RWD100

Evaluation of an Electronic Medical Records-Based Spinal Muscular Atrophy Registry for Outcomes Research Readiness: Results from a Comprehensive Gaps Assessment

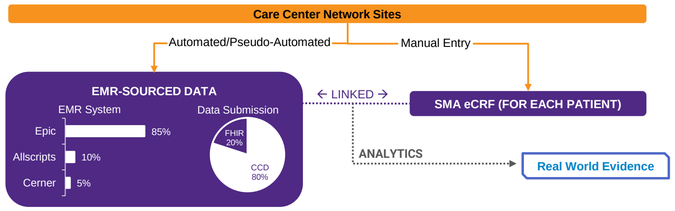
Sarah M Whitmire MS[1], Lisa Belter MPH[1], Alesa Monk MBA[1], Mary Curry ND[1], Jason Colquitt[2], Lorraine Marchand MA, MBA[2], Aaron Kamauu MD, MPH[2], Amanda Shields[2], and Mary Schroth MD, FAAP, FCCP[1]

[1] Cure SMA, Elk Grove Village, IL, US; [2] Across Healthcare, Carrollton, GA, US

Background

- Spinal Muscular Atrophy (SMA) is a rare genetic neurodegenerative disease, and real-world data (RWD) that can be used to understand the current disease landscape, evaluate outcomes, and create evidence to improve the standards of care is limited.
- In 2018, Cure SMA launched the SMA Clinical Data Registry (CDR), a RWD source with the goal of improving SMA standards of care.
- Cure SMA is a nonprofit patient advocacy organization that funds and directs comprehensive research that drives breakthroughs in treatment, advances access to high-quality care, provides practical support programs, and advocates for the needs of the SMA community.
- The CDR includes electronic medical record (EMR)-sourced data linked to an SMA-specific electronic case report form (eCRF) [Figure 1].

Figure 1: High Level Structure of the CDR



- As of March 2023, the CDR is comprised of data from ~850 consented patients with SMA who are receiving care from 20 of the SMA Care Center Network (CCN) sites across the United States; however, documentation patterns, EMR systems, and data transfer method vary across data-contributing sites, which may impact data quality.
- We sought to explore gaps in CDR processes to understand steps needed to improve data quality and to increase the confidence of future CDR analysis designs to support various use cases.

Methods

- We created a customized framework based on the CDR that included criteria that an EMR-sourced registry should have in place.
- In addition to internal knowledge and experience with the CDR, content from existing RWD frameworks, such as the Agency for Healthcare Research and Quality (AHRQ)[1,2], the Learning Health Network[3], and other published resources were incorporated[4].
- The compiled criteria fit within four categories:
  - **1. DOCUMENTATION AND SUPPLEMENTARY RESOURCES**
    criteria related to documentation of processes, guides, and training materials needed for transparency, consistency of analyses, and reduction of biases.
  - **2. DATA QUALITY**
    criteria related to ensuring high data quality, accuracy, completeness, and timeliness.
  - **3. FIT FOR PURPOSE (IDENTIFYING ELIGIBLE PATIENTS AND ASSESSING OUTCOMES)**
    criteria related to processes that need to be in place to set up and run analyses in a population of interest.
  - **4. INTEROPERABILITY, COMMUNICATION, AND ISSUE RESOLUTION:**
    criteria related to how information flows into the Clinical Data Registry, data-related communication back to the sites, and how to efficiently resolve issues.
- After the framework was complete, we assessed if the CDR met each criterion; If there was opportunity for improvement, a recommendation was provided on how to fill the gap.
- Finally, external RWD experts were consulted to add criterion, prioritize gaps, and provide recommendations for timing/implementation.

Results – Customized Framework

- In this assessment, we identified 44 criteria [Tables 1-4]:
  - Documentation and supplementary resources: 13
  - Data quality: 16
  - Fit for purpose: 9
  - Interoperability: 6

Table 1: Criteria in the Documentation and Supplementary Resources Category

DOCUMENTATION AND SUPPLEMENTARY RESOURCES
Full data dictionary: Data dictionary containing tables, variables, calculated variables, keys, data type, and for eCRF, skip logic/allowable choices - updated regularly
Data relational schema diagram: Document outlining all the tables, variables within them, and how the tables link together
Data quality plan: A plan outlining all quality control checks that are put in place, definitions for success, and frequency of checks
Registry eligibility requirements: Outlined eligibility requirements to participate in the registry and a way to identify individuals that do not meet those requirements
Required vs optional data: Clear documentation stating which variables are required for transfer
Best practices/ Data management manual: Best practices document for analyses as well as a data management manual to document how to handle and resolve various data issues
Data strategy document: Document outlining 3–5-year goals for the data and outlines the end goals for the CDR (e.g., time to treat analyses, multidisciplinary care)
eCRF user manual: A comprehensive manual that outlines how to complete the eCRF
Definition of all eCRF metrics: Clear and comprehensive operational definitions outlined for each question
Data/platform issues log: A way to track or log for all platform or data issues and resolutions
Data sharing standard operating procedure (SOP): Document outlining the de-identification procedures, transformations, data cleaning, and transfer method for data sharing of patient level data with external parties
CDR training materials: Training resources outlining setup, biases, limitations, etc. for anyone who is not familiar with the database
Data sharing data dictionary: Data dictionary for any external party receiving a de-identified data cut

Table 2: Criteria in the Data Quality Category

DATA QUALITY
Missing data checks: Data completeness checks run on a regular basis to flag missing data at multiple levels
eCRF question requirements: Method for determining complete forms vs partially complete forms and reducing the amount of missing data
External completeness benchmarks: Completeness benchmarks to appropriate gold standards
Ensure relevant data in tables: Ensure each table within the registry (e.g., patient, procedure, medication) contains appropriate data
Data quality by transfer type: Understand if there are any quality issues by data transfer method: CCD, FHIR (directly from EMR), FHIR (data warehouse or RedCap)
eCRF validation checks at entry: Method for flagging invalid data as it's being input into the electronic form
Invalid data flags: Method for flagging invalid data if it does get input into the registry
EMR data accuracy: Implementation of checks comparing data found in EMR to other sources (eCRF, patient chart, etc.)
Identify duplicate patients: A method/algorithm to identify patients that appear as two separate IDs (same patient consented at 2 sites contributing data)
QC automation (scalability): Automation of all quality control checks
Timely and consistent data: Consistent and timely data transfers from all sources
Accessible archived data: Archived versions of the full database saved regularly so that historical versions can be accessed anytime
Data tracking over time: Track the number of data points from each site longitudinally
eCRF data entry tracking: Tracking the individual who input the data into the eCRF
Data quality by transfer type: Comprehensive understanding of any issues with a particular method of data submission (e.g., FHIR vs CCD)

Table 3: Criteria in the Fit for Purpose Category

FIT FOR PURPOSE (IDENTIFYING ELIGIBLE PATIENTS AND ASSESSING OUTCOMES)
Linking tables: Method for linking patients together longitudinally across tables
Consistent identifiers: Consistent unique IDs that do not change over time
Missing tables per patient: Flagging patients from sites that do not contribute specific tables
Last updated date: Identifying last date of data transfer per patient
Inactive patient flag: Method for identifying inactive patients
Duplicate patient flags: Method for flagging duplicate patients for de-identified datasets where patient identifiers are not present
Measurable outcomes of interest: Dataset contains measures of interest that can be measured and analyzed
Metric standardization: Standardization or harmonization of metrics for analysis (e.g., coding systems such as ICD-10)
Data generalizability: Understanding of the generalizability of the dataset and how well it reflects the US population with SMA

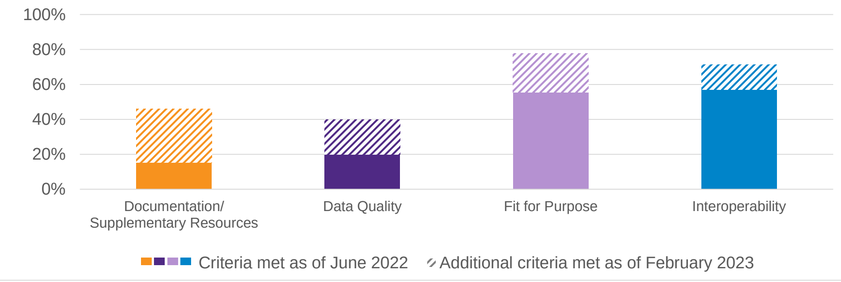
Table 4: Criteria in the Interoperability Category

INTEROPERABILITY, COMMUNICATION, AND ISSUE RESOLUTION
Site data access: CCN access to individual data for their site as well as aggregated data and graphics for assessing target outcomes and quality metrics
CCN ad hoc queries: A process for ad hoc queries able to be completed if requested by CCN site
Collect SMA data from EMR: All sites using automated data transfer methods with limited / no reliance on manual entry through supplementary eCRF
Regular site data quality touchpoints: Regular touchpoints with the CCN sites to discuss data-related issues
Site data-related point of contact: Identification of at least one individual to serve as a contact for IT/data related questions
Registry platform point of contact: A singular point of contact for the sites to reach out to for troubleshooting data issues

Results – CDR Scoring

- At the time of the gaps assessment (June 2022), the CDR had met 14/44 criteria; however, 10 additional criteria have since been met [Figure 2].

Figure 2: Criteria Met by the CDR within Each Gaps Assessment Category



- Once the gaps assessment and prioritization exercises were complete, we identified five high priority improvement areas, which encompassed multiple criteria [Figure 3].

Figure 3: High Priority Improvement Areas Identified for the CDR

Conclusion

- This CDR gaps assessment was meant to be used as a starting point and is a dynamically evolving assessment.
- The creation of any novel dataset with data sourced from multiple sources/pathways will always be met with unavoidable challenges, but an analysis of the system and processes surrounding the dataset can help reduce risks and improve quality and accuracy.
- This analysis highlights the importance of having the right processes in place to ensure the data is high quality, reliable, and appropriate for the research question of interest to minimize bias and maximize impact.

Acknowledgements

- Thank you to the SMA community for sharing their data and supporting the Cure SMA Care Center Network.
- Thank you to the Cure SMA Care Center Network for their commitment to improving care for people with SMA and contributing consented patient data to the SMA Clinical Data Registry. The Cure SMA Care Center Network includes 29 SMA Care Centers across the US who provide multidisciplinary care for people with SMA.
- Funding was provided by the Cure SMA Real World Evidence Collaboration, which includes Novartis Gene Therapies, Biogen, and Genentech/Roche.

The Cure SMA Real World Evidence Collaboration was established in 2021 to leverage the experience, expertise and resources of pharmaceutical and biotechnology companies and nonprofit organizations involved in development of SMA therapeutics to guide the future direction of real world data collection and use in SMA.

References

1. 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. Available from: https://www.ncbi.nlm.nih.gov/books/NBK208616/
2. Gliklich RE, Leavy MB, Dreyer NA, editors. Tools and Technologies for Registry Interoperability, Registries for Evaluating Patient Outcomes: A User's Guide, 3rd Edition, Addendum 2 [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2019 Oct. Report No.: 19(20)-EHC017-EF. PMID: 31891455.LHN
3. Lannon, C, Schuler, CL, Seid, M, et al. A maturity grid assessment tool for learning networks. *Learn Health Sys*. 2021; 5:e10232. https://doi.org/10.1002/lrh2.10232
4. Miksad, R.A. and Abernethy, A.P. (2018), Harnessing the Power of Real-World Evidence (RWE): A Checklist to Ensure Regulatory-Grade Data Quality. Clin. Pharmacol. Ther., 103: 202-205. https://doi.org/10.1002/cpt.946