <https://www.cdisc.org/standards/therapeutic-areas/duchenne-muscular-dystrophy/duchenne-muscular-dystrophy-therapeutic>.
14. Huntington's Disease. CDISC. <https://www.cdisc.org/standards/therapeutic-areas/huntingtons-disease>.
15. Conrado, D.J. *et al.* Towards regulatory endorsement of drug development tools to promote the application of model-informed drug development in Duchenne muscular dystrophy. *J. Pharmacokinet. Pharmacodyn.* **46**, 441–455 (2019).
16. Neville, J. *et al.* Development of a unified clinical trial database for Alzheimer's disease. *Alzheimers Dement.* **11**, 1212–1221 (2015).
17. Conrado, D.J., Karlsson, M.O., Romero, K., Sarr, C. & Wilkins, J.J. Open innovation: towards sharing of data, models and workflows. *Eur. J. Pharm. Sci.* **109S**, S65–S71 (2017).
18. European Medicines Agency (EMA). Qualification opinion of a novel data driven model of disease progression and trial evaluation in mild and moderate Alzheimer's disease. <https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/draft-qualification-opinion-novel-data-driven-model-disease-progression-trial-evaluation-mild_en.pdf> (2013).
19. US Food and Drug Administration (FDA). Quantitative drug development tool for trial simulation in cognitive trials in mild to moderate dementia of the Alzheimer's type. <https://www.fda.gov/media/98856/download> (2013).
20. European Medicines Agency (EMA). Letter of support for Model-based CT enrichment tool for CTs in aMCI. <https://www.ema.europa.eu/en/documents/other/letter-support-model-based-ct-enrichment-tool-cts-amci_en.pdf> (2018).
21. Romero, K. *et al.* The future is now: model-based clinical trial design for Alzheimer's disease. *Clin. Pharmacol. Ther.* **97**, 210–214 (2015).
22. US Food and Drug Administration (FDA). The voice of the patient: Huntington's disease. <https://www.fda.gov/media/96350/download> (2016).
23. US Food and Drug Administration (FDA). The Voice of the Patient: Parkinson's Disease. <https://www.fda.gov/media/124392/download> (2016).
24. Haas, M. *et al.* Big data to smart data in Alzheimer's disease: Real-world examples of advanced modeling and simulation. *Alzheimers Dement.* **12**, 1022–1030 (2016).
25. Rogers, J.A. *et al.* Combining patient-level and summary-level data for Alzheimer's disease modeling and simulation: a β regression meta-analysis. *J. Pharmacokinet. Pharmacodyn.* **39**, 479–498 (2012).

26. Ito, K. *et al.* Understanding placebo responses in Alzheimer's disease clinical trials from the literature meta-data and CAMD database. *J. Alzheimers Dis.* **37**, 173–183 (2013).
27. Perrone, R.D. *et al.* Total kidney volume is a prognostic biomarker of renal function decline and progression to end-stage renal disease in patients with autosomal dominant polycystic kidney disease. *Kidney Int. Rep.* **2**, 442–450 (2017).
28. Perrone, R.D. *et al.* A drug development tool for trial enrichment in patients with autosomal dominant polycystic kidney disease. *Kidney Int. Rep.* **2**, 451–460 (2017).
29. European Medicines Agency (EMA). Qualification opinion: total kidney volume (TKV) as a prognostic biomarker for use in clinical trials evaluating patients with autosomal dominant polycystic kidney disease (ADPKD). <https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/qualification-opinion-total-kidney-volume-tkv-prognostic-biomarker-use-clinical-trials-evaluating_en.pdf> (2015).
30. US Food and Drug Administration (FDA). Qualification of biomarker - total kidney volume in studies for treatment of autosomal dominant polycystic kidney disease. <https://www.fda.gov/media/93105/download>.
31. US Food and Drug Administration (FDA). FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource (Food and Drug Administration (US), Silver Spring, MD, 2016).
32. US Food and Drug Administration (FDA). Table of surrogate endpoints that were the basis of drug approval or licensure. <https://www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure>.