Development of a multicenter EEG research consortium and data platform to advance Pediatric Quantitative EEG Research

Craig A. Press*[1], Giulia M. Benedetti*[2], Dana B. Harrar[3], Raquel Farias-Moeller[4], Stuart Tomko[5], Anuj Jayakar[6], Brian Appavu[7], Lindsay A. Morgan[8], Carlos Castillo-Pinto[8], Ajay X. Thomas[9], Daniel Davila-Williams[9], Stephanie MD Rau[2], Maelyn Fulton[2], Darshana Parikh[1], Nicholas S. Abend[1], Cecil Hahn[10], James Riviello[9], Saptharishi Lalgudi Ganesan[11], Darrell De Freitas[12], Yomi Dare[12], Joost Wagenaar[12]

1. Children's Hospital of Philadelphia, University of Pennsylvania; 2. C. S. Mott Children's Hospital, University of Michigan; 3. Children's National Hospital, George Washington University School of Medicine; 4. Children's Wisconsin, Medical College of Wisconsin; 5. St. Louis Children's Hospital, Washington University School of Medicine; 6. Nicklaus Children's Hospital; 7. Phoenix Children's Hospital, University of Arizona College of Medicine Phoenix; 8. Seattle Children's Hospital, University of Washington; 9. Texas Children's Hospital, Baylor College of Medicine; 10. The Hospital for Sick Children, University of Toronto; 11. Children's Hospital at London Health Sciences Centre, Western University; 12. University of Pennsylvania.

RATIONALE

Electroencephalography (EEG) is the most commonly used tool for acute brain monitoring. Quantitative EEG (qEEG) allows quantitative analysis, visualization, and bedside display of EEG data. Challenges to EEG research include de-identification of EEG files, inefficient system to share EEGs, lack of secure data platforms for waveform and clinical data, and non-standardized analytic pipelines.

METHODS

10 large pediatric academic centers convened to create a consortium to develop infrastructure for creating:

- Regulatory infrastructure connected to and managed within the research consortium, the Pediatric Quantitative EEG Strategic Taskforce (PedQuEST);
- Data extraction pipeline automating EEG data collection, deidentification, and transfer;
- State-of-the-art research platform for EEG storage and analysis linked to clinical data.

RESULTS

Figure 1: PedQuEST Data and Regulatory Management

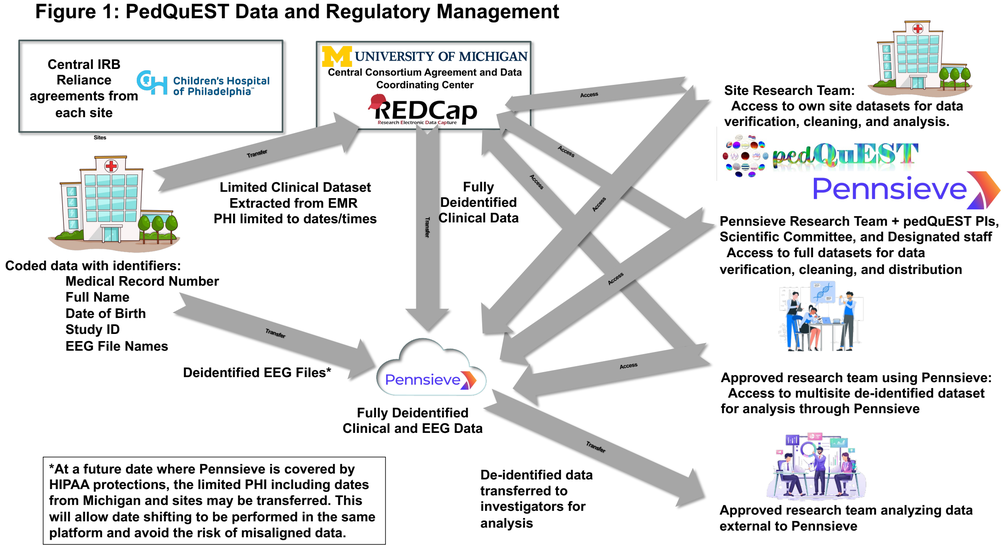
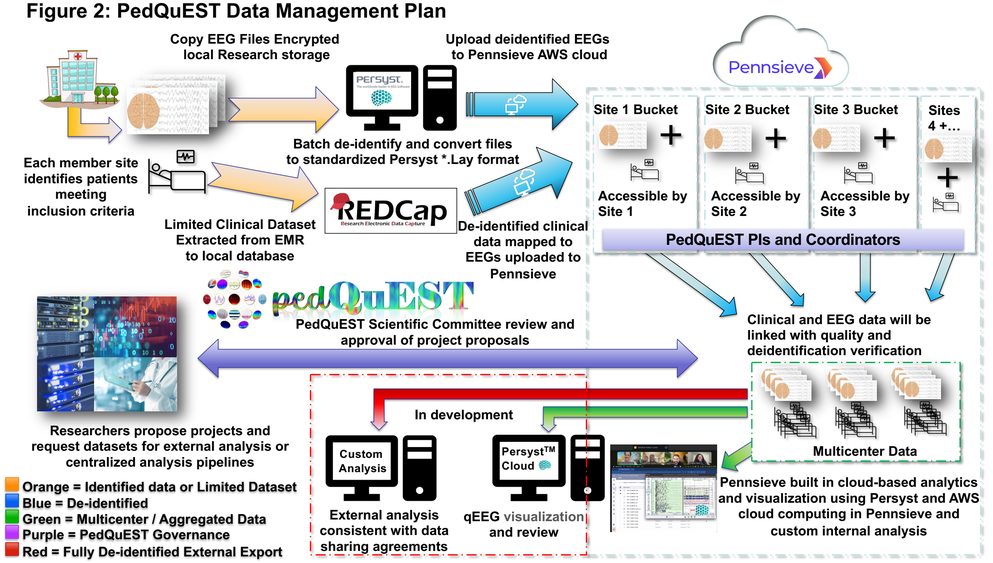


Figure 2: PedQuEST Data Management Plan



RESULTS

Regulatory:

- Centralized IRB and data-sharing agreement across 10 sites (Fig. 1)
- Governance for current and future projects
- Scientific Committee evaluates proposals and oversee studies
- Rapid site onboarding and new study activation
- Mentorship and regulatory support for investigators using PedQuEST.

Data Extraction Pipeline:

- Python-based pipeline using Persyst (15b, Research) on virtual compute nodes indexes all clinical EEGs and metadata
- Standardized de-identification and format conversion (*.lay,*.dat) for all sites.
- Preserves timestamps with date shifting while allowing realignment using days from birth
- Persyst "flat" layout ensures compatibility with clinical and research analytic platforms.
- Data integrity verified during transfer process

Data Platform:

- Collaborating with the Wagenaar lab, PedQuEST @ Pennsieve is a cloud system for multisite data curation, integration, and analysis (Fig 2).
- Investigators to access their own data securely and explore multisite datasets.
- Supports analytic pipelines and external data sharing.
- Developing centralized custom pipelines
- Clinical-platform processing and collaboration will enable rapid translation of qEEG research to bedside care.

Examples and Use Cases

- All sites have virtual machines to process EEGs running PedQuEST software
- 8/9 sites have uploaded data
- First data connection created at BCM-TCH
- Processed ~3.5K EEGs from 350 patients, ~3.75TB of data with minimal interaction over 4 days (<4hrs of work) vs. ~3.6 months of 40/hr/week effort assuming 10min/file
- Infrastructure supporting both a K-Award and PERF grant – subsequently added new site to IRB and consortium
- Two additional grants submitted using PedQuEST infrastructure
- Local nodes processed hundreds of EEGs for local studies
- Virtual Cloud based access to fully deidentified EEG for visualization across sites and centrally.

CONCLUSIONS

The development of a multicenter EEG data platform with efficient large scale data sharing and IRB regulatory approval for pilot and future studies eliminates many of the obstacles that have hindered pediatric EEG and qEEG research while maintaining data safety and integrity. The structure and governance will enable research that requires many subjects and EEG files from multicenter collaborations.

Funding and Acknowledgements

The Pediatric Epilepsy Research Foundation funded this work through their Grant for Infrastructure.

We also appreciate the support provided from Persyst, Michael Guess, Mark Sheuer, Nicolas Nierenberg, and Kady Colabrese, assisting with the ability to use software to universally read EEG formats and standardized file conversion.