



ORIGINAL RESEARCH

Interoperability of Leigh Syndrome Patient Registry Data with CDISC Standards

Parag Shiralkar*, Pallavi Bakare*, Kasey Woleben[†] and Sophia Zilber[†]

Introduction: Leigh syndrome (LS) is a rare, severe neurometabolic disorder and a type of primary mitochondrial disease. The Cure Mito Foundation is a nonprofit foundation founded in 2018 by parents of affected children. The organization's mission is to unite the global LS community to accelerate patient-centered research, treatments, and cures. Cure Mito has launched a global patient registry for Leigh syndrome in September 2021 to better understand the disease, facilitate clinical trials recruitment, and build a strong global community.

Objectives: Patient data is collected in registries for getting a real-world view of patient reported outcomes, and to improve the understanding of the disease. Establishment of interoperability of registry data with regulatory submission standards like CDISC is essential to accelerate the development of therapies. The project described here has been undertaken to establish such interoperability for Leigh syndrome patient registry data.

Methods: Establishing interoperability consisted of assessment of all data elements and collected data, alignment of data elements and variables with CDASH domains and variables, terminology alignment, data transformation, and compliance assessment.

Results: Data assessment has been done, domain map has been developed, and data has been converted to CDASH standard.

Conclusion: Cure Mito Foundation and Sumptuous Data Sciences, LLC completed a project focused on interoperability of Leigh syndrome patient registry data with CDISC standards, converting the data to CDASH standards, with SDTM conversion in progress. This project is a key milestone for all stakeholders who collect real world data in rare diseases.

Keywords: Leigh syndrome; mitochondrial disease; patient registry; CDISC; CDASH; SDTM; patient advocacy; cure mito; data standards

Introduction

Leigh syndrome (LS) is a rare and severe neurometabolic disorder and a type of primary mitochondrial disease. Symptoms usually start between 3 months and 2 years of age. LS can also occur in teenagers and adults, although this is less common. Newborns with Leigh syndrome seem healthy at birth and may reach major milestones on time. The disease interferes with function of the central nervous system causing children with Leigh syndrome to experience loss of abilities to walk, talk, and swallow. Leigh syndrome affects around 1 in 40,000 individuals and can be caused by over 110 gene mutations in mitochondrial and nuclear DNA. Treatments for Leigh syndrome are focused on managing symptoms and providing supportive care.^{1,2}

Background

The Cure Mito Foundation is an all-volunteer nonprofit foundation that was founded in 2018 by parents of children with Leigh syndrome.³ The organization's mission is to unite the global Leigh syndrome community to accelerate patient-centered research, treatments, and cures. Cure Mito Foundation has built a patient registry for Leigh syndrome in partnership with Coordination of Rare Diseases at Sanford (CoRDS).⁴ CoRDS provides a disease agnostic patient registry platform for rare diseases, with data being available to researchers and industry at no cost and providing a straightforward way to contact patients. Data for over 2000 rare diseases is available on the CoRDS platform. Patient registry is ethically approved by the CoRDS Institutional Review Board and listed on ClinicalTrials.gov (NCT01793168). CoRDS registry platform was established in 2010, and Sanford Health was founded in 1894.

Leigh syndrome patient registry was started in September 2021, and quickly has become the largest patient registry for Leigh syndrome with over 270 participants from at

* Sumptuous Data Sciences, LLC, US

[†] Cure Mito Foundation, US

Corresponding authors: Parag Shiralkar (Parag.Shiralkar@sumptuous-ds.com); Sophia Zilber (sophia@curemito.org)

least 35 countries enrolled as of August 2023. Results have been shared at multiple conferences. This includes the poster presentations at Mitochondria-Related Drug Development Summit 2022 and 2023,⁵ Mitochondrial Medicine Symposium 2022,⁶ mitoNice conference 2022,⁷ annual conference of National Organization of Rare Diseases (NORD) Breakthrough Summit 2022,⁸ Pharmaceutical User Exchange Computational Sciences Symposium (PHUSE -CSS) 2022,⁹ Mitochondrial Medicine – Therapeutic Development conference 2022,¹⁰ and Euromit 2023.¹¹ All results shared to date are available on the Cure Mito Foundation website.¹²

Cure Mito Foundation made it a priority to keep open and transparent communication with the patient community, making sure that all information about the registry, such as how data is shared and used, the purpose of any collected information, as well as the results, are available.¹³

The Clinical Data Interchange Standards Consortium (CDISC) is dedicated to helping to improve clinical research by driving meaningful and efficient research through data standardization.¹⁴ This data standardization enables the rapid design, build, analysis, and submission of clinical trials. CDISC standards are mandatory for submissions to the FDA and PMDA.

Cure Mito Foundation has partnered with Sumptuous Data Sciences, LLC, a specialized data solutions provider to assess interoperability of Leigh syndrome patient registry data with CDISC standards. The goals of the project are to assess the extent to which registry data can be aligned with CDISC standards and what challenges are encountered during such conversion and to share the learning experiences widely in order to benefit other mitochondrial and rare disease real-world data collections. In step one of the project, we looked into alignment of the registry data with the CDISC data acquisition standards (CDASH). Step two is aimed to transform the registry data to CDISC Submission standards as outlined in CDISC Study Data Tabulation Model (SDTM). Establishment of interoperability can help academia, researchers, and industry to understand the data and its use for clinical research and to analyze and report data in commonly accepted data standards and models.

Methods

The registry consisted of two surveys: general and Leigh syndrome specific. The general survey is provided by CoRDS, is the same for all rare diseases on the CoRDS platform and uses Common Data Elements (CDE) provided by the National Institutes of Health (NIH).¹⁵ The Leigh syndrome survey was developed by the Cure Mito Foundation with input from Matthew Klein, MD, CEO of PTC Therapeutics, the industry partner conducting clinical trials in mitochondrial disease and Cure Mito medical and scientific advisory board members.

Information collected included: demographic, mitochondrial disease diagnosis, loss of milestones, disease management, specialists seen, history of symptoms, healthcare utilization and infections within three and 12 months, caregiver burden, and health-related quality of life.

Establishing interoperability consisted of the following steps:

- Step 1: Assessment of all data elements
- Step 2: Review and assessment of collected data
- Step 3: Alignment of data elements with CDASH domains
- Step 4: Variable alignment with CDASH domain variables
- Step 5: Terminology alignment
- Step 6: Data transformation and compliance assessment
- Step 7: Preliminary variable alignment with SDTM domain variables

Although the data transformation process is technology agnostic, because of the availability of the application, our current efforts of the data transformation program development have been in the SAS® system.

Results

Assessment of data collected in Leigh syndrome global patient registry has been done, a domain map has been developed in order to organize the data specific to CDASH requirements, and data has been converted to CDASH standard.

Overall, 339 data fields have been mapped to the CDASH domains of AP, BE, CE, DD, DM, DS, MH, PF and QS as shown in **Table 1**. In the next step of the project, which is already underway, this data will be mapped to the corresponding SDTM domains. Those variables that don't map to the listed domains constitute 'dataset specific variables', which are to be retained to avoid loss of data and will be mapped to supplemental qualifier datasets (SUPPQUAL).

Discussion

Patient registry data provides critical information about patient reported outcomes, phenotypes, quality of life of patients, and the caregiver burden. These data are especially valuable in the case of rare diseases, where patient numbers are small.^{16,17} Registry data also has potential to support therapeutic and drug development processes.

While patient registries have traditionally been developed and maintained by academic researchers and in some cases by industry sponsors, increasingly patient registries are developed and led by patient organizations. This is particularly true in the context of rare diseases, where babies and children are often profoundly impacted, and families face a dearth of information and support. In such circumstances, parents or family members feel compelled to swiftly assume roles as doctors, researchers, fundraisers, and caregivers simultaneously, all in their quest to effectively advocate for and help their loved one.

Patient registries developed by patient organizations have many advantages. They can make it possible to easily share data with researchers or industry, facilitate clinical trial recruitment, understand patient perspectives and disease burden, build a strong global community, and

Table 1: Mapping of Leigh syndrome patient registry data to CDASH domains.

Domain	Domain description	Type of data mapped to the domain	Leigh syndrome patient registry data that mapped to domain	Number of data fields mapped to the CDASH domain	Percent of data fields aligning directly to SDTM main domain in preliminary mapping*
AP	Associated person domain	Data collected about persons other than the subject under study like parents, caregivers, etc.	Information about parent/legally authorized representative Education of mother/father	26	84.6%
BE	Biospecimen Events	Collection and storage information like culture start and end date about the biospecimen.	Information about biospecimen previously donated	8	50%
CE	Clinical events	Clinical events of interest that are not classified as adverse events as per protocol requirements.	Information regarding rare disease symptoms and age when started	40	97.5%
DD	Death details	Additional details collected when a death occurs like cause of death, when the death occurred.	Cause and date of death	4	100%
DM	Demographics	Includes essential fields for a subject like birth date, sex, race.	Demographic variables such participant state, country, race, ethnicity, sex, and other	18	77.7%
DS	Disposition	Accounts all information about a subject into protocol milestones and Disposition Events.	Permissions to be contacted for research	6	100%
MH	Medical history	Medical History of the subject at the start of the trial.	Information regarding patient's diagnosis, age at diagnosis	71	64.7%
PF	Genetics Findings	Pharmacogenomics of interest.	Information regarding genetic mutation	9	100%
QS	Questionnaire	Standardized data collection instruments in form of Questionnaires, Scales.	Health-related quality of life, symptom history, healthcare utilization, loss of milestones, disease management, infections, caregiver burden.	157	62%

*The percentages are provided based on the preliminary mapping process and may slightly change when the final mapping is fully completed.

much more. Achieving these objectives can prove more challenging in other types of registries, especially those established by academic researchers, as there are often significant barriers to sharing data and results, and it is rarely possible to enroll patients from across the globe – a critical aspect for rare diseases.¹⁸

At the same time, there are challenges. FDA guidances that exist for patient registries, natural history studies, or real-world data are written for the industry sponsors, not for patient organizations.^{19,20} When it comes to registries led by patient organizations, there are resources, toolkits, and papers that may help.^{21,22,23,24} Additionally, two of the

authors of this paper have started and are co-leading a PHUSE working group dedicated to creating awareness and education about best practices for collecting regulatory quality data from the industry perspective, improving data quality, empowering patients and patient groups with knowledge and information, and building bridges between industry, academics and patient communities.²⁵ However, there are no formal FDA guidelines outlining how to develop such registries, what information should be collected, who has the ultimate accountability of such registries and how to be sure that the registry is an effective tool that is helping patients.

Since Cure Mito Foundation’s registry data incorporates elements of caregiver burden, phenotypes, as well as health-related quality of life as reported by patients or caregivers of patients suffering from Leigh syndrome, we found that establishment of interoperability of such data with CDISC standards is beneficial for the following reasons:

- The phenotype-specific assessments and findings can help sponsors recruit patients within a very specified target patient population.
- This data can also help supplement assessments of clinical outcomes by providing additional information.
- Interoperability can help provide additional information that can supplement data linkages for the mainstream clinical trial data obtained from other real-world data sources.²⁶
- This data can also be considered in the context of patient reported outcomes as a ‘planned or unplanned clinical event’.

The usability of this registry data with reference to the above benefits can be maximized if this data is interoperable with CDISC standards. In this context, the alignment of Leigh syndrome registry data with the Clinical Data Acquisition Standards Harmonization standards (CDASH) as well as Study Data Tabulation Model standards (SDTM) is important.

The graphic shown in **Figure 1** from “CDISC 360: Evolving our standards toward end to end automation”²⁷ illustrates the flow of data using CDISC standards.

CDASH establishes a standard way to collect data across studies. During a clinical trial, case report forms (CRFs) will usually be structured in such a way that data will be captured following CDASH guidelines and be aligned with industry standards from the beginning. This is usually not done in a patient registry, natural history study, or other types of research studies. However, according to CDASH implementation guide, version 2.2, “although the CDASH standard was originally developed for use in regulatory submissions, it can be used by any organization or individual involved in the collection, preparation, and analysis of clinical research data that may also be used for other purposes, including publication, warehousing, and meta-analyses”²⁸.

After collecting data in CDASH standard, data is typically converted to SDTM format. SDTM provides a standard for organizing and formatting data to streamline processes in collection, management, analysis, and reporting.²⁹ During a clinical trial or study, SDTM data is usually transformed into ADaM (Analysis Data Model) format to meet specific scientific and medical objectives; however, for our project, immediate conversion is not required, though we will have the capability to convert SDTM datasets to ADaM if deemed necessary in the future.

Leigh syndrome patient registry data has been successfully mapped to CDASH domains, as shown in **Table 1**. For all domains, either all or the majority of data will map directly to the main SDTM domain²⁹ with the rest of the data to be mapped to Supplemental Qualifier (SUPPQUAL) datasets, which allow us to add non-standard data to our SDTM.³⁰ The process of SDTM mapping is already well underway and close to completion.

Our data mapping result is a significant milestone for this registry and holds the potential to serve as a crucial stepping stone for advancing data collection efforts within the mitochondrial disease and rare disease community. The attainment of data interoperability provides a significant advantage for subsequent stages of data curation and validation processes, ensuring its effective utilization as a supplementary resource in mainstream randomized clinical trials.

It is important to understand that the extent to which a specific survey would map to CDISC is a result of the design of the initial surveys. In other words, not every patient reported survey will map to CDISC this well. In our case, we attribute the success of this project to several contributing factors. Firstly, the utilization of NIH common data elements in the general survey facilitated a seamless conversion to CDISC standards. Secondly, the thorough design of the Leigh syndrome survey, where questions were separated into sections that resembled CDISC domains, also played a crucial role in its straightforward conversion to CDISC. For this reason, the establishment of formal guidelines for designing surveys for patient foundations is especially important. Third, Cure Mito Foundation has been analyzing data from the start of data collection as well as partnering and collaborating with data experts throughout, ensuring high quality for both the data and surveys. The authors are not aware of other projects where patient reported data collected by a patient advocacy

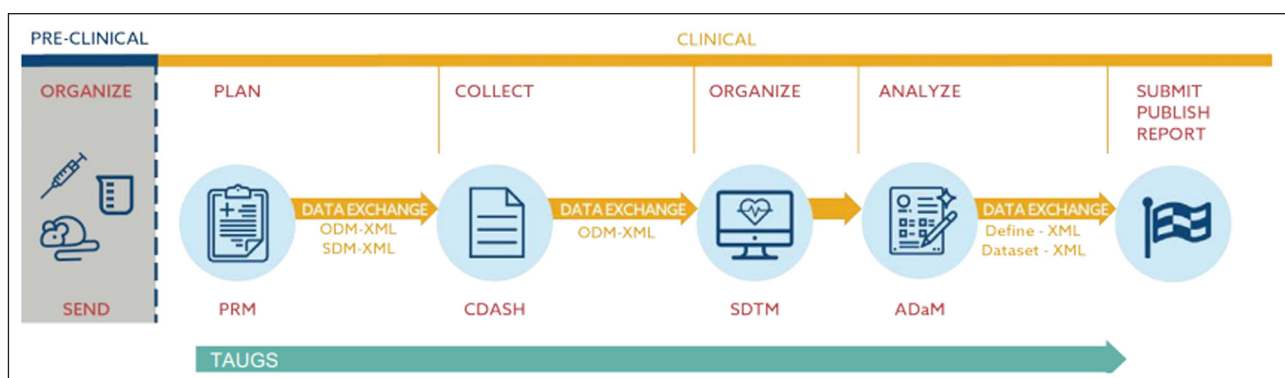


Figure 1: CDISC Standards in the Clinical Research Process.

organization has been converted to CDISC, neither in mitochondrial disease or in rare disease overall. Therefore, this is likely the first such project of its kind.

As a next step, Cure Mito Foundation, and Sumptuous Data Sciences, LLC is continuing discussions and assessments of usability of the registry data in the context of data linkages, exploratory endpoint assessments and identification of measurable and valid phenotypes in case of Leigh syndrome. They also continue to collaborate and get guidance and recommendations from regulatory bodies. Additionally, the Cure Mito Foundation will continue to work with their industry partners to better understand next steps regarding use of this data.

Conclusion

The Cure Mito Foundation and Sumptuous Data Sciences, LLC successfully completed a project focused on interoperability of Leigh syndrome patient registry data with regulatory submission standards. This project is likely the first such attempt, and therefore is an important learning experience as well as a key milestone for patient advocacy organizations and researchers who collect patient registry and natural history data in rare diseases. Next steps include further discussions with regulatory bodies and industry partners for development of mutually agreeable process for data curation and submission to regulatory bodies.

Competing Interests

The authors have no competing interests to declare.

References

1. **Rahman S.** Leigh syndrome. *Handb Clin Neurol*. 2023; 194: 43–63. PMID: 36813320. DOI: <https://doi.org/10.1016/B978-0-12-821751-1.00015-4>
2. **About Leigh Syndrome.** Available at <https://aboutleighsyndrome.com/>
3. **Cure Mito Foundation.** Available at <https://www.curemito.org/>
4. **Coordination of Rare Diseases at Sanford (CoRDS).** Available at <https://research.sanfordhealth.org/rare-disease-registry>
5. **Mitochondria-Related Drug Development Summit.** Available at <https://mitochondria-drug-development.com/>
6. **Mitochondrial Medicine Symposium.** Available at https://www.umdf.org/symposium_2022/
7. **MitoNICE.** Available at <https://www.myology2022.org/>
8. **NORD Breakthrough Summit.** Available at <https://nordsummit.org/>
9. **PHUSE/FDA Computational Science Symposium (CSS) 2022.** Available at <https://www.cdisc.org/events/education/external-events/2022/09/phusefda-computational-science-symposium-css-2022>
10. **Mitochondrial Medicine – Therapeutic Development.** Available at <https://coursesandconferences.wellcomeconnectingscience.org/event/mitochondrial-medicine-therapeutic-development-20221121/>
11. **Euromit 2023.** Available at <https://euromit2023.eu/>
12. **Leigh syndrome patient registry results.** Available at <https://www.curemito.org/results>
13. **Meskó B, deBronkart D.** Patient Design: The Importance of Including Patients in Designing Health Care. *J Med Internet Res*. 2022; 24(8): e39178. DOI: <https://doi.org/10.2196/39178>
14. **Clinical Data Acquisition Standards Harmonization (CDASH).** <https://www.cdisc.org/standards/foundational/cdash>
15. **NIH Common Data Element repository.** <https://cde.nlm.nih.gov/home>
16. **Gliklich RE, Dreyer NA, Leavy MB, editors.** Registries for Evaluating Patient Outcomes: A User's Guide [Internet]. 3rd edition. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014 Apr. 1, Patient Registries. Available at <https://www.ncbi.nlm.nih.gov/books/NBK208643/>.
17. **Chinnery PF, Falk MJ, Mootha VK, Rahman S.** Editorial: Mitochondrial medicine special issue. *J Inherit Metab Dis*. 2021; 44(2): 289–291. DOI: <https://doi.org/10.1002/jimd.12374>
18. **Denton N., Molloy M, Charleston S, et al.** Data silos are undermining drug development and failing rare disease patients. *Orphanet J Rare Dis*. 2021; 16(161). DOI: <https://doi.org/10.1186/s13023-021-01806-4>
19. **U.S. Department of Health and Human Services Food and Drug Administration.** Rare Diseases: Natural History Studies for Drug Development Guidance for Industry. Draft Guidance. March 2019. Available at <https://www.fda.gov/media/122425/download>
20. **U.S. Department of Health and Human Services Food and Drug Administration.** Patient-Focused Drug Development: Incorporating Clinical Outcome Assessments Into Endpoints For Regulatory Decision-Making. Draft Guidance. April 2023. Available at <https://www.fda.gov/media/166830/download>
21. **Jones C.** Rare Disease Registries: Advancing Disease Understanding, Treatments, and Cures. Available at https://globalgenes.org/wp-content/uploads/2019/01/GG_toolkit_rare-registries-web-hyperlinked.pdf
22. **Registries for Evaluating Patient Outcomes: A User's Guide: 4th Edition.** 2020. Available at <https://effectivehealthcare.ahrq.gov/products/registries-guide-4th-edition/users-guide>
23. **U.S. Department of Health and Human Services, National Institutes of Health,** National Center for Advancing Translational Sciences, RADAR Rare Diseases Registry Program. Available at <https://registries.ncats.nih.gov/module/setting-up-your-registry/decide-how-to-collect-and-store-your-data/overview/>
24. **Wicks P, Wahlstrom-Edwards L, Fillingham S, et al.** So You Want to Build Your Disease's First Online Patient Registry: An Educational Guide for Patient Organizations Based on US and European Experience. *Patient* 2023; 16: 183–199. DOI: <https://doi.org/10.1007/s40271-023-00619-w>

25. **Newton N.** Best Data Practices for Rare Disease Patient Foundations and Researchers. 2023. Available at <https://advance.phuse.global/display/WEL/Best+Data+Practices+for+Rare+Disease+Patient+Foundations+and+Researchers>
26. **U.S. Department of Health and Human Services Food and Drug Administration.** Real-World Data: Assessing Registries to Support Regulatory Decision-Making for Drug and Biological Products Guidance for Industry. Draft guidance for industry. November, 2021. Available at <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-registries-support-regulatory-decision-making-drug-and-biological-products>
27. **Van Reusel P, Hume S, Cohen B.** CDISC 360: Evolving our standards towards end to end automation. 2020. Available at https://www.cdisc.org/sites/default/files/2020-05/CDISC_360_Project_Description.pdf
28. **CDASH Implementation Guide v2.2.** Available at <https://www.cdisc.org/standards/foundational/cdash/cdashig-v2-2/>
29. **CDISC Submission Data Standards Team.** Study Data Tabulation Model Implementation Guide: Human Clinical Trials. v3.3 (Final). 2018. Available at <https://www.cdisc.org/standards/foundational/sdtmig/sdtmig-v3-3/html>
30. **Supplemental Qualifiers.** CDISC Online course, available at <https://www.cdisc.org/education/course/supplemental-qualifiers>

How to cite this article: Shiralkar P, Bakare P, Woleben K, Zilber S. Interoperability of Leigh Syndrome Patient Registry Data with CDISC Standards. *Journal of the Society for Clinical Data Management*. 2023; 4(1): 1, pp. 1–6. DOI: <https://doi.org/10.47912/jscdm.244>

Submitted: 16 February 2023

Accepted: 11 July 2023

Published: 22 December 2023

Copyright: © 2023 SCDM publishes JSCDM content in an open access manner under a Attribution-Non-Commercial-ShareAlike (CC BY-NC-SA) license. This license lets others remix, adapt, and build upon the work non-commercially, as long as they credit SCDM and the author and license their new creations under the identical terms. See <https://creativecommons.org/licenses/by-nc-sa/4.0/>.



Journal of the Society for Clinical Data Management is a peer-reviewed open access journal published by Society for Clinical Data Management.

OPEN ACCESS The Open Access icon, which is a stylized 'O' with a circular arrow inside, indicating that the content is freely available.