The LORIS MyeliNeuroGene Rare Disease Database for Natural History Studies and Clinical Trial Readiness

Aaron Spahr
McGill University Health Centre
Zaliqa Rosli
McGill University Health Centre
Mélanie Legault
McGill University Health Centre
Luan Tran
McGill University Health Centre
Simon Fournier
McGill University Health Centre
Helia Toutounchi
McGill University Health Centre
Lama Darbelli
McGill University Health Centre
Cecile Madjar
McGill University Health Centre
Cassandra Lucia
McGill University Health Centre
Marie-Lou St-Jean
McGill University Health Centre
Samir Das
McGill University Health Centre
Alan Charles Evans
McGill University Health Centre
Geneviève Bernard (genevieve.bernard@mcgill.ca)
McGill University Health Centre https://orcid.org/0000-0002-9634-2966

Research Article

Keywords: leukodystrophy, rare diseases, information management systems, databases, registry, natural history, outcome measures, clinical trials, biomarkers

Posted Date: May 6th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-491571/v1

Version of Record: A version of this preprint was published at Orphanet Journal of Rare Diseases on July 23rd, 2021. See the published version at https://doi.org/10.1186/s13023-021-01953-8.

The LORIS MyeliNeuroGene Rare Disease Database for Natural History Studies and Clinical Trial Readiness

Aaron Spahr[1,2,3,4,5], **Zaliqa Rosli**[6], **Mélanie Legault**[6], **Luan T Tran**[1,2,3,4,5], **Simon Fournier**[1,2,3,4,5], **Helia Toutounchi**[1,2,3,4,5], **Lama Darbelli**[1,2,3,4,5], **Cécile Madjar**, **Cassandra Lucia**[1,2,3,4,5], **Marie-Lou St-Jean**[1,2,3,4,5], **Samir Das**[6], **Alan C Evans**[6], **Geneviève Bernard**[1,2,3,4,5,+]

1. Department of Neurology and Neurosurgery, McGill University, Montréal, Québec Canada
2. Department of Pediatrics, McGill University, Montréal, Québec, Canada
3. Department of Human Genetics, McGill University, Montréal, Québec, Canada
4. Department of Specialized Medicine, Division of Medical Genetics, McGill University Health Centre, Montréal, Québec, Canada
5. Child Health and Human Development Program, Research Institute of the McGill University Health Center, Montréal, Québec, Canada
6. McGill Centre for Integrative Neuroscience, Montreal Neurological Institute, McGill University, Montréal, Québec, Canada

+ Corresponding Author: genevieve.bernard@mcgill.ca

Abstract

Background: Rare diseases are estimated to affect 150-350 million people worldwide. With advances in next generation sequencing, the number of known disease-causing genes has increased significantly, opening the door for therapy development. Rare disease research has therefore pivoted from gene discovery to the exploration of potential therapies. With impending clinical trials on the horizon, researchers are in urgent need of natural history studies to help them identify

surrogate markers, validate outcome measures, define historical control patients, and design therapeutic trials. **Results**: We customized a browser-accessible multi-modal (e.g. genetics, imaging, behavioral, patient-determined outcomes) database to increase cohort sizes, identify surrogate markers, and foster international collaborations. Ninety data entry forms were developed including family, perinatal, developmental history, clinical examinations and diagnostic investigations, neurological evaluations (i.e. spasticity, dystonia, ataxia, etc.), disability measures, parental stress, and quality of life. A customizable clinical letter generator was created to assist in continuity of patient care. **Conclusions**: Small cohorts and underpowered studies are a major challenge for rare disease research. This online, rare disease database will be accessible from all over the world, making it easier to share and disseminate data. We have outlined the methodology to become Title 21 Code of Federal Regulations Part 11 Compliant, which is a requirement to use electronic records as historical controls in clinical trials in the United States. Food and Drug Administration compliant databases will be life-changing for patients and families when historical control data is used for emerging clinical trials. Future work will leverage these tools to delineate the natural history of several rare diseases and we are confident that this database will be used on a larger scale to improve care for patients affected with rare diseases.

Keywords: leukodystrophy, rare diseases, information management systems, databases, registry, natural history, outcome measures, clinical trials, biomarkers

Background

According to the World Health Organization (WHO), the definition of a rare disease is one that affects every 1 in 2,000 people or less. Global prevalence of these approximately 8,000 rare genetic disorders is estimated to affect between 150 and 350 million people [1-7]. Historically, rare diseases have been notoriously difficult to diagnose due to their heterogeneous phenotypes

and genotypes [8]. Since only around 5% of all rare diseases have an FDA-approved treatment, many orphan diseases utilize off-label indications of medications approved for other purposes [9]. However, an incredible amount of advancement has been accomplished over the last decade using rapidly evolving genetic technologies, including with the most recent use of next generation sequencing (NGS), to identify genes causing these diseases [6]. The description of novel rare disease entities and the identification of novel disease-causing genes have opened the door for studies investigating disease pathogenesis and potential therapeutic approaches [6]. Rare disease research has therefore pivoted from gene discovery towards investigating potential treatments [6].

With impending clinical trials on the horizon, rare disease researchers are realizing a tremendous need for natural history data [10, 11]. The goal of a natural history study is to recruit patients for longitudinal analysis of natural disease progression [12]. The data gathered is used to help identify surrogate markers, determine the best outcome measures to be used in potential therapeutic trials, and can serve as the control arm of a clinical trial [13-16]. Natural history studies result in incredible amounts of information being collected, including clinical, behavioral, sociodemographic, genetic, imaging, and patient and family reported outcomes.

This diversity and quantity of data can be difficult to manage, so rare disease researchers must begin to utilize information management systems, or databases, to facilitate natural history studies. Rare disease research relies heavily on international collaboration and data sharing in order to recruit large patient populations to obtain adequate statistical power [6, 17]. Therefore, utilizing an online database can uniquely benefit rare disease research more than other disease research fields where significant patient populations are more prevalent.

If rare disease databases are going to be successful in future clinical trials, they must adhere to local and international regulations for electronic records. Title 21 Code of Federal Regulations

(CFR) Part 11 published in 1997, from the U.S. Food and Drug Administration, outlines what is considered trustworthy, reliable record keeping. These regulations apply to any FDA-regulated industry, such as pharmaceutical companies, medical device manufacturers, biotechnological companies, and clinical research organizations. We chose to adhere to all general requirements that will be detailed below in the Methods section.

There are a variety of different databases available to aid researchers such as RedCap [18], Deduce [19], HID [20], DFBIdb [21], LONI [22], MIND [23], NeuroLOG [24], etc. We elected to customize the Longitudinal Online Research and Imaging System (LORIS) [25-28] to help organize data and facilitate international collaborations when conducting multi-site natural history studies because of its strong track record and the fact that it is open source. Here, we detail below how our group used LORIS and 21 CFR Part 11 guidelines to set up workflows and developed the LORIS MyeliNeuroGene Database for Rare Diseases to lead us to clinical trial preparedness in the coming years.

Results

An iteration of LORIS was installed and configured for the MyeliNeuroGene Research Group at the Research Institute of the McGill University Health Centre. This database is easily accessible via a web browser and multi-modal, with the ability to capture genetics data, medical history, medical imaging, detailed assessments of cognition and motor function, and patient-derived outcomes, among other things.

Within LORIS, data entry forms, or instruments, were created using the “Instrument Builder” module. Using the workflow found in the Methods section, 90 LORIS instruments were

created. Detailed phenotyping including family history, perinatal history, developmental history, clinical evolution, time to event, neurological examination, neuropsychological assessment, etc. were developed in conjunction with other parent- and patient-reported outcomes such as quality of life, disability, and stress. Of the 90 instruments, 62 had scoring algorithms developed to aid in data processing. The resulting instruments are summarized in Table 1.

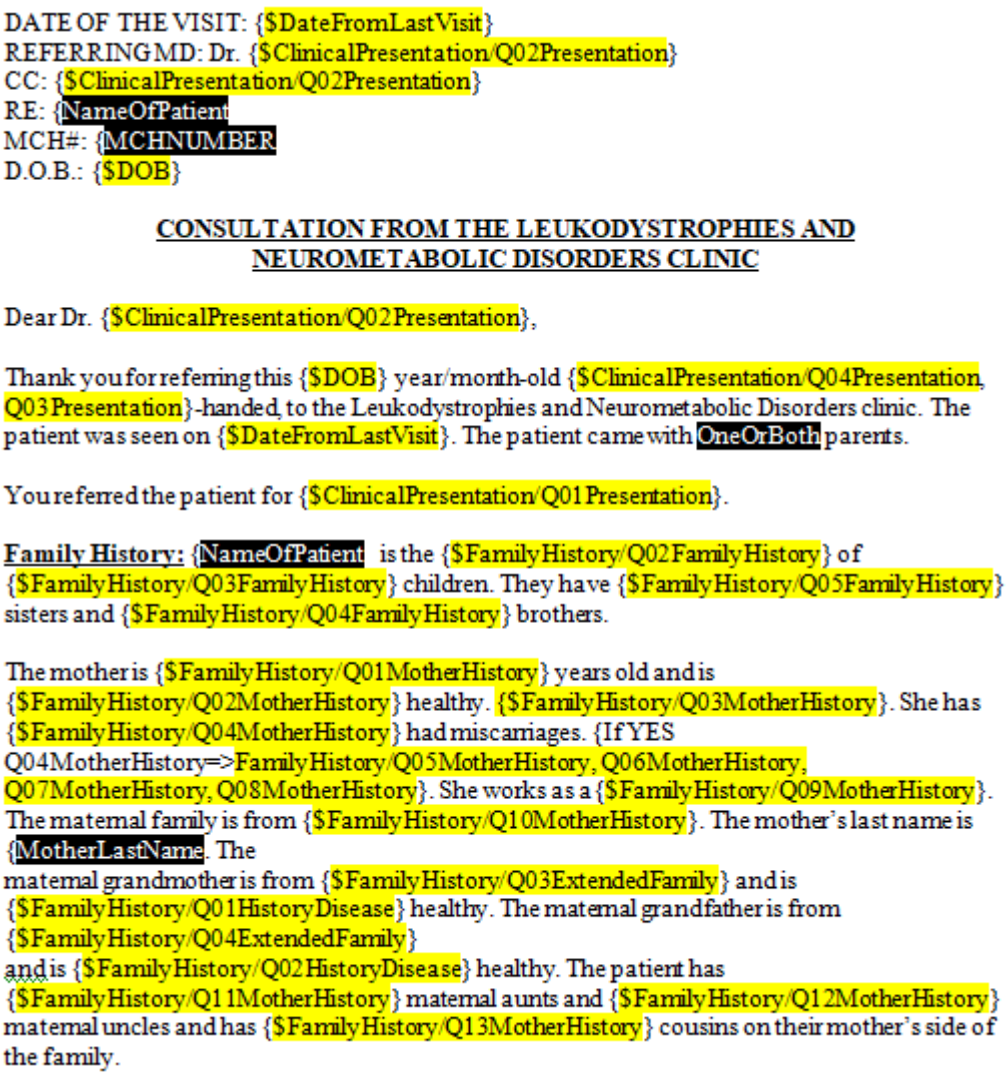
Table 1: Developed Instruments of the MyeliNeuroGene Loris Database

Instrument	**Purpose**
Family History	Inheritance pattern
Perinatal History	Disease Onset/Progression
Developmental History	Disease Onset/Progression
Investigations	Diagnostic Odyssey
Demographics	Sociodemographic variables
Clinical Presentation	Disease Onset/Progression
Primary Diagnosis	Disease Onset/Progression
Gross Motor Function Measure - 88	Measure changes in motor function
Leiter-3 Intelligence Scale	Measure changes in intelligence
Neuropsych Examinations	Measure changes in cognition
Rehabilitation	PT, OT, SLT, etc. used
Clinical Evolution	Disease Onset/Progression
Time to Event	Disease Milestones
Clinical Examination	Disease Onset/Progression
Swallowing Scales	VFSS and FEES evaluation
MRI Analyses	Disease Onset/Progression
Modified Ashworth Scale (MAS)	Measure changes in spasticity
Fahn Marsden Scale (F-M)	Measure changes in dystonia
Global Dystonia Scale (GDS)	Measure Changes in dystonia
Guy's Neurological Disability Scale (GNDS)	Measure disability and ADL
Gross Motor Function Classification System (GMFCS)	Characterize gross motor function
Communication Function Classification System (CFCS)	Characterize communication function
Manual Ability Classification System (MACS)	Characterize fine motor function
Eating and Drinking Ability Classification System (EDACS)	Characterize eating function
Scale for the Assessment and Rating of Ataxia	Measure changes in ataxia
Non-communicating Children's Pain Checklist - Revised	Measure parent reported pain
Parent Reported Stress Questionnaires	Measure parental stress
Health-Related Quality of Life Questionnaires	Measure patient’s quality of life

Abbreviations: PT: physical therapy; OT: occupational therapy; SLT: Speech and language therapy; VFSS: Video fluoroscopic swallow study; FEES: Fiberoptic endoscopic evaluation of swallowing; MRI: Magnetic resonance imaging; ADL: Activities of daily living

One thousand patients and family members with rare diseases have been included into LORIS and assigned unique identifiers. This includes activation of enrollment, informed consent designation, external identifier logging, and family relationship mapping.

In addition, a dynamic letter generator is currently in development to assist in forwarding patient information to other physicians. The tool compiles the patient's data, entered via the phenotyping instruments, into a Clinical Examination Letter. In place of the database field names, highlighted in yellow in Figure 1, an instance of the letter renders the patient data for the corresponding field. The Clinical Examination Letter can be exported as an editable word document that details patient information, such as family history, clinical evolution, time to event and future plans for investigations. This letter can then be sent to the referring physicians for continuity of care, and has the advantage of not duplicating work done by the data entry clinician; as the clinician sees the patient and enters the data in the LORIS MyeliNeuroGene Database, the clinical note is auto-populated.



DATE OF THE VISIT: {$DateFromLastVisit}
REFERRING MD: Dr. {$ClinicalPresentation/Q02Presentation}
CC: {$ClinicalPresentation/Q02Presentation}
RE: {NameOfPatient
MCH#: {MCHNUMBER
D.O.B.: {$DOB}

CONSULTATION FROM THE LEUKODYSTROPHIES AND NEUROMETABOLIC DISORDERS CLINIC

Dear Dr. {$ClinicalPresentation/Q02Presentation},

Thank you for referring this {$DOB} year/month-old {$ClinicalPresentation/Q04Presentation, Q03Presentation}-handed, to the Leukodystrophies and Neurometabolic Disorders clinic. The patient was seen on {$DateFromLastVisit}. The patient came with OneOrBoth parents.

You referred the patient for {$ClinicalPresentation/Q01Presentation}.

Family History: {NameOfPatient is the {$FamilyHistory/Q02FamilyHistory} of {$FamilyHistory/Q03FamilyHistory} children. They have {$FamilyHistory/Q05FamilyHistory} sisters and {$FamilyHistory/Q04FamilyHistory} brothers.

The mother is {$FamilyHistory/Q01MotherHistory} years old and is {$FamilyHistory/Q02MotherHistory} healthy. {$FamilyHistory/Q03MotherHistory}. She has {$FamilyHistory/Q04MotherHistory} had miscarriages. {If YES Q04MotherHistory=>FamilyHistory/Q05MotherHistory, Q06MotherHistory, Q07MotherHistory, Q08MotherHistory}. She works as a {$FamilyHistory/Q09MotherHistory}. The maternal family is from {$FamilyHistory/Q10MotherHistory}. The mother's last name is {MotherLastName. The maternal grandmother is from {$FamilyHistory/Q03ExtendedFamily} and is {$FamilyHistory/Q01HistoryDisease} healthy. The maternal grandfather is from {$FamilyHistory/Q04ExtendedFamily} and is {$FamilyHistory/Q02HistoryDisease} healthy. The patient has {$FamilyHistory/Q11MotherHistory} maternal aunts and {$FamilyHistory/Q12MotherHistory} maternal uncles and has {$FamilyHistory/Q13MotherHistory} cousins on their mother's side of the family.

Figure 1: Screenshot of the LORIS MyeliNeuroGene dynamic letter generator: Yellow highlights customizable variables for the clinical letter generator. Black highlighted variables represent information that is not stored in LORIS and must be filled in by the physician.

Discussion

Most patients affected with rare diseases, from mildly to severely affected, support data sharing to promote research, healthcare, and knowledge transfer [17]. We have built and customized a LORIS database and detailed our workflow to aid rare disease researchers to create their own information management system, electronic health records, or database. There is a major need and benefit to sharing data in rare disease research. De-identifying and sharing information

allows rare disease researchers to efficiently study disorders by collaborating and minimizing redundant studies [29], and by maximizing sample sizes.

An exportable dynamic letter generator has also been developed to save time when examining patients referred to the clinic. Patients with a rare disease who come to the Montreal Children's Hospital undergo a battery of tests that can take up to two days to complete. All information is stored in the LORIS MyeliNeuroGene Database and can be exported in the form of a Clinical Examination Letter detailing all results, impressions, and plans to help treat the patients. This letter is then sent back to the referring physician for continuity of care. When this letter is written by hand it takes a few hours and introduces numerous chances for human error. Exporting the letter from quality-controlled instruments reduces this error and saves researchers and physicians' time.

In addition to the clinical phenotyping instruments and dynamic letter generator, we have outlined, for the first time, the methodology to become Title 21 Code of Federal Regulations Part 11 Compliant, which is a requirement to use electronic records as historical controls in clinical trials in the United States [30, 31]. To our knowledge, our manuscript is the first to outline the requirements to adhere to 21 Code of Federal Regulations Part 11 Compliance. Future work will leverage the tools developed in this project to delineate the natural history of several rare diseases and will hopefully be used by clinicians and researchers around the globe.

Conclusions

A major obstacle in rare disease research is overcoming small cohorts. Developing an online database that international collaborators can access and contribute to from all over the world is invaluable for increasing cohort sizes, discerning surrogate markers, and improving natural

history data. Using this FDA compliant natural history data to validate outcome measures will be life-changing for patients and families because it will lead to historical control data that can be used in emerging clinical trials.

Methods

Title 21 Code of Federal Regulations Part 11 Compliance (Part 11 Compliance)[32]

To adhere to Part 11 Compliance regulations, the LORIS MyeliNeuroGene Database has been customized to include additional security measures such as time stamped audit trails. We are currently implementing the electronic signatures and the 2-factor authentication. There is a gap in scientific literature detailing workflow and database development. As such, we will summarize the general requirements of Part 11 Compliance below and how they were implemented into our database.

Training Verification: Users are required to have their credentials (e.g. education, training, experience) verified before performing tasks within the database. Written policy must be signed holding users accountable and responsible for their electronic signatures (discussed further below). This written policy must be stored, and a hard copy sent to the Office of Regional Operations (HFC-100), 5600 Fishers Lane, Rockville, MD 20857.

Biometrics: This is a method of verifying an individual's identity based on a measurement of the individual's physical features (i.e. fingerprints, etc.) or repeatable action that are unique to that person. In our case, we chose to use a unique pin separate from an authorized user's password for 2-factor authentication.

Closed system: The **MyeliNeuroGene** database is a closed environment, meaning that access to the system is controlled by the same people who are responsible for the content of the electronic records. This includes the researchers and principal investigator. Operational audits on the system are done on a routine basis. Time stamp audit trails are tracked for each authorized user to trace creation, modification, or deletion of any instrument, visit, or other electronic record. User access is hierarchical, meaning some users do not have full access to the database and may only have "read" or "write" access only. The database also must ensure that no user has the same pin or password, and that pins and passwords are periodically checked and changed to prevent unauthorized use. If unauthorized use occurs, there are immediate system security notifications. Per Canadian predicate rules, records must be stored for 25 years after study completion. United States record retention rules require storage for a minimum of 10 years.

Quality Control: Processing pipelines must ensure data fits specific parameters and types. This is discussed in depth under the Methods section "LORIS Database and Workflow".

Electronic signature: This includes any combination of text, graphic, data, audio, or other information that is represented in digital form by the database. Electronic signatures must include printed names of the signers, dates and times, meanings (e.g. approval, creation, reviewing), and an internal audit trail. These signatures are legally binding. Authority checks are completed every month to ensure only authorized users may sign, input, output, or modify records.

Digital signature: A digital signature combines the electronic signature and its corresponding cryptographic authentication, usually a pin and/or password that is used to verify the identity of the signer. It cannot be copied or pasted to or from another document, making it inexorably linked to the signed document. To not become cumbersome, continuous signing periods only require the first to be 2 factors authenticated with a biometric identification and password.

193 *External Auditing*: It is highly recommended that after database development a third-party
194 auditor inspects the system and documentation put in place. Auditors alert parties of any gaps or
195 shortcomings and can advise developers of what needs to be changed for full compliance with
196 local and international regulations. This will be organized for the MyeliNeuroGene database.

197 **LORIS Database and Workflow**

198 *Architecture:* LORIS is a web-based data and project management software that stores
199 demographic, clinical, behavioral, genetic, imaging, and patient-related outcomes accessible from
200 any computer browser connected to the internet [25]. Multiple sites can enter, organize, and
201 validate data under one management framework. Longitudinal data is organized around the
202 "Subject Profile". Clinical examination, imaging data, outcome measures, and metadata are
203 organized by "Visits". All stored information is de-identified and can be queried by an authorized
204 user. Source documentation can be uploaded and affiliated with each visit. Quality control is
205 ensured by automated scoring of clinical, behavioral and patient-reported outcomes, validating
206 data types (string vs numerical), and requiring double data entry where necessary.

207 *Workflow*: To properly set up our rare disease database, we first began by drafting a data
208 dictionary in the form of an Excel sheet. This spreadsheet outlined all of the data entry forms, or
209 instruments, that would be developed using the LORIS Instrument Builder module detailed below.
210 After instrument creation, participant enrollment and data entry can begin, with query and
211 dissemination details tackled later. An overview of the workflow can be found in Figure 2.

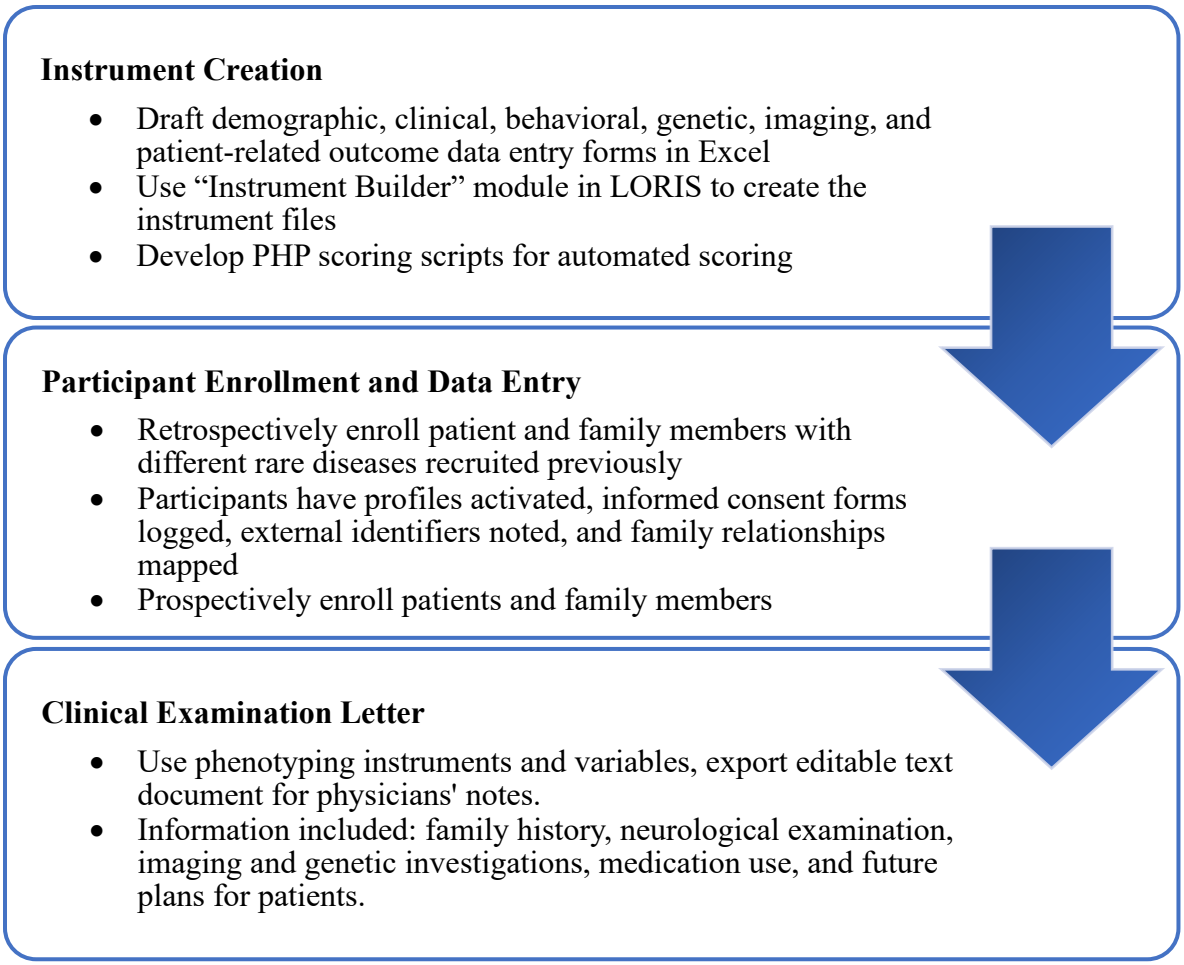


Figure 2: Database development workflow to create instruments, scoring algorithms, enroll patients, enter data, and output information into a clinical examination letter

Instrument Builder: Within LORIS are different modules to help researchers with no computer science or programming experience. The Instrument Builder module aids in the creation of demographic, clinical phenotyping, behavioral, genetic, imaging, and patient-related outcome measures. Each instrument can be customized with specific information such as a "Header", "Label", and "Scored Field" that give the instrument title, background information, and automatically calculated scoring respectively.

Data entry can be standardized using a "Textbox", "Text area", "Dropdown", "Multiselect", "Date", and "Numeric" question entry. Each question is assigned a variable name "Question Name", for calculations and data querying, and "Question Text" which asks the

pertinent question at hand. For Dropdown questions, instrument specific options can be added for every question.

Instrument Creation: Instruments were first planned and drafted using Excel in the form of a Data Dictionary. Columns consisted of Question Names, type of question (e.g. Numeric, Dropdown, etc.), Question Text, Question Options (available choices), and Formulas (for later calculations). Each row represented one question. Using the Data Dictionary and the Instrument Builder module on LORIS, each instrument was created: demographic forms, clinical phenotyping (i.e. spasticity and dystonia measures, gross and fine motor, eating, and drinking function, ataxia, intelligence, disability, swallowing evaluations etc.), behavioral , genetic, imaging (i.e. MRI analyses), and patient-related outcomes (i.e. health-related quality of life, parental stress, pain characterization, etc.). Instruments' files were then uploaded onto the MyeliNeuroGene private repository on GitHub as Pull Requests for review.

Scoring Algorithms: After instrument completion, a PHP scoring script was developed for instruments that required them. Automatic scoring reduces human error and dramatically decreases time spent on calculations. Scoring scripts were also uploaded onto the GitHub repository for review.

Instrument Implementation: After instruments and scoring scripts were reviewed and modified, the Pull Requests on GitHub were approved and the instruments made available on an insulated LORIS staging server where beta testing occurred. After testing was completed, instruments were pushed to the LORIS production server for instrument pipeline completion and data entry.

Participant Enrollment: Before data entry could be completed, Subject Profiles had to be entered. Our group has consented more than 1,000 patients and family members with different rare diseases since 2011, and patient and family recruitment is ongoing. To create a new profile, "Date of Birth", "Sex", "Site" (in the case of a multi-site study), and "Project" must be entered. Projects can be separated into different studies such as natural history, imaging, genetic, or even clinical trials assessing therapeutics. A new Subject Profile, or candidate, generates two identifier codes, a DCCID and a PSCID which are unique LORIS identifiers.

After the creation of the Subject Profile, each candidate was activated in the study, designated for which informed consent form was signed, and mapped to any external identifier codes. Under "Participant Status", we tracked the participant's status in the study (e.g. Active, Death, Lost to Follow-up, etc.). Comments can be entered with both time, date, and author history tracked in the internal audit trail. "Consent Status" tracks the latest signed Research Ethics Board (REB) approved informed consent form. Finally, mapping the "External Identifier" is crucial for future correspondence with family doctors and other collaborators.

Data Entry: "Create time point" allows for data entry of clinical, behavioral, and patient determined outcomes that were created during the Instrument Creation process. It also enables uploading of any imaging data collected. We customized our time points to correspond to the age of the patient. For instance, a participant's birth date would be time point T000, and a follow-up appointment 6 months later would be time point T006. A prenatal examination 1 month before a T000 examination would be designated as T-001. The steps to creating a time point can be seen in Figures 3, 4, and 5.

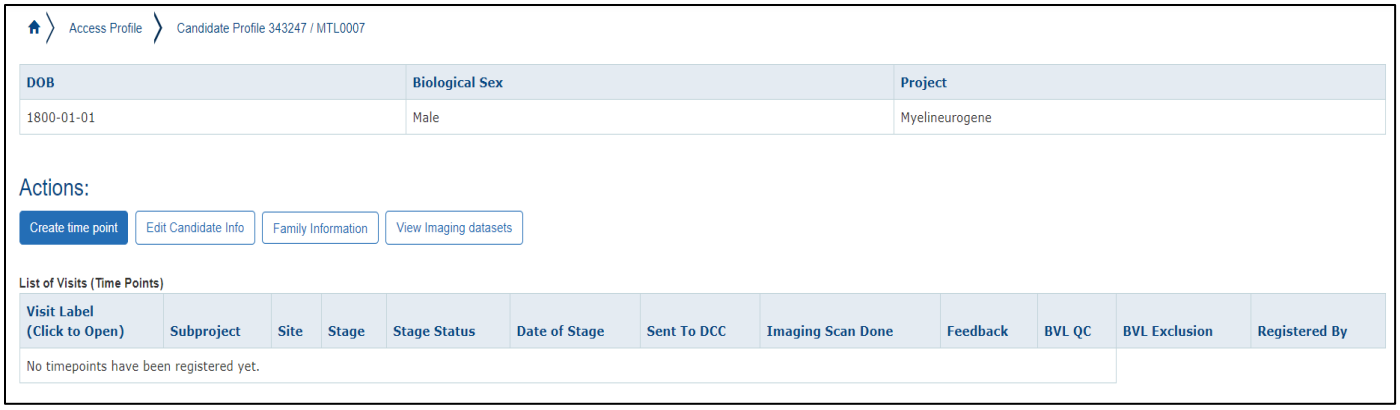


Figure 3: Creating longitudinal time points for patient visits

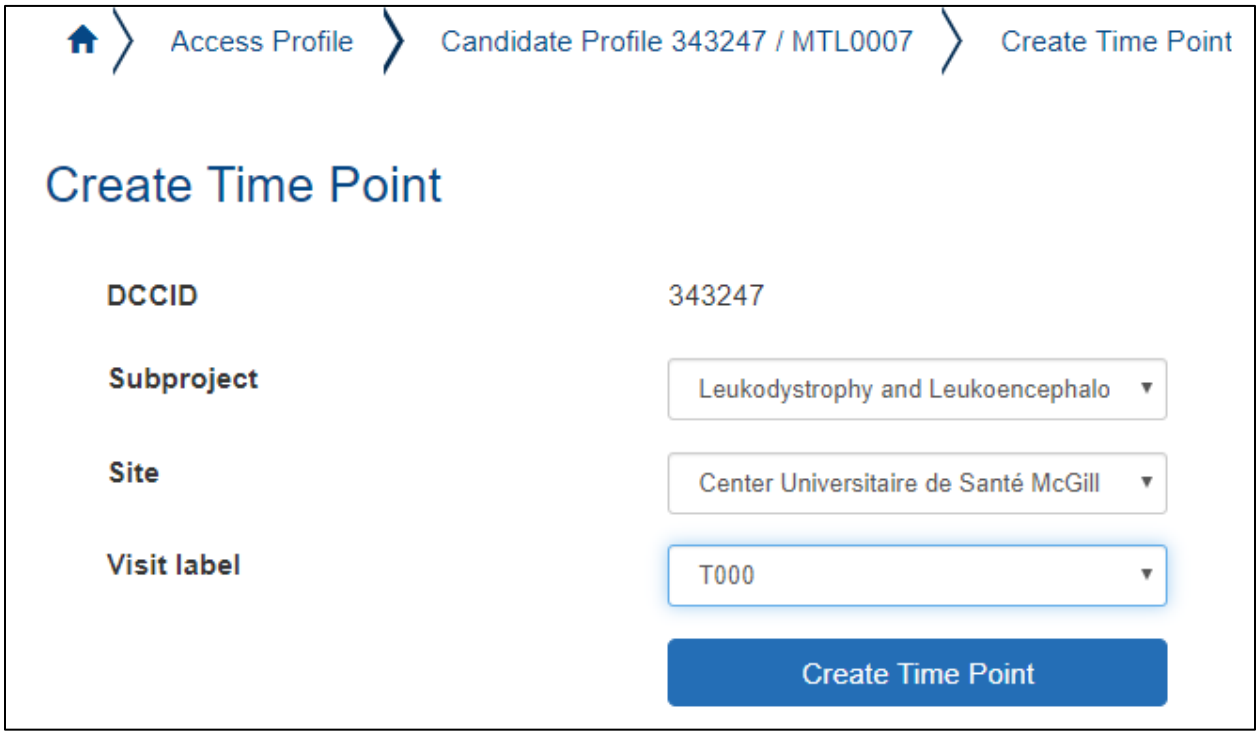


Figure 4: Associating time points with subprojects and study sites

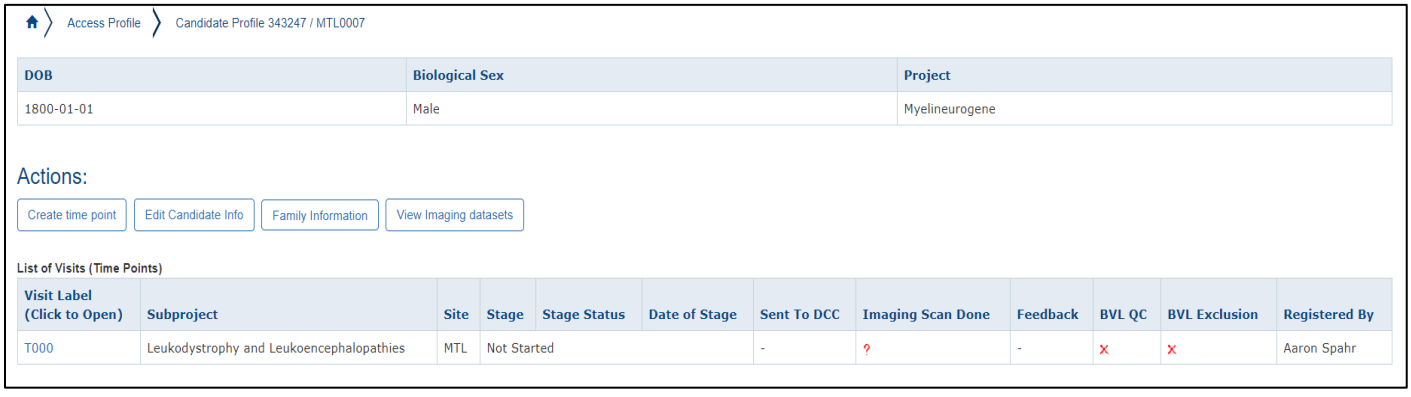


Figure 5: Visualizing time point information in the LORIS Candidate Profile

Selecting time point T000 opens a page for all instruments developed to work on our database (Figure 6). Time points can be customized so that only specific instruments are available to participants at specific ages. Entering multiple visits allows for prospective tracking.

Behavioral Battery of Instruments

Instrument (Click To Open)	Data Entry	Administration	Feedback	Double Data Entry Form	Double Data Entry Status
Family History	In Progress		-		
Perinatal History			-		
Developmental History			-		
Investigations			-		
Molecular Genetics			-		
Demographics			-		

Figure 6: Test battery of instruments customized for each participant based on time point and age appropriateness

Family Information: We have further customized LORIS to include Family Relationship information. Linking de-identified individuals allows us to link a given patient's disease characteristics to his/her parents' reported measures such as parental stress or patient/parents/sibling's quality of life. It also allows us to organize family genetic results when next generation sequencing (NGS) investigations are being conducted as well as any family/parent reported outcomes.

Declarations

Ethics approval and consent to participate: Written and informed consent was obtained from all research participants. This study was approved by the Research Ethics Board of the McGill University Health Center Research Institute (11-105-PED, 2019-4972)

Consent for publication: All participants have given consent for publication.

Availability of data and materials: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests: Our group, the MyeliNeuroGene Lab, is a collaborator of Dr. Alan C Evans' research group, the McGill Centre for Integrative Neuroscience, who developed LORIS, a free and open-source web-accessible database solution for multi-modal data and multi-site studies.

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article : This work was supported by the Montreal Children's Foundation (Estate of Daphne Dale Townsend), Fondation Les Amis d'Elliot, Fondation le Tout pour Loo, and Leuco-Action. Dr. Bernard has received the New Investigator Salary Award from the Canadian Institutes of Health Research (2017-2022). Aaron Spahr has received funding from the Desjardins Studentship in Child Health Research through the Research Institute of McGill University Health Centre (2018-2019), the Healthy Brains for Healthy Lives Graduate Student Fellowship (2019-2020), as well as a Graduate Excellence Award from the Integrated Program in Neuroscience at McGill University (2018-2019). None of the funding sources was relevant for study design, collection of data, analysis and interpretation of data, or writing of this manuscript.

Author's contributions: AS, ZR, ML, SD, ACE, GB were responsible for the conception and design of this project. AS developed, all of the instruments, with contributions from HT and expert guidance from GB, ZR, LD and ML. AS, ZR, and ML developed the scoring scripts. CL and MS helped perform quality control on the scoring scripts. CM developed the imaging platform. AS retrospectively entered over 1,000 participants and was responsible for initial data entry. AS generated the figures and the tables, with contributions from SF. AS drafted the initial manuscript, with contributions and modifications from SF and LT. ZR, ML, CM, SD, LT, SF, HT, LD, ACE, and GB provided feedback and all authors approved the current submitted version.

Acknowledgements: The authors wish to thank all the patients and families for their participation, time, and patience to complete questionnaires. The authors also wish to thank all collaborators and clinicians who referred patients, research would not be possible without you.

Author's information: GB is a pediatric neurologist and clinician scientist leading the Leukodystrophy and Neurometabolic Disorders Clinic at the McGill University Health Centre Research Institute.

List of Abbreviations

ADL	Activities of Daily Living
CFCS	Communication Function Classification System
CFR	Code of Federal Regulations
DCCID	LORIS Identifier
EDACS	Eating and Drinking Ability Classification System
FDA	Food and Drug Administration
FEES	Fiberoptic Endoscopic Evaluation of Swallowing
GDS	Global Dystonia Scale
GMFCS	Gross Motor Function Classification System
GNDS	Guy's Neurological Disability Scale
HID	Human Clinical Imaging Database
LORIS	Longitudinal Online Research and Imaging System
LONI	LONI Image Data Archive
MACS	Manual Ability Classification System
MAS	Modified Ashworth Scale
MRI	Magnetic Resonance Imaging
NGS	Next Generation Sequencing
OT	Occupational Therapist
PHP	PHP: Hypertext Preprocessor
PSCID	LORIS Identifier
PT	Physical Therpist
REB	Research Ethics Board
SaLT	Speech and Language Therapist
VFSS	Video Fluoroscopic Swallow Study

References

1. Aymé S, Urbero B, Oziel D, Lecouturier E, Biscarat AC. Information on rare diseases: the Orphanet project. La Revue de médecine interne. 1998;19:376S-7S.

2. Baird PA, Anderson TW, Newcombe HB, Lowry RB. Genetic disorders in children and young adults: a population study. American journal of human genetics. 1988;42(5):677.

3. Humphreys G. Coming together to combat rare diseases. SciELO Public Health; 2012.

4. McKusick VA. Mendelian Inheritance in Man and its online version, OMIM. The American Journal of Human Genetics. 2007;80(4):588-604.

5. Wakap SN, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. European Journal of Human Genetics. 2019:1-9.

6. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nature Reviews Genetics. 2013;14(10):681.

7. Sawyer SL, Hartley T, Dyment DA, Beaulieu CL, Schwartzentruber J, Smith A, et al. Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. Clinical genetics. 2016;89(3):275-84.

8. E Samuels M. Saturation of the human phenome. Current genomics. 2010;11(7):482-99.

9. Sardana D, Zhu C, Zhang M, Gudivada RC, Yang L, Jegga AG. Drug repositioning for orphan diseases. Briefings in bioinformatics. 2011;12(4):346-56.

10. Griggs RC, Batshaw M, Dunkle M, Gopal-Srivastava R, Kaye E, Krischer J, et al. Clinical research for rare disease: opportunities, challenges, and solutions. Molecular genetics and metabolism. 2009;96(1):20-6.

11. Helman G, Van Haren K, Bonkowsky JL, Bernard G, Pizzino A, Braverman N, et al. Disease specific therapies in leukodystrophies and leukoencephalopathies. Molecular genetics and metabolism. 2015;114(4):527-36.

12. Prevention CfDCa. Principles of Epidemiology in Public Health Practice, Third Edition: An Introduction to Applied Epidemiology and Biostatistics. cdc.gov: U.S. Department of Health and Human Services; 2006.

13. Hobbs BP, Sargent DJ, Carlin BP. Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. Bayesian analysis (Online). 2012;7(3):639.

14. Neuenschwander B, Capkun-Niggli G, Branson M, Spiegelhalter DJ. Summarizing historical information on controls in clinical trials. Clinical Trials. 2010;7(1):5-18.

15. Pocock SJ. The combination of randomized and historical controls in clinical trials. Journal of chronic diseases. 1976;29(3):175-88.

16. Viele K, Berry S, Neuenschwander B, Amzal B, Chen F, Enas N, et al. Use of historical control data for assessing treatment effects in clinical trials. Pharmaceutical statistics. 2014;13(1):41-54.

17. Courbier S, Dimond R, Bros-Facer V. Share and protect our health data: an evidence based approach to rare disease patients' perspectives on data sharing and data protection - quantitative survey and recommendations. Orphanet journal of rare diseases. 2019;14(1):175.

18. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. Journal of biomedical informatics. 2009;42(2):377-81.

19. Horvath MM, Winfield S, Evans S, Slopek S, Shang H, Ferranti J. The DEDUCE Guided Query tool: providing simplified access to clinical data for research and quality improvement. Journal of biomedical informatics. 2011;44(2):266-76.

20. Ozyurt IB, Keator DB, Wei D, Fennema-Notestine C, Pease KR, Bockholt J, et al. Federated web-accessible clinical data management within an extensible neuroimaging database. Neuroinformatics. 2010;8(4):231-49.

21. Adamson CL, Wood AG. DFBIdb: a software package for neuroimaging data management. Neuroinformatics. 2010;8(4):273-84.

22. Dinov I, Van Horn J, Lozev K, Magsipoc R, Petrosyan P, Liu Z, et al. Efficient, distributed and interactive neuroimaging data analysis using the LONI pipeline. Frontiers in neuroinformatics. 2009;3:22.

23. Bockholt HJ, Scully M, Courtney W, Rachakonda S, Scott A, Caprihan A, et al. Mining the mind research network: a novel framework for exploring large scale, heterogeneous translational neuroscience research data sources. Frontiers in neuroinformatics. 2010;3:36.

24. Gibaud B, Kassel G, Dojat M, Batrancourt B, Michel F, Gaignard A, et al., editors. NeuroLOG: sharing neuroimaging data using an ontology-based federated approach. AMIA Annual Symposium Proceedings; 2011: American Medical Informatics Association.

25. Das S, Zijdenbos AP, Harlap J, Vins D, Evans AC. LORIS: a web-based data management system for multi-center studies. Frontiers in neuroinformatics. 2011;5:37.

26. Das S, Glatard T, MacIntyre LC, Madjar C, Rogers C, Rousseau ME, et al. The MNI data-sharing and processing ecosystem. Neuroimage. 2016;124(Pt B):1188-95.

27. Das S, Glatard T, Rogers C, Saigle J, Paiva S, MacIntyre L, et al. Cyberinfrastructure for Open Science at the Montreal Neurological Institute. Frontiers in Neuroinformatics. 2017;10(53).

28. Das S, Lecours Boucher X, Rogers C, Makowski C, Chouinard-Decorte F, Oros Klein K, et al. Integration of “omics” Data and Phenotypic Data Within a Unified Extensible Multimodal Framework. Frontiers in Neuroinformatics. 2018;12(91).

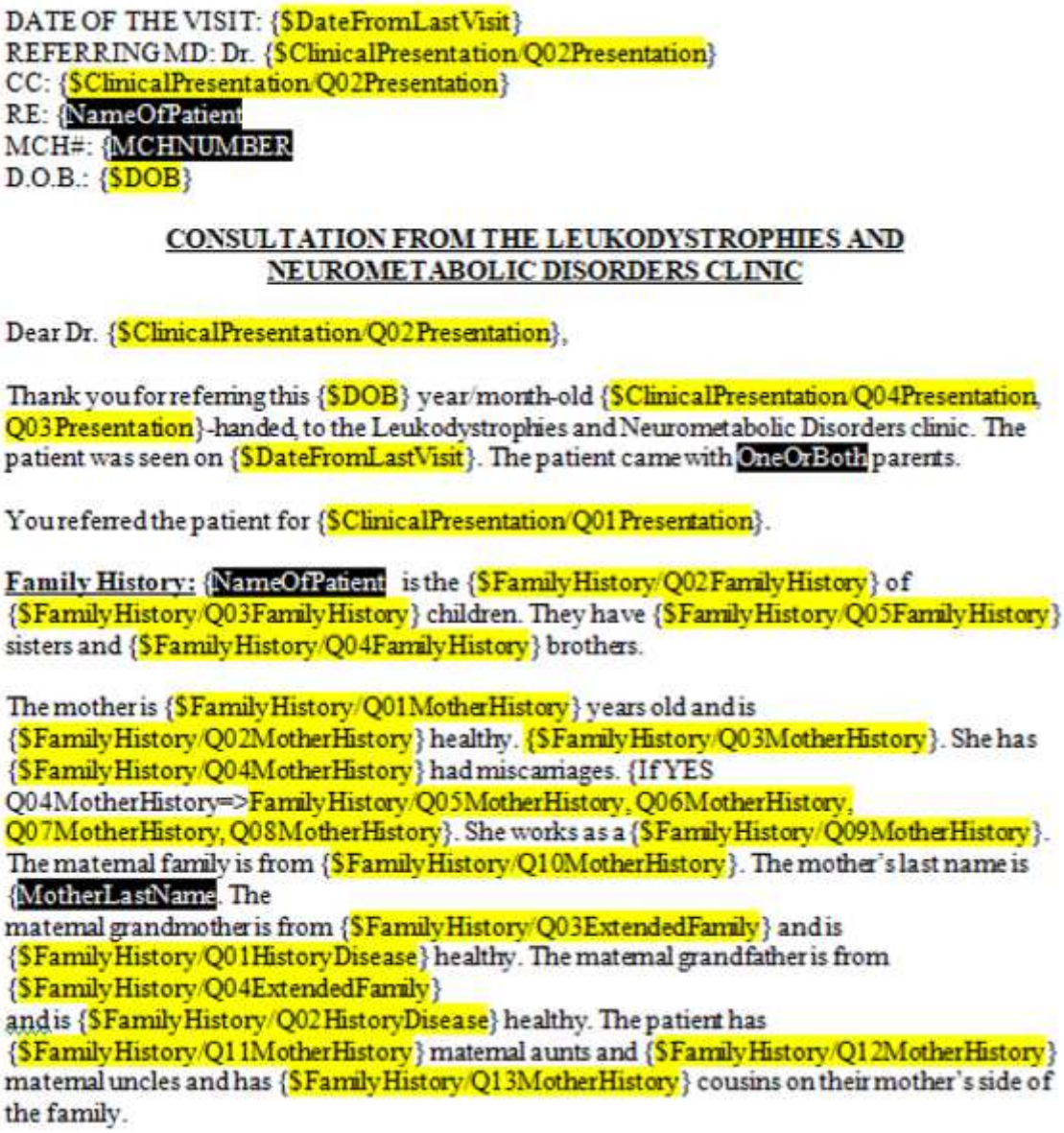
29. El Emam K, Rodgers S, Malin B. Anonymising and sharing individual patient data. bmj. 2015;350:h1139.

30. Administration USFaD. Guidance for Industry: Part 11, Electronic Records; Electronic Signatures - Scope and Application. In: Services USDoHaH, editor. fda.gov: U.S. Department of Health and Human Services; 2003. p. 12.

31. Administration UFaD. Rare diseases: natural history studies for drug development guidance for industry fda.gov: U.S. Department of Health and Human Services; 2019 [Available from: https://www.fda.gov/media/122425/download.

32. Administration USFD. Electronic Code of Federal Regulations [Webpage]. 62 FR 13464: U.S. Department of Health & Human services; 1997 [updated January 14, 2020.

Figures



DATE OF THE VISIT: {$DateFromLastVisit}
REFERRING MD: Dr. {$ClinicalPresentation/Q02Presentation}
CC: {$ClinicalPresentation/Q02Presentation}
RE: {NameOfPatient
MCH#: {MCHNUMBER
D.O.B.: {$DOB}

CONSULTATION FROM THE LEUKODYSTROPHIES AND NEUROMETABOLIC DISORDERS CLINIC

Dear Dr. {$ClinicalPresentation/Q02Presentation},

Thank you for referring this {$DOB} year/month-old {$ClinicalPresentation/Q04Presentation, Q03Presentation}-handed, to the Leukodystrophies and Neurometabolic Disorders clinic. The patient was seen on {$DateFromLastVisit}. The patient came with OneOrBoth parents.

You referred the patient for {$ClinicalPresentation/Q01Presentation}.

Family History: {NameOfPatient is the {$FamilyHistory/Q02FamilyHistory} of {$FamilyHistory/Q03FamilyHistory} children. They have {$FamilyHistory/Q05FamilyHistory} sisters and {$FamilyHistory/Q04FamilyHistory} brothers.

The mother is {$FamilyHistory/Q01MotherHistory} years old and is {$FamilyHistory/Q02MotherHistory} healthy. {$FamilyHistory/Q03MotherHistory}. She has {$FamilyHistory/Q04MotherHistory} had miscarriages. {If YES Q04MotherHistory=>FamilyHistory/Q05MotherHistory, Q06MotherHistory, Q07MotherHistory, Q08MotherHistory}. She works as a {$FamilyHistory/Q09MotherHistory}. The maternal family is from {$FamilyHistory/Q10MotherHistory}. The mother's last name is {MotherLastName. The maternal grandmother is from {$FamilyHistory/Q03ExtendedFamily} and is {$FamilyHistory/Q01HistoryDisease} healthy. The maternal grandfather is from {$FamilyHistory/Q04ExtendedFamily} and is {$FamilyHistory/Q02HistoryDisease} healthy. The patient has {$FamilyHistory/Q11MotherHistory} maternal aunts and {$FamilyHistory/Q12MotherHistory} maternal uncles and has {$FamilyHistory/Q13MotherHistory} cousins on their mother's side of the family.

Figure 1

creenshot of the LORIS MyeliNeuroGene dynamic letter generator: Yellow highlights customizable variables for the clinical letter generator. Black highlighted variables represent information that is not stored in LORIS and must be filled in by the physician.

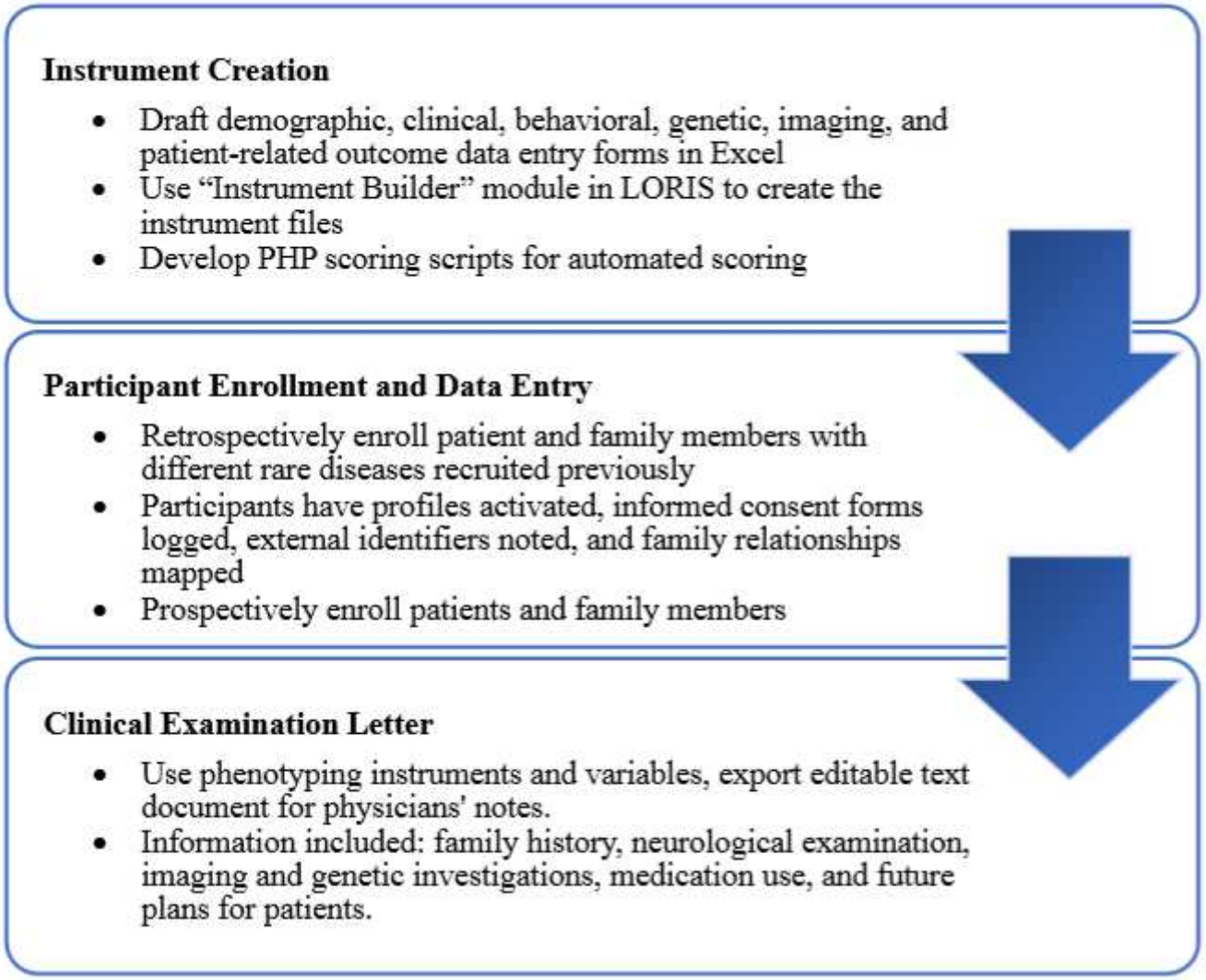


Figure 2

Database development workflow to create instruments, scoring algorithms, enroll patients, enter data, and output information into a clinical examination letter

Figure 3

Creating longitudinal time points for patient visits

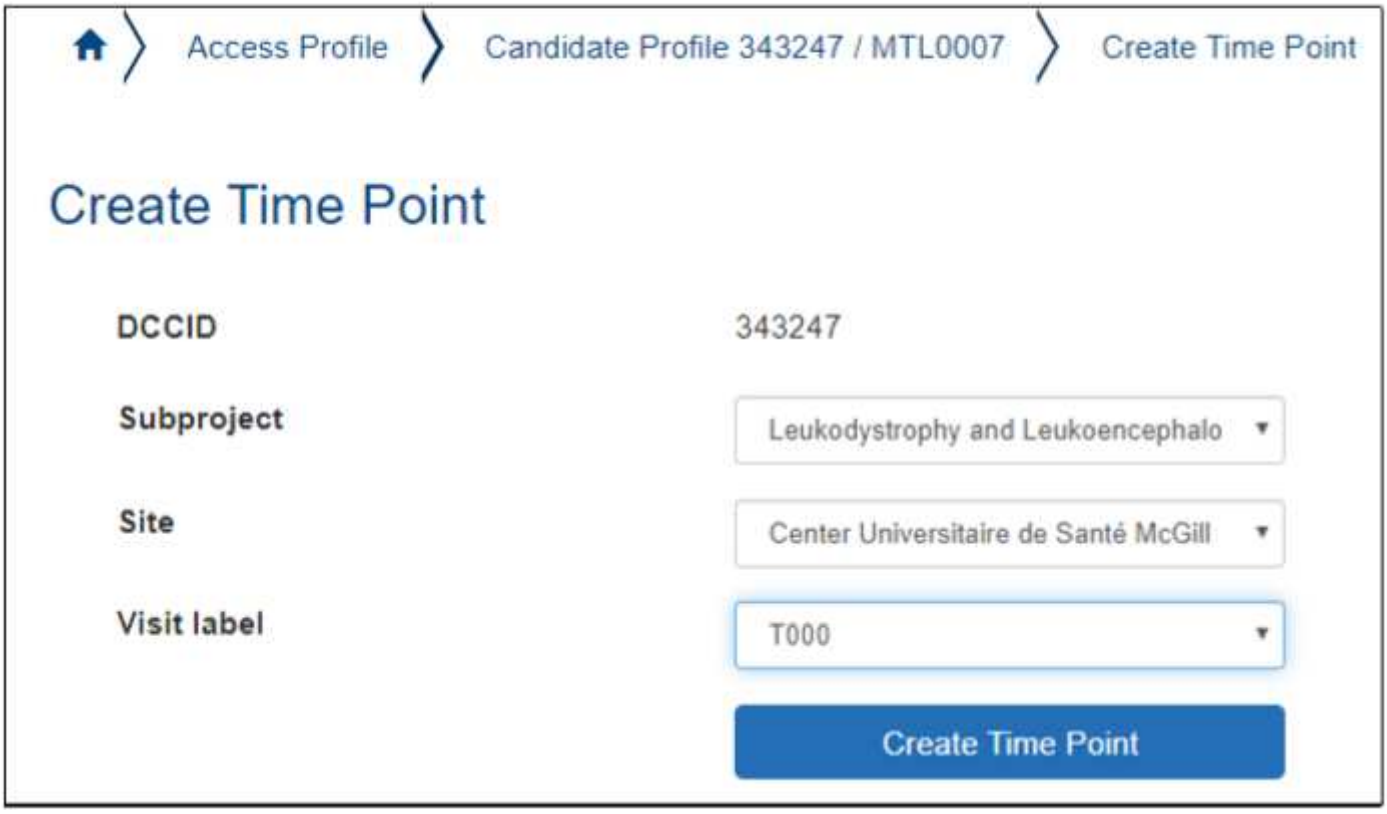


Figure 4

Associating time points with subprojects and study sites

Figure 5

Visualizing time point information in the LORIS Candidate Profile

Behavioral Battery of Instruments

Instrument (Click To Open)	Data Entry	Administration	Feedback	Double Data Entry Form	Double Data Entry Status
Family History	In Progress		-		
Perinatal History			-		
Developmental History			-		
Investigations			-		
Molecular Genetics			-		
Demographics			-		

Figure 6

Test battery of instruments customized for each participant based on time point and age appropriateness